ARUP

Oxera

Review of rail cross-industry interfaces, incentives, and structures

Options to reduce industry net cost

Prepared for
VfM Review team

September 30th 2010

Although this report was commissioned jointly by the Department for Transport (DfT) and the Office of Rail Regulation (ORR), the findings and recommendations are those of the authors and do not necessarily represent the views of the DfT and the ORR. While the DfT and the ORR have made all reasonable efforts to ensure the information in this document is accurate, the DfT and the ORR do not guarantee the accuracy, completeness or usefulness of that information; and cannot accept liability for any loss or damages of any kind resulting from reliance on the information or guidance this document contains.

Department for Transport
Great Minster House
76 Marsham Street
London SW1P 4DR
Telephone: 0300 330 3000
Website: www.dft.gov.uk

Office of Rail Regulation
1 Kemble Street
London
WC2B 4AN
Telephone: 020 7282 2000
Website: www.rail-reg.gov.uk

Executive summary

Oxera and Arup have been commissioned by the Department for Transport (DfT) to provide input into the Value for Money (VfM) Review. The scope of our work covers GB rail industry interfaces, incentives and structures and focuses on:

– a review of the current industry structure, contractual arrangements, interfaces and processes;
– an assessment of the impact of current arrangements on costs, incentives, decision-making and delivery;
– a review of best practice on cross-industry incentives and interfaces from other sectors;
– recommendations on how rail industry interfaces, incentives and structure can be improved to reduce industry net cost.

The first stage of our work investigated nine industry process case studies in order to understand the interfaces present between parties involved in the delivery of those processes, and the incentives across those interfaces. The processes studied in the case studies were:

– performance (delays, cancellations) improvement;
– possessions;
– station management, operation and upkeep;
– the specification, procurement and contract enforcement with respect to rail franchises;
– development of the timetable;
– the procurement and ongoing provision of rolling stock;
– long-term planning;
– the delivery of renewals and enhancements, including cross-industry projects;
– the allocation of capacity. This stage involved detailed research into legislation, regulatory documents and other relevant publications, supported by interviews with key industry players involved in delivering these processes within the current industry arrangements.

We assessed the interfaces and incentives associated with these processes with respect to incentive alignment and the degree of transaction costs. This assessment highlighted a number of areas of considerable concern:

– the franchise agreement, which is central to a number of incentive misalignments; in particular, fostering a focus on the contract rather than the end-user, and detaching franchised train operating companies (TOCs) from the costs and benefits of infrastructure interventions;
– incentives on Network Rail, both in relation to cost efficiency, and making the best use of the available capacity: Network Rail's cost base is an important component of the whole industry's cost, while it is itself relatively insulated from the revenue performance of its users;
– a lack of effective investment coordination in the sector, which increases direct and opportunity costs;
– stations, which might be considered a microcosm of current issues in the sector, where existing arrangements are complicated, time-consuming and appear unsatisfactory to all stakeholders.

A general conclusion to emerge from this review has been that, while transaction costs (defined as direct staff costs, as opposed to opportunity-based transaction costs) are significant in absolute cash terms, they do not in themselves appear to contribute as much to

net cost and general industry outcomes as do incentives. A number of sources of evidence supported this conclusion, including stakeholder interviews, evidence from other sectors (where the benefits of incentive alignment at the point of industry restructuring have far outweighed disbenefits associated with additional transaction costs), and evidence from studies of rail in Great Britain and elsewhere.

That is not to say that transaction costs should not be reduced in the sector; our conclusion is, instead, that specifically targeting a reduction in transaction costs to the bare minimum (through, for example, some form of vertical integration) rather than viewing them in the context of the whole cost base will not deliver the step change in industry net cost that is needed in today's constrained funding environment.

Two further general conclusions have emerged from this review, and they are closely linked to one another. First, it is clear that for any proposal for change to be successful, it must not demand a 'one-size-fits-all' approach—ie, it must not require the same change to the industry regardless of the rail markets being served. This principle of 'horses for courses' is a common thread throughout the options we have developed. Second, it is important to ensure that incentives are aligned to enable the provision of rail services to reflect changes in the markets being served.

Having identified these issues, our study then turned its attention to drawing best-practice examples and precedent from other sectors and other railways, and using this to develop a series of base options, recognising that ultimately elements of more than one option may be combined to reach the final solution. These options were designed to address the above issues, either alone or in combination, and ranged from incremental change within the current regulatory and legislative frameworks, to more radical structural reform. In all, seven options were evaluated against a counterfactual (base position) of no change to the status quo.

– 'Option 1a'—incremental change within the Control Period 4 (CP4) regulatory framework and existing legislative arrangements.

– 'Option 1b'—radical franchise reform, focusing on realignment of incentives between Network Rail and TOCs, and allowing Network Rail to raise equity.

– 'Horizontal separation'—structural separation of Network Rail into regional or market-based infrastructure managers (IMs, responsible for day-to-day operations, maintenance, renewals and enhancements), and a 'shallow' system operator (SO) responsible for planning, timetable development and national operations. This option focuses on enhancing cost efficiency and better use of the network, while minimising operational interfaces.

– 'Vertical integration'—the creation of vertically integrated rail concessions, along similar geographies as in the market-oriented horizontal separation option.

– 'Railways Agency'—taking Network Rail into public ownership, and creating a Railways Agency to let infrastructure concessions and passenger rail franchises. This option has the overall focus of introducing contestability into network activities, and would also see TOCs allocating an enhancements budget over the life of a franchise.

– 'Nationalisation'—the creation of state-owned passenger operations (OpCo) and IM companies.

The first stage in option evaluation was to identify the impact that each would have on different industry players, and how responsibilities for different aspects of the railway would be distributed among these organisations. The second stage described the specific problems that the option would address by the changes proposed in the first stage, and used the

available evidence to test hypotheses around potential efficacy. Options were then ranked on the basis of potential savings in **annual** net cost resulting from incentive realignment (on a scale ranging from low: £10m–£100m, to medium: £100m–£500m, to high: £500m+), and the likely duration and cost associated with transition. The results of this exercise are as follows.

– The franchise reform option offers a high positive impact on net cost. Its main focus is on bridging the incentives gap between infrastructure cost and revenues from end-users, and also benefits from the equity financing of Network Rail. This option builds on the incremental change option.

– Horizontal separation has a medium to high positive impact on net cost. Its main focus is on incentivising infrastructure cost efficiency and revealing the costs of providing rail services in different markets, enabling more market-oriented interventions. The SO function also offers incentives for better use of the network. Horizontal separation was evaluated on both a 'regions' basis and a 'markets' basis—in practice, the distinction between the two options in terms of implementation and evaluation is limited.

– Vertical integration has a medium impact on net cost following a costly transition period, enabling incentive realignment between operations and infrastructure, but at a cost of lost competition.

– The Railways Agency option has a medium impact on net cost following a costly transition period. Its focus is on providing contestability with respect to infrastructure interventions.

– The incremental change option has a low impact on net cost. Its focus is on change within the regulatory and legislative status quo, incurring moderate costs, but enabling only moderate benefits.

– Nationalisation has a negative impact on net cost due to the loss of competition and the creation of a public sector monopoly.

The ranking demonstrates that our radical franchise reform and horizontal separation options provide the greatest opportunities for savings in net cost, and involve proportionate time and costs associated with transition. The key assumptions involved in each of these options are as follows (options are listed broadly in terms of expected net benefit).

Radical franchise reform

– That providing TOCs with a long-term business and a different form of risk profile, where they are able to take 'normal' decisions to grow revenues and minimise costs, would deliver considerable net cost savings.

– That more accountable operators would make good investment and day-to-day decisions around stations, service patterns and rolling stock.

– That incentivising TOCs in respect of Network Rail's costs would deliver greater efficiencies, more quickly. Similarly, Network Rail's stronger volume incentives would enable operators to maximise demand.

– That Network Rail introducing equity would strengthen cost-efficiency incentives.

– That operators would be better able to respond to demand shocks, and would require lower bid margins than under the status quo, except for the infrastructure cost risk they would assume.

Horizontal separation

– That the increased interfaces would be outweighed by the efficiency potential from separation, and that these interfaces might be ameliorated by the SO function, and some franchise mergers.

– That the transition from Network Rail to the new IMs/SO could be worked out smoothly, especially in relation to Network Rail debt, the recruitment of senior staff, and employee relations.

– That effective, high-level incentives could be introduced on the SO for maximising the use of the network.

– That comparative competition would deliver benefits of the order seen in gas distribution networks (DNs), despite Network Rail reorganising to facilitate benchmarking between delivery units in the counterfactual (and despite the presence of subsidy). It might be the case that the benefits are even greater than seen in the gas DNs context, given alignment with markets and TOCs.

– That retaining the planning function within the SO (mini-Network Rail) could be efficient and effective, and that such a body could command industry respect in this function.

It should be noted that, due to Network Rail equity issuance being included in the franchise reform option, the sums in the above ranking are **not** strictly additive. However, some addition of the net cost savings would be appropriate if both options were to be introduced.

Other options may be considered preferable in specific circumstances as a means of building on this preferred solution. In particular, a 'horses-for-courses' approach could enable the most appropriate solution to emerge for particular local geographies, and for particular rail markets.

The study team has developed the following potential roadmap for introducing preferred options, should the more detailed work required to refine the options and pin down more precisely their costs and benefits (more than is possible in an eight-week study), reinforce our conclusions.

– Stations: we propose that the forthcoming franchise re-letting exercises offer opportunities to test the market in relation to alternative station management models. These would focus in particular on 99-year full repairing leases for the franchisee, with an increase in Network Rail's managed stations portfolio (see the franchise reform option); and another model, whereby Network Rail focuses on station fabric, with TOCs taking responsibility for customer-facing assets.

– Fares regulation: our suggested move towards economic regulation within the DfT and Transport Scotland fares policy can be started immediately, for implementation in 2011.

– Radical franchise reform: implemented via franchise change for TOCs with over two years to run, and the rest changed at re-letting. An evolution through the first few franchises would ease implementation, and enable the detail of risk-sharing arrangements to be developed. A licence change would be required to enable Network Rail revenue sharing.

– Horizontal separation: begin the process of separating price controls immediately. In terms of implementation, one option would be short-term nationalisation, followed by restructuring to split out SO and IM functions, and for concessions to be established. An alternative would be for Network Rail to carry this out itself via the sales of separated businesses, as occurred in the gas DN case. Our expectation is that this could be

implemented for 2014, and available for examination during PR13, although further work to develop the detail would be required.

A number of elements of this programme can be started or implemented, and would begin to deliver benefits, within a year:

– transfer of station management to TOCs;
– changes in fares regulation, so that fares are regulated on an economic basis;
– new franchise agreement for upcoming competitions;
– separation of price controls.

Contents

List of tables

List of figures

1 Introduction

Oxera and Arup were commissioned by the Department for Transport (DfT) to review rail cross-industry interfaces, incentives and structures. This report presents the results of our analysis.

The scope of this work was to provide:

– a review of the current industry structure, contractual arrangements, interfaces and processes;
– an assessment of the impact of current arrangements on costs, incentives, decision-making and delivery;
– a review of best practice on cross-industry incentives and interfaces from other sectors;
– recommendations on how rail industry interfaces, incentives and structure can be improved to reduce industry net cost.

Our findings are part of the overall work that the DfT and Office of Rail Regulation (ORR) are undertaking as part of their Value for Money (VfM) Review, for which the terms of reference were published on December 9th 2009.[1]

Our work relates only to the scope set out above, and does not address all of the avenues of inquiry being pursued by the VfM Review team. Our work reflects the time, information and explanations made available to us, and we recognise that the VfM Review team will evaluate the conclusions of this report in the context of the wider terms of reference of the VfM Review process, which is due to conclude and issue its findings in spring 2011.

Our approach to addressing the agreed scope of work within the time available to us (less than three months) is summarised in Figure 1.1. The Oxera–Arup team reviewed industry practices with a view to identifying a list of processes and their associated interfaces that have material consequences for incentive alignment and net cost in the industry.[2] A list of nine industry process case studies—set out in the left-hand column of Figure 1.1—was agreed with the Review team. The impact of these processes was assessed using a variety of different evidence sources, including interviews with a range of industry stakeholders.

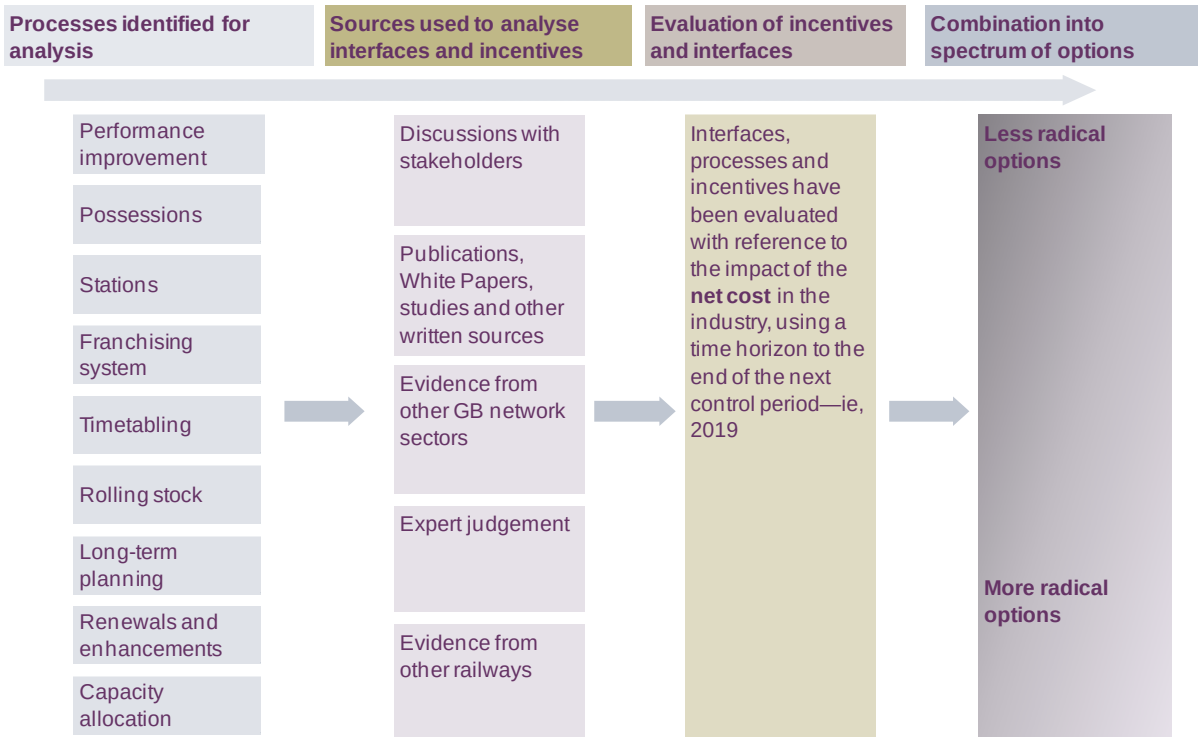
This evaluation focused on incentive alignment and transaction costs across industry interfaces. Areas of significant alignment and misalignment were identified, and were used as the basis on which a spectrum of coherent, self-contained options was constructed to redress associated failings and to protect and reinforce successes in the context of existing levels of safety. These options ranged from incremental changes within the existing regulatory framework (such as the Control Period 4 (CP4) settlement) to more radical changes that would constitute wide-ranging structural interventions.

The spectrum of options was evaluated against a list of criteria intended to elicit their implications for industry net cost, allowing for a degree of recombination between elements of options as appropriate.

Oxera and Arup are pleased to acknowledge the assistance rendered to our work by the wider industry. We are grateful for the interviews, workshops, discussions and written submissions that were provided in the course of this project.

[1] The terms of reference for the VfM work are available here: http://www.dft.gov.uk/pgr/rail/strategyfinance/valueformoney.
[2] As such, as some important 'outward-facing' processes were outside the scope of this study.

Figure 1.1 Description of the project process



Source: Oxera/Arup.

This paper elaborates on each of the steps described in Figure 1.1 as follows.

– The remainder of this section providers a brief overview of the context of this study, touching on the structure of the GB rail industry, and presents a case study that illustrates the role of incentives and structure in determining net cost outcomes in the industry.

– Section 2 identifies areas where incentive misalignment and/or transaction costs affect the outcomes associated with the nine process case studies studied.

– Section 3 describes the means by which options for change were constructed, and the criteria used to evaluate them.

– Section 4 describes and evaluates options for change, and concludes with a set of options for change.

The appendices contain supporting material to this analysis:

– Appendix 1 briefly reviews the general literature on incentives and vertical structure as a means to motivate the exploration of these issues elsewhere in the paper, and reviews literature on the studies of international rail industry structure

– Appendix 2 contains a review of the contractual arrangements and interfaces for the nine key industry process case studies set out in Figure 1.1.

– Appendix 3 contains a review of relevant experiences from other sectors, focusing on three GB network utility industries (energy, aviation and water).

1.1 Industry structure and context

The British rail industry was radically restructured at the time of privatisation, which led to a fragmented industry with a large number of interfaces.[3]

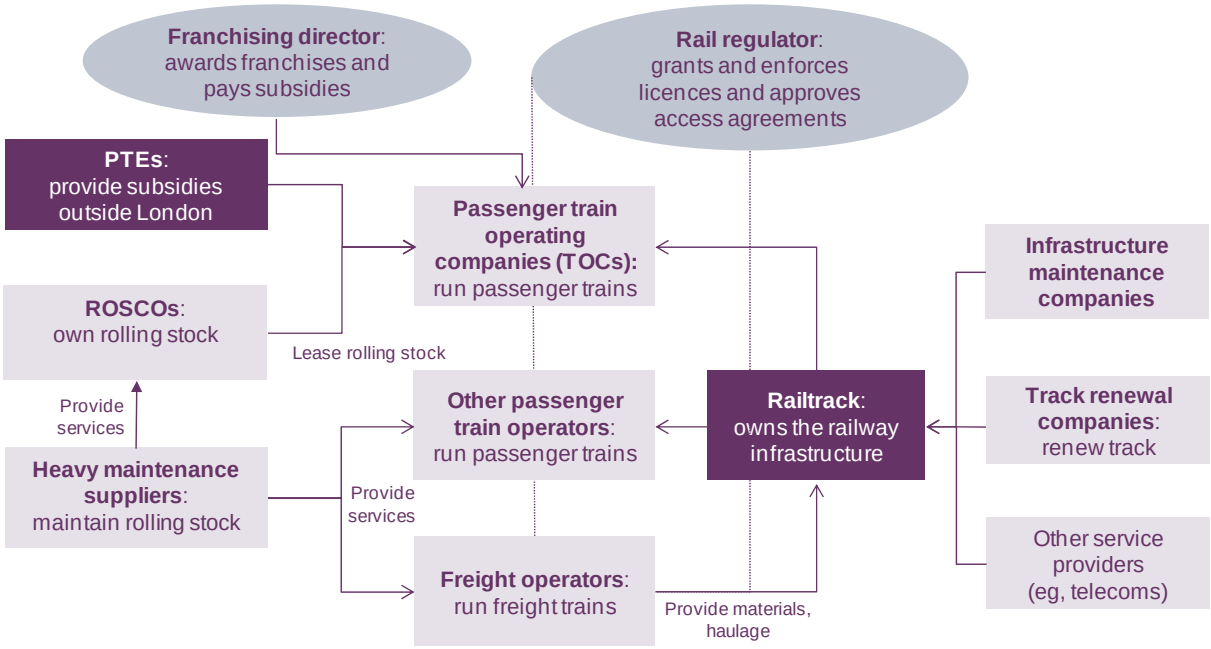
The enduring reforms of the 1990s that gave effect to privatisation and the new industry structure had two main characteristics: the first being the separation of the ownership of infrastructure and the operation of the rail network; and the second that the rail industry became a complex contractual matrix. In particular, contracts to operate passenger services were initially negotiated with train operating companies (TOCs), which were divided into franchises. Subsequently, these franchises were let by competition.

The former British Rail was restructured into one track authority (Railtrack), 25 franchised passenger TOCs, seven freight operating companies (FOCs), three rolling stock leasing companies (ROSCOs) and 70 ancillary businesses, with additional regulatory bodies responsible for the administration of the network.

On the administrative side there was: the Office of Passenger Rail Franchising (OPRAF), responsible for franchising passenger services; the Office of the Rail Regulator, responsible for issuing licences to run services, approving franchising agreements and enforcing domestic competition law; and HM Railway Inspectorate (HMRI), the safety regulator.

Figure 1.2 summarises the structure of the industry immediately after privatisation.

Figure 1.2 British rail industry structure after privatisation



Source: Thompson, L.S. (2004), 'Privatizing British Railways: Are there Lessons for the World Bank and its Borrowers?', *Transport Papers*, T.S. Board, Washington, D.C., The World Bank.

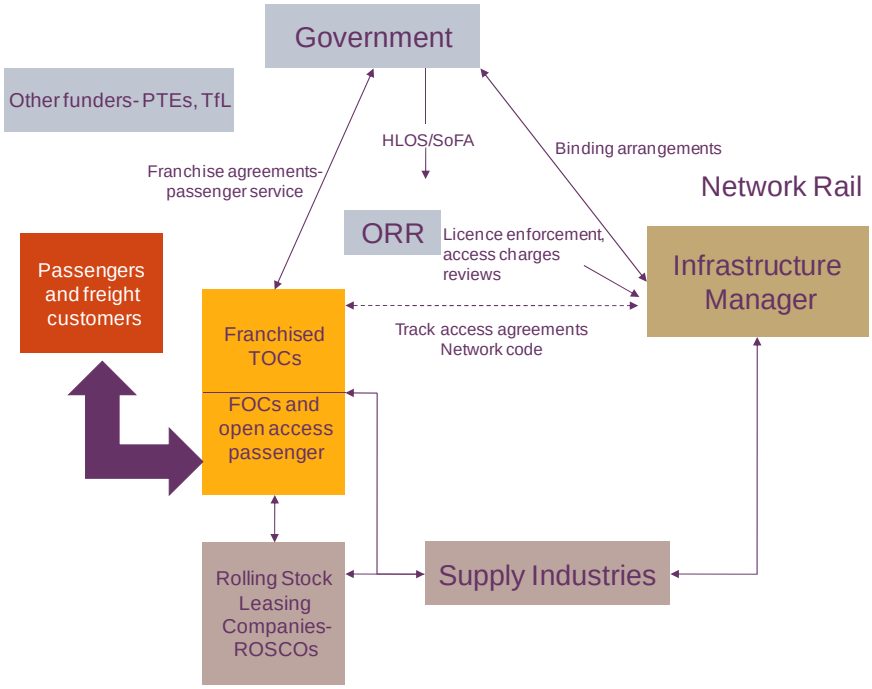
Since privatisation, the industry has undergone a number of structural changes, although the fundamental separation between infrastructure ownership and track operation remains. Post-1997, the creation of the Strategic Rail Authority (SRA) was placed at the centre of the Labour government's restructuring of the industry. The SRA had a wide mandate—from

3 For further discussion, see Harris, N. and Godward, E. (1997), *The Privatisation of British Rail*, London: Railway Consultancy Press.

developing a programme to deliver policy objectives, to regulating passenger network benefits and franchising passenger services, to allocating funds to the Rail Passenger Councils, the Passenger Transport Executives and to the freight industry.

The industry structure changed again after the Railtrack special administration and the transfer of infrastructure management to Network Rail. The current main stakeholders in the industry are the DfT (and other funders, including Transport Scotland), Network Rail, the ORR, and the TOCs, FOCs and ROSCOs (and their associated supply chain).

Figure 1.3 British rail industry structure



Source: ORR.

To some extent this is a simpler industry structure than that which prevailed following privatisation. Nevertheless, the industry remains fragmented, and its functioning depends on the successful operation of a number of interfaces.

1.2 Incentives, industry structure and net cost

The section briefly reviews the conditions under which interfaces and a separated industry structure might be expected to deliver appropriate outcomes, and the reasons for concerns about net cost in the GB rail industry in this context. This will be illustrated by means of a case study on track-friendly trains.

Box 1.1 Identifying and defining an incentive

In everyday language, an incentive is a thing that encourages and motivates someone to do something; any decision that individuals (or companies) make consists of weighing different options and picking the one that yields the highest benefit. The decision process of firms therefore involves comparing the benefits associated with different actions (at the same time, any project is compared with the firm's outside option; namely, to undertake no new action). This concept is captured by the incentive-compatibility constraint, which means that actions will not be undertaken where they do not result in net benefits accruing to the party undertaking the action.

No profit-maximising firm would take any action that violates a relevant incentive-compatibility constraint. For example, a TOC on a seven-year franchise may not find it incentive compatible to contribute to the funding of a major station redevelopment if it is not guaranteed the prospect of recovering a commercial return on the investment (because the franchise may have been reassigned before returns can be captured).

That aside, it is sometimes advisable for society as a whole that firms take actions that would otherwise be incompatible with their particular incentive constraint. Regulators and other authorities cannot oblige firms to provide services at a loss for a sustained period of time. What regulators and governments do in practice is to operate directly on a firm's incentive constraint in order to render socially desirable—yet incentive-incompatible—courses of action incentive-compatible. This is attained through incentive schemes, which are collections of rules ensuring that the best way for firms to make a profit is to abide by societal interests. For instance, franchise specifications that compensate for unprofitable routes and penalise high delays by rescinding the franchise could be seen as examples of this thinking in the GB rail industry.

The nature of the incentives prevailing in an industry such as rail is therefore an important driver of outcomes, since it is incentive schemes (in the form of contractual obligations or a price settlement) that oblige companies to undertake particular types of activity that are judged to be desirable.

Similarly, industry structure can also have an impact on net cost, since it may enhance or frustrate the effectiveness of particular types of incentive, as well as having more general effects that are distinct from incentives per se. For example, a fragmented industry structure—such as GB rail—may make efficient coordination of activities more difficult to achieve and increase industry net cost. This industry structure may create more barriers to investment than other structures, since different participants may not be confident about recovering a return on their investment if cooperation with multiple external parties is required. It is notable that the experience of separation in GB rail has been significantly more problematic than the experience in other sectors.[4]

These issues are illustrated below by means of a case study on track-friendly trains, which highlights some of the issues with incentives and interfaces in the GB rail industry.

1.2.1 Case study: Track-friendly trains[5]

Background

Rolling contact fatigue (RCF) is damage to rails (typically crack formation around the gauge corner of the rail) arising from wheel–rail contact. It is a major cause of the need for maintenance and renewal (principally premature rail replacement, rail defect removal and grinding), and in extreme cases can lead to failure of the track. The incidence of RCF has increased significantly in recent years, driven by characteristics of modern rolling stock—in particular, increasing weight and more aggressive steering characteristics (driven by the need for greater stability at high speed and lower vehicle maintenance costs).

Network Rail states that it has increased its rail grinding programme and taken other mitigating actions in response to this, but this has been at the expense of increased maintenance expenditure. There is no other economic, technical solution available on the 'rail' side of the wheel–rail interface, and this has led to a search for a solution on the 'wheel' side to mitigate the stress effects.

The problem has been particularly noticeable in the Wessex area for three reasons.

[4] This conclusion emerges from the review of experiences in other sectors that is described in Appendix 3.
[5] We gratefully acknowledge information provided by Network Rail in drafting this case study.

– A large fleet replacement programme was undertaken in Wessex from 2002, where the old Mk1 slam door stock was replaced in its entirety. This old stock had a relatively benign steering characteristic and generated little RCF. Approximately 700 new Desiro vehicles replaced the old stock in a short period of time.

– The new Desiro fleet, as operated by South West Trains (SWT), has a particularly aggressive steering characteristic and causes much more RCF than the predecessor stock.

– Third-rail electrification in the area means that it is difficult to use a high-output track grinder; instead, a bespoke fleet of smaller grinders is in use, increasing maintenance time and cost.

The proposed solution

The particular Desiro characteristic that causes the problem is the stiffness of the rubber primary yaw suspension (PYS) bushes used. This high stiffness improves stability at high speeds and successfully minimises vehicle maintenance costs, but causes much greater damage to the rail.

A few years ago, Freudenberg Schwab, a German firm, developed a hydraulic PYS bush which gives the same vehicle stability at high frequencies, but provides a 'softer' response at lower frequencies, significantly reducing RCF. This PYS bush can be retrofitted to Desiro vehicles, and probably to many other modern multiple units and coaches, and could be fitted to new rolling stock. (It is unlikely to provide a solution for complex bogies such as seen in locomotives and tilting trains). The track-friendly trains project was set up to forward the retrofitting of this new bush to Desiro vehicles (and potentially other vehicle types) as the main vehicle-based solution to RCF (although other solutions are being progressed as well).

Business case

The cost of the new bushes is approximately €1,500 each (approximately £1,250), giving a component cost of £10,000 per vehicle. They are around five times the cost of conventional rubber bushes, giving an incremental cost per vehicle of around £8,000. If the bushes are fitted during a normal cyclical maintenance programme there are no incremental fitting costs. For the entire Desiro fleet of some 1,400 vehicles nationwide, incremental fitment costs would therefore be of the order of £11m. To fund this change, Network Rail states that it has established an incentive for TOCs via reductions in the Variable Track Access Charge (VTAC) for modified vehicles: this applies to any modification that makes the trains more track-friendly.

The estimated annual rail maintenance savings are estimated by Network Rail to be significant—of the order of £10–£20m per annum—resulting in a favourable whole-industry business case.

Interface issues

In spite of the very favourable whole-industry business case, taking this initiative forward has been problematic because of the clear disconnect between where the costs fall (TOCs/ROSCOs) and where the physical benefits lie (Network Rail).

The current industry mechanisms do give a means of transferring some of the benefit back to the TOCs: the reduction in damage to the track means that TOCs will see a reduction in VTAC. Considerable work has gone into establishing the framework for charges—the mechanism has been endorsed by the ORR and provides a clear basis for TOCs to develop their business case, although there is some simplification inherent in the network-wide nature of the charges. However, under this arrangement it will take around three years before a typical TOC sees a payback on the PYS bush investment. This immediately means that it is only attractive for franchises which have a number of years to run (this includes SWT). However, even for these franchises, the fact that the benefits will not begin to accrue until the

next control period means that there is no certainty of any payback beyond the next periodic review as a change in the structure of charges could have the effect of wiping out the reduction in VTAC.

Network Rail, the ORR and TOCs have held discussions regarding a mechanism for ensuring that TOCs continue to see the reduction in VTAC, but no firm way forward has emerged from these. It is therefore proposing to offer SWT a letter of comfort, effectively guaranteeing the TOC's benefits over the life of the franchise, and taking the risk of any adverse effect of the periodic review itself in order to achieve the benefits of the scheme.

Thus, although the scheme is likely to go ahead, it has taken some considerable time to reach a means of implementation, with associated opportunity cost (the new bushes could have been fitted as long as three years ago).

Conclusions

This case study highlights that in the current industry structure, projects that have a clear overall business case, but with costs and benefits falling to different parties, can be difficult to implement, as even when there are incentive mechanisms they can fail to operate as needed, and there is no clear owner of the industry business case. For example, the particular causes of problems in this case appear to be attributable to the manner in which the DfT applies protection to changes in charges in the latest model form contract.

1.3 Conclusion

This section has introduced the scope of this study, which focuses on the interfaces, incentives and structures of the GB rail industry. These three features interact with each other to produce the observed outcomes in the industry. The case study on track-friendly trains provides one example of where net cost seems to be higher than it would be under alternative approaches. The following section presents the findings of a review of nine key industry process case studies to identify where interfaces, incentives and structures might be altered to secure reductions in industry net costs.

2 Assessment of the impact of current arrangements on incentives and costs

This section summarises identified misalignments and transaction costs in industry processes. A fundamental issue is that normal commercial incentives do not operate in many areas of the rail industry. One could characterise the industry's 'profit and loss' (P&L) account as having an artificial disjoint between the passenger revenues (which are the main focus of TOC attention) and infrastructure cost (which is the main focus of attention for Network Rail).

Long-run incentive alignment may therefore mean that each organisation is spending its own money (as opposed to another industry player's) when it invests or incurs a cost, and is earning money directly from its customers when it provides a service or product.

More specifically, the main findings of this work in relation to particular process case studies are as follows.

- There seem to be substantial incentive misalignments in the sector, relating to:
    - the current franchise agreement;
    - incentives on Network Rail to reduce costs;
    - a lack of investment coordination in the sector;
    - stations.
- Transaction costs (defined as direct costs involved in managing interfaces—largely staff costs) appear to represent a relatively small proportion of total costs in the sector, and specifically targeting them for reduction (eg, via a structural alternative) is unlikely to substantially reduce net cost.
    - Many of these costs would also be present in alternative industry structures.
    - This is not to say that transaction costs are at an optimal level—they can and should be reduced where possible (eg, in relation to certain consultations).

2.1 Case studies on current arrangements: incentives and costs

This section provides an overview of findings on the performance of nine key industry process case studies. The case studies identified and agreed with the VfM Review team and in consultation with stakeholders were:

- performance improvement;
- possessions;
- stations;
- the franchising system;
- timetabling;
- rolling stock;
- long-term planning;
- renewals and enhancements;
- capacity allocation.

These process case studies were assessed with respect to their impact on incentives and costs. The assessment revealed a number of areas where incentive misalignments were

especially problematic. These related to the franchising system and stations processes, and two areas that cut across a number of processes (including franchising and stations): incentives on Network Rail for cost efficiency, and on investment coordination.

These findings emerged from a review of a wide body of evidence that is described and analysed in Appendix 2, and which is used as the basis for the summary conclusions on each of the nine processes that are presented below.

2.1.1 Performance improvement

Objectives and incentive regime

Performance improvement is concerned with maintaining and increasing performance levels (delays and cancellations). The main interfaces are between government and the industry in setting output requirements, and between Network Rail and the operators (freight and passenger).

Long-term incentives in this area relate to the Public Performance Measure (PPM) and operate via: Joint Performance Improvement Plans (JPIPs) and their contractualisation (in franchise agreements and the Network Code); Schedule 8 of the Track Access Agreement (TAA); and the regulation of the network by staff 'on the ground' to the PPM.

Outcomes: incentives and transaction costs

Broadly speaking, and based on evidence relating to performance and a range of discussions with stakeholders, performance improvement is an example of an industry success story, with the focus on the PPM delivering considerable benefits over time. However, concerns still remain in particular areas, including: whether current levels of performance should be improved rather than maintained; whether Schedule 8 payments accurately reflect TOC revenue; the change in the scope of Schedule 8 from liquidated damages to claims; and whether existing measures help to end very poor performance.

In addition to the issue of whether existing performance levels should be improved or maintained at today's levels, incentives in this area could be improved through revisiting trade-offs between capacity, journey time and performance. This could be achieved through changing the emphasis of incentives to reflect passenger priorities (eg, a focus for long-distance operators on avoiding cancellations and serious lateness), while maintaining overall performance at appropriate levels and through incentivising operators (eg, via the franchise agreement) to cooperate with Network Rail to identify whole-industry solutions to performance challenges (building, where appropriate, on JPIP initiatives), as opposed to relying on infrastructure changes (eg, FCC's improvements to London–Cambridge services).[6]

A number of transaction costs arise—eg, the complexity of the system adds to resource time cost (the delay attribution guide comprises 90 pages). A large staffing resource is dedicated to this area (approximately 300 to 500 staff at Network Rail and operators for delay attribution). However, interviews and other analysis (eg, assessments of resource costs in this area for other industry configurations) have suggested that many of these staff members would still be required under alternative market structures. For example, in the pre-privatisation structure of British Rail, delay attribution was still a significant activity, with department heads allocated delay minutes. It was also noted that interviewees stated that delay attribution is a useful piece of management information for delivering continuous improvement in performance. According to interviewees, a common regime for all TOCs also

[6] This refers to joint FCC/DfT work that examined short-term means of increasing capacity on chronically overcrowded Cambridge services. Given timescales available for the work, and other constraints, a number of non-infrastructure options were pursued, including timetabling and rolling stock deployment. The experience suggests that alternatives to infrastructure-based solutions can deliver value for passengers, and can help improve incentive alignment; this is not to suggest that infrastructure-based solutions are never appropriate, but that alternatives can deliver desirable outcomes at lower cost and with greater speed, depending on circumstances.

helps to control overall transaction costs, although it is noted that the PEARS system is nearing the end of its life.

2.1.2 Possessions

Objectives and incentive regime

The purpose of the possessions regime is to ensure that operators are compensated for planned restrictions in use of the network, and to minimise disruption to customers. The main interfaces arise between Network Rail, Network Rail's supply chain, and operators.

As part of the CP4 settlement, Network Rail was given targets relating to possession disruption indices (PDIs), which are intended to ensure that Network Rail minimises disruption to its customers resulting from taking possession of the network. Since CP4, Joint Network Availability Plans (JNAPs) have started to be developed between Network Rail and TOCs, the first being put in place with Arriva CrossCountry. A key objective of the JNAPs is to improve collaboration between TOCs and Network Rail in this area, including through the endorsement of the possession plans by operators. The plans also facilitate the monitoring of progress against plan.

Schedule 4 of the TAA sets out how train operators (passengers and freight) are compensated for the cost of disruption caused by planned possessions. Compensation is intended to be broadly reflective of the degree of disruption and the period of notice given to operators (and, therefore, their customers) while striking a balance between accuracy and simplicity to minimise transaction costs.

Network Rail's Schedule 4 cost is funded through an access charges supplement (ACS) paid by TOCs. The ACS and changes to the formula are subject to clause 18.1/Schedule 9A of franchise agreements (no net loss/no net gain).

Outcomes: incentives and transaction costs

Stakeholders have concerns about elements of the possessions process, but on the whole suggest that the outcomes of the possessions regime are—eventually—acceptable, even if the process for reaching the outcome is itself laborious.

Interviews with stakeholders, and the review of this process, described in Appendix 2, have also indicated specific areas of concern, including:

– the time it takes to negotiate short-term variations to the timetable to enable possessions to be taken;
– the incentives on Network Rail to optimise the duration of each possession given the difficulties faced by external parties in establishing how long (and at what cost) Network Rail should take to efficiently undertake each possession;
– the manner in which Schedule 4 interacts with increasing the times each day when trains are running (ie, does Schedule 4 compensate only with regard to the existing timetable, or does it not disincentivise the amount of time in which trains could run);
– how Network Rail plans its possessions optimally across track renewal, maintenance and large capital projects—Network Rail now has resources devoted to making this happen.

The fundamental issue is whether it is possible in the current arrangements both to maximise network availability, and possessions productivity (work done per possession) and efficiency (unit rates). In practice, the changes that might give effect to this could include the following.

– Where possible, making the timetabling impact of certain possessions more predictable, thus avoiding lengthy negotiations.

– Strong incentives on Network Rail to optimise possession plans (eg, one approach would require Network Rail to 'rent out' lines it is having to close, much like utility companies may have to pay lane rental).

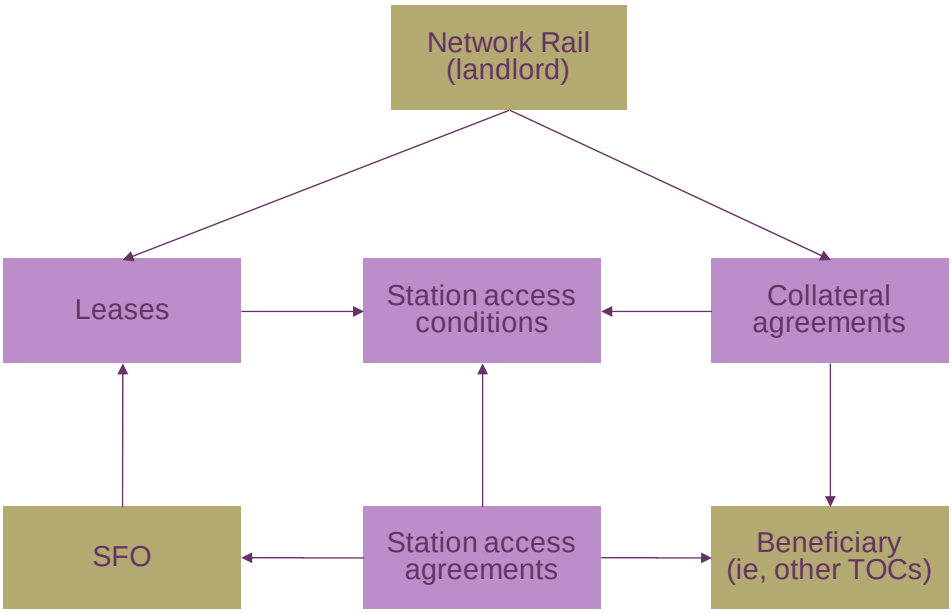
– Incentivising operators (eg, via the franchise agreement) to cooperate with Network Rail in identifying ways of jointly maximising availability, and possessions productivity and efficiency.

2.1.3 Stations

Objectives and incentive regime

The objectives of the stations contractual matrix are to provide for an efficient level of maintenance for station assets, to enable third-party access to stations, and to impart price signals and contractual protection for all concerned parties. Some of the contractual interfaces between parties are illustrated in Figure 2.1.

Figure 2.1 Stations interfaces



Note: SFO is station facility owner.
Source: Based on ORR, see http://www.rail-reg.gov.uk/server/show/nav.224.

Network Rail is the landlord of almost all stations and is funded (by the DfT or the relevant government authority) to repair, maintain and renew the relevant parts of stations.[7] Network Rail must also carry out specified enhancements as part of its Control Period settlement and as part of other DfT schemes, such as Access for All. In addition to owning and maintaining most stations, Network Rail is responsible for managing several of the larger stations.[8] The day-to-day operation of stations is usually leased out via station leases to TOCs. The lessee TOC then becomes the Station Facility Owner (SFO), with a licence to operate the station and is normally responsible for light maintenance.

Outcomes: Incentives and transaction costs

In practice, the existing arrangements are fraught with complexity, especially in relation to responsibilities. This complexity was communicated in interviews with stakeholders, and

[7] A small number of stations are owned by third parties, such as St Pancras, Prestwick International and Warwick Parkway.
[8] Under the Station Licence granted to Network Rail Infrastructure Ltd, Schedule 2, Network Rail may manage Glasgow Central, Edinburgh Waverley, Leeds, Manchester Piccadilly, King's Cross, London Bridge, Birmingham New Street, Waterloo, Paddington, Liverpool Street, Charing Cross, Victoria, Euston, Liverpool Lime Street, Canon Street, Gatwick Airport and Fenchurch Street stations.

emerged from the review of the system of contracts, change procedures, and financial flows described in Appendix 2.

The consequences of the complexity of the contractual matrix in which stations operate can lead to suboptimal investment by Network Rail (given incentives around dilapidation expenditure), and suboptimal asset management decisions (since improvements may not be taken forward, and life-expired assets are maintained to avoid removal procedures) and poor asset knowledge. Split responsibilities for station development (design, development, regulation and funding) can lead to delays, avoidable cost increases, and potentially missed opportunities. These missed opportunities include demand-side impacts, in which revenue could be grown (to reduce net cost) by improving retail offerings and potentially working with local authorities to support demand management/revenue generation through improved multi-modal opportunities and integration with the local area.

The overall situation could be improved through regulatory interventions that require one party to have responsibility for management and repair, subject to non-discrimination protections, and it is noted that the DfT has been considering allocating TOCs a full repairing lease for the majority of stations (with the first instance expected to be part of the Greater Anglia franchise specification), an approach which has also been advocated by the Association of Train Operating Companies (ATOC). Additionally, and particularly for stations, where considerable engineering work is required for structural upkeep, an increase in Network Rail's portfolio of managed stations might be considered, and perhaps be combined with some refocusing of Network Rail's commercial incentives (and possibly recreating a separate 'Major Stations' organisation).

One approach would be for Network Rail to be responsible for maintaining, repairing and renewing the building fabric, and TOCs for maintaining, repairing and renewing passenger-facing facilities.

2.1.4 Franchising system

Objectives and incentive regime

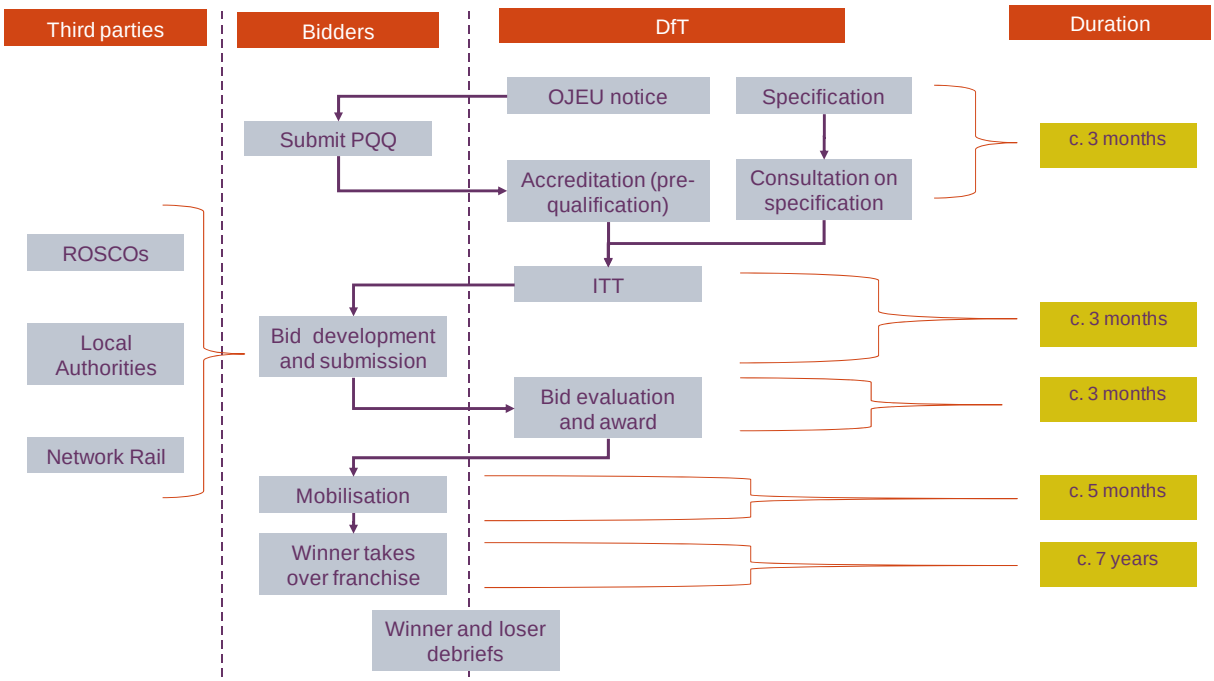
The franchising system is intended to allow for cost efficiency by securing competition for private sector delivery of passenger train operations, within a framework that is intended to allow for socially desirable requirements to be specified and delivered by winning bidders.

Franchise agreements have changed slightly with the different versions of franchising policy (eg, those regimes administered by the OPRAF, the SRA, and the DfT). The current franchise agreement details the parameters of the franchise, ranging from the commencement date and duration of the franchise, to the specific obligations of the franchisee and the DfT. It covers requirements such as the service level commitment and the concessionary schemes that the franchisee is obliged to offer.

In addition to the main document, there are often multiple appendices covering risk assumptions, target revenue, priced options submitted in the franchise bid, service quality benchmarks and provisions for expected major changes during the franchise, such as the 2012 Olympic and Paralympic Games, or Thameslink upgrade. Following the award of a franchise through the tendering process, the TOC will enter into a franchise agreement with the relevant authority.

A schematic representation of the interfaces involved in franchise tendering is provided in Figure 2.2.

Figure 2.2 Example of tendering interfaces



Source: Oxera/Arup.

Franchise agreements reflect government minimum train specifications, fares and ticketing policy, quality of service, and any required enhancements. The agreement is monitored and enforced by the relevant authority.

Outcomes: incentives and transaction costs

The government has recognised some of the problems with franchising, and has stated that through an 'appropriate and a better aligned balance of incentives and risks, private operators are more likely to commit their own resources to investment in our railways and deliver greater efficiencies'.[9] ATOC has also identified that significant cost savings (several hundred million pounds per annum) are available from reforms to the franchising system.

These findings and suggestions as to the possible scale of cost savings connected with reforms are reflected in the findings of this study, which indicates that the franchise agreement is at the centre of a number of significant incentive misalignments, including the following.

- It is currently designed to transfer risk, especially infrastructure-related risk, away from operators (keeping bid margins lower);
- It delivers a business model involving high fixed costs, limiting TOC flexibility to respond to (local) market conditions;
  - this magnifies revenue risk linked to the macroeconomy.
- The 'cap and collar' system can lead to gaming behaviour in bidding;[10]

[9] Department for Transport (2010), 'Reforming Rail Franchising', p. 6.
[10] This is because the system gives an incentive to push up premia in the latter years of the franchise in order to increase NPV, and leads to 'safer' investments being concentrated in the early years in order to ensure that the TOC gets the benefit of these investments.

– The franchising system disincentivises cooperation with Network Rail to improve whole-industry outputs, and to reduce industry whole-life costs As a consequence, the ORR may have to intervene more frequently between Network Rail and TOCs than is desirable;

– The system encourages TOCs to manage to the contract, as opposed to focusing on desirable outputs and on the passenger.

The interaction of the system of fares with demand management is also seen as problematic. For example, regulation of off-peak fares ('Savers') on longer-distance flows can, effectively, put a cap on off-peak advance purchase fares, and prevents the TOC from increasing prices on more attractive off-peak trains. Relatedly, regulated restrictions on Savers mean that TOCs cannot price up to demand on Sunday afternoons, which in some cases is a time of peak intra-week demand. Currently TOCs have limited incentives to manage demand as a means of removing the need to strengthen trains. This was offered as a central option in 2006/07 franchise bid rounds, but the 'priced option' alternative of remaining with the status quo was accepted by the DfT.

In addition, 'one-size-fits-all' fares regulation does not necessarily fit with the different markets served by train operators. The recent Oxera–Arup study for the DfT[11] suggested that fare elasticities are higher than previously thought, and that the elasticities of fares basket products are now closer together, limiting the scope for demand-reflective pricing (such as 'Ramsey pricing') in the fares basket (although other forms of basket management remain possible).

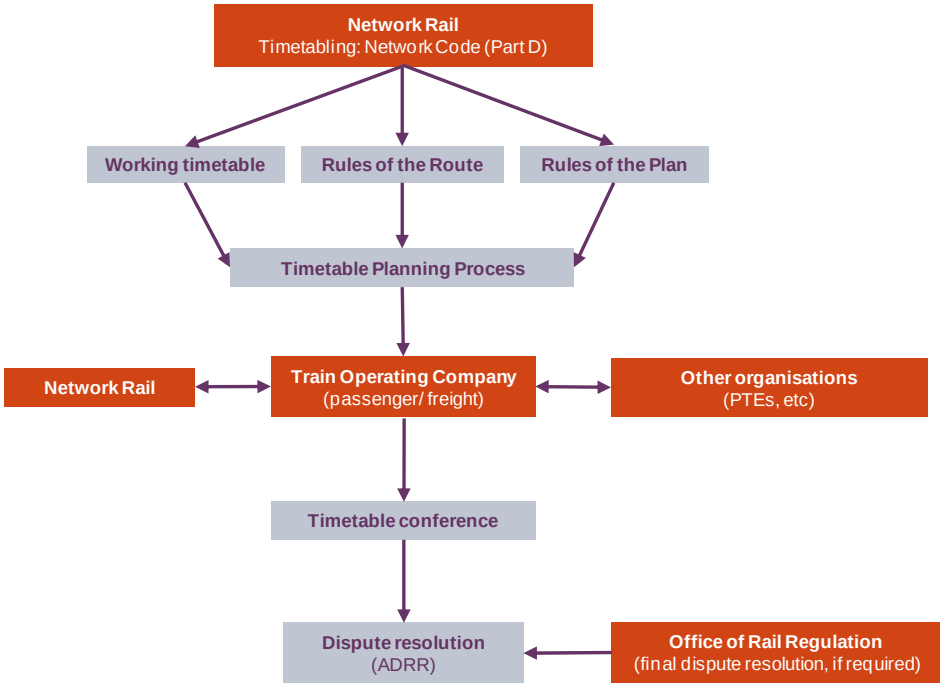
2.1.5 Timetabling

Objectives and incentive regime

The objective of the timetabling process is to allocate train paths. The bid/offer process enables on-rail competition, and allows for dialogue between Network Rail, the TOCs, and ORR. The principal interfaces exist between Network Rail and the operators, and, especially at franchise specification, between the DfT and Network Rail. The main interfaces are summarised in the figure below.

[11] As yet unpublished.

Figure 2.3 Timetabling interfaces



Source: Oxera/Arup.

Outcomes: incentives and transaction costs

A key issue with timetabling is that its iterative nature[12] means that timetabling staff at both Network Rail and the operators arguably spend longer working on the process than would be the case if fewer interfaces existed. Also, stakeholders have suggested that the extent of manual intervention could be reduced, and greater reliance placed on IT-based approaches to timetabling.

There is also a degree of misalignment of incentives: Network Rail's incentive to produce an operationally robust timetable may cut across operators' commercial timetabling aspirations, while the timetable requirements in the franchise agreement limit the ability of TOCs to respond to market changes in their timetabled services.

The high number of interfaces, guidelines, consultations, and dispute resolution processes make the system of timetabling very time-consuming. The process absorbs a high number of staff resources.

2.1.6 Rolling stock

Objectives and incentive regime

Arrangements relating to rolling stock are intended to provide safe rolling stock that makes best use of existing capacity. Key interfaces arise between Network Rail and operators, between ROSCOs and operators, between ROSCOs and vehicle manufacturers, and between vehicle maintainers and operators.

Outcomes: incentives and transaction costs

A number of concerns were raised by stakeholders, and emerged from the review of this process case study.

– Potential duplication in the process between compatibility (safety and interoperability) and vehicle change (commercial impact) procedures.

[12] Described in detail in Appendix 2.

– Incentives on franchised TOCs to obtain the best deal on rolling stock seem to be limited by a lack of bargaining power at franchise replacement, and during the franchise, the DfT's 'hard-wiring' of rolling stock fleets within franchise contracts contributes to this problem.

– The model in the GB freight sector and the models used in continental Europe, seem to offer alternatives that are worthy of further examination.

Incentives could, conceivably, be improved by:

– involving Network Rail fully in the design process, regardless of who procures rolling stock;
– incentivising TOCs to reflect market conditions in their procurement and use of rolling stock throughout the franchise—this would involve a move away from centrally planned procurement and cascades, except in clear cases of market failure.

2.1.7 Long-term planning

Objectives and incentive regime

The objectives of long-term planning processes in the industry are self-explanatory: to provide long-term information on the industry and market environment, and to develop and implement potential means of addressing changes. A key aspect of long-term planning is the role played by the government in establishing the outputs it is willing to buy from the industry through the High-level Output Specification (HLOS) and Statement of Funds Available (SoFA) process, and the ORR review of the costs to be incurred by Network Rail in fulfilling this demand. All industry stakeholders have some interface with the processes of long-term planning.

Outcomes: incentives and transaction costs

Perhaps inevitably, long-term planning involves a large number of stakeholders and this increases interfaces. There also appears to be some degree of duplication of consultation between Network Rail and the ORR.

More generally, there is a degree of misalignment between the network planning framework (based on WebTag) and the statutory duties of the ORR around capacity allocation. A further risk of misalignment arises because delivery of some aspects of the HLOS is the joint responsibility of both Network Rail and the TOCs.

The main issues with long-term planning are being covered in more detail in the SDG-CEPA study on leadership, planning and decision-making.

2.1.8 Renewals and enhancements

Objectives and incentive regime

The basic objectives of renewals and enhancement activities are to meet the requirements of the HLOS, other legislative and technical requirements, and more generally, to meet user needs. The interfaces therefore cut across almost all stakeholders in the industry. Legislation, the Network Code, and bilateral contractual agreements all play a role in defining how and when renewals and enhancement activity is undertaken.

Outcomes: incentives and transaction costs

There appear to be limited incentives for wider industry input and collaboration in this area, resulting from:

– the mismatch of enhancement benefits and enhancement costs over time, and between parties/interfaces (eg, costs may be incurred upfront by Network Rail, the benefits of which may be realised by train operators a number of months or years hence);

– the indirect way in which specific project capital expenditure (CAPEX) is charged to operators over time via the 'averaged' track access charges;
– the lack of financial benefits arising in connection with schemes that enhance wider societal outcomes, and the fact that users generally do not pay directly for enhancements;
– more generally, the difficulties of investment coordination across these interfaces.

There are also concerns that network change procedures are a barrier to reducing net cost. This has been a particular issue since the ORR's decision on the appeal to NV33,[13] which considerably widened the scope of network change (this also has implications for current performance arrangements, as mentioned above). This is some distance from the original objective of 'Network Change' procedures, which related to protecting operators from significant and permanent/semi-permanent decrements to the infrastructure, harming their commercial interests.

There is also a perception in some cases that strategic positions have changed during the planning of cross-industry projects (eg, IEP; see Foster (2010), 'A review of the Intercity Express Programme', June).

Recent experience from large cross-industry projects, such as Global System for Mobile Communications – Railway (GSM-R), has identified enablers for the delivery of complex projects:

– parties interact according to clearly defined roles and responsibilities—there are fewer conflicts;
– parties actively collaborate via ideas and actions to improve project (industry) net present value (NPV) outcomes;
– the process is underpinned by a rational flow of funds and risk allocation;
– the development process is quicker, more transparent, and evaluates the best options. Inter-generational equity is preserved by optimising between short-term disruption and long-term benefits;
– delivery is more efficient.

More generally, incentives might be improved by:

– full operator engagement with the delivery plans associated with the Periodic Review process (eg, through disapplication of Clause 18.1/Schedule 9A provisions for all changes but those to fixed access charges);
– incentivising operators (eg, via the franchise agreement) to cooperate with Network Rail in identifying ways of maximising industry benefits and minimising industry whole-life costs from enhancements;
– TOCs having a role in leading the specification of enhancements under longer franchises.

2.1.9 Capacity allocation

Objectives and incentive regime

Capacity allocation processes in the industry attempt to secure efficient use of network capacity, optimising it from an operational perspective and from the perspective of customers. The main interfaces arise in the form of track access contracts, covering access rights to the network held by the operator. Network Rail's Railway Operational Code (ROC) sets out the operation of train services on the network in accordance with the working timetable.

[13] RR 2 2003 Case No: 2003/02 On Appeal from The Network and Vehicle Change Sub-Committee of the Access Dispute Resolution Committee.

For capacity allocation, a process exists whereby all parties seeking to secure track access paths have to apply through Network Rail to gain the required timetable paths. This process can be iterative, with several permutations of capacity allocations being developed before a final timetable is agreed.

Outcomes: incentives and transaction costs

Network Rail is incentivised through Section H of the Network Code to regulate train services on the network in accordance with the working timetable and to restore operations following incidents—if it fails to do this, it will miss specified performance targets and will incur financial penalties. Likewise, TOCs are incentivised through their commercial objectives to seek optimal capacity allocation and to maximise their day-to-day performance so as not to incur financial penalties.

However, a key source of incentive misalignment is that Network Rail is incentivised to allocate capacity optimally on an operational basis, whereas TOCs seek to optimise on a commercial basis. There is also no guarantee that the optimal balance between performance and capacity utilisation will be achieved.

There are significant time and resource costs associated with capacity allocation, although elements of this overall cost have already been described above in relation to timetabling.

2.2 Transaction costs

Transaction costs associated with industry processes have been touched on above. A general conclusion to emerge from this review was that, while transaction costs are significant in absolute cash terms, they do not appear to contribute as much in their own right to net cost and general industry outcomes as do incentives. A number of sources of evidence supported this conclusion.

The first source was interviews with stakeholders. These interviews were generally discussed in the context of the work on the mapping of interfaces, and the conclusions on interfaces and transaction costs that had emerged.

In general, although not in all cases, there was agreement that transaction costs were not in themselves the core driver of industry net cost. A number of stakeholders told us that staff costs constituted the largest proportion of overall transaction costs, and expected most, if not all, staff to be retained in similar functions across these processes under alternative industry structures. For example, Network Rail employs hundreds of people in its train planning function and many of these staff members would still be required under alternative industry structures, even if employed by a different organisation.

Another source of evidence was ex post analysis of restructuring in other sectors, as well as experiences in international rail markets. Ex post studies of the experience of separation in utility industries, such as electricity and gas (and proposed future restructurings in the England and Wales water sector), have generally found significant benefits from restructuring, in spite of the accompanying increase in transaction costs; more detail on the experience of other sectors is presented in Appendix 3.

These studies imply that, if overall market design and the associated set of incentives are appropriate, then end-users enjoy net benefits relative to the relevant counterfactual (which might typically be an integrated industry structure with lower levels of competition), and in particular that *industry* costs are lower overall, despite the increase in *transaction* costs.

One especially relevant study in this context is that of Merkert (2009).[14] Merkert found that transaction costs in the GB rail sector increased after Hatfield to a higher level than before privatisation, and had continued to increase up to that point (transaction costs here are defined as management and administrative staff). However, even at their highest level, they have remained less than 5% of total TOC operating costs. This figure of 5% (which is itself at the upper bound of recent experience) indicates that, while transaction costs are not insignificant, a reduction relative to most plausible counterfactual industry structures would secure, at best, modest savings in net cost.

None of these positions supports, or is intended to support, a conclusion that the overall level of transaction costs in the industry as it now stands is 'optimal'; there should be continued effort by stakeholders to secure reduction in these costs wherever possible, but this is likely to be within the context of wider cost-reduction initiatives. For example, Network Rail is continuing to review the number of staff involved in delay attribution, and has been successful in reducing this number in response to the CP4 efficiency challenge. Network Rail's 'Transformation Programme' can be seen as a means of achieving the CP4 outputs in a sustainable way.

Another perspective is provided by the 2008 BSL study[15] of Network Rail's renewal efficiency gap, which found that less than 10% of the gap with peers was attributable to 'transaction costs' (£70m against a gap of £846m). A quote in this study from contractors with international exposure identified a perception that transaction costs were high in this area, although one could potentially identify this quote as arising from a lack of incentives on different parts of the GB rail industry to control costs: 'no compliance with plan, last-minute changes with major cost implications, freezing of plans and deadlines (eg, in Switzerland) could provide substantial gains'. Interviews with Network Rail suggested that the company will place a greater reliance on 'freezing' of plans in response to the CP4 settlement, which indicates both movement toward best practice in this area, and a need to control the transaction costs that exist between Network Rail and its suppliers.

It is also worth noting that these findings on transaction costs largely reflect the direct cost associated with transactions. In practice, there may be some opportunity costs associated with these transactions. For example, staff may spend time managing procedural or other aspects of transactions, instead of spending time on more value-added activities that improve the operation of the rail network. This aspect of transaction costs is inherently more difficult to evaluate than direct costs.

However, a number of considerations suggest that a more in-depth evaluation of opportunity costs would not significantly alter the general conclusions described above. The first is the issue of whether opportunity costs would also arise under counterfactual restructuring scenarios, and whether opportunity costs associated with increased transactions have increased in sectors that have moved from integrated to separate industry models.

On the former issue, there are no apparent grounds for regarding the character of opportunity costs that arise under the current industry structure as being fundamentally different to those arising under counterfactuals. Similarly, studies of other sectors that measure restructuring benefits are also likely to capture the opportunity cost effect (whether directly or indirectly) and yet still identify net benefits. Nevertheless, the existence of opportunity costs associated with the existing set of transactions and interfaces could merit further study, including in relation to the interface costs that may exist between government and different industry players (eg, TOCs, ROSCOs).

[14] Merkert, R (2009), 'Changes in transaction costs over time – the case of franchised train operating companies in Britain', Working paper, Cranfield University.
[15] BSL (2008), 'Rail Infrastructure Cost Benchmarking', April, p. 36.

2.3 Conclusions and key areas for change

Research for this project into nine key industry process case studies, based on interviews and reviews of a variety of evidence, has found that many incentives across those case studies considered are misaligned. This misalignment inhibits investment coordination, often greatly reduces focus on the end-user, and is likely to lead to poor decision-making.

There does appear to be a strong case for reducing net cost in the sector in the medium term (ie, up to the end of the CP5 control period) through incentive realignment—be it via alterations to existing arrangements or structural change. One theme that has emerged in several areas is that better alignment to passenger markets would improve incentive alignment within the wider industry.

The misalignments that appear to have particularly serious consequences are associated with:

– the current franchise agreement;
– cost control by Network Rail;
– lack of investment coordination in the sector;
– stations.

The **franchise agreement** is central to a number of incentive misalignments. The system does not encourage a focus on customers, or on cooperation with Network Rail to improve whole-industry outputs or industry whole-life costs.

There is some evidence that the PR08 settlement is succeeding in driving change and reducing cost at Network Rail, but **Network Rail's overall cost efficiency** is an important component of the whole industry's cost. Moreover, it is perhaps not desirable that the regulator, instead of the industry as a whole, is driving this efficiency performance. The review of industry process case studies reveals that Network Rail has limited incentives in certain areas:

– trading off capacity utilisation against performance (incentives around performance are much stronger than those around improving frequency, or journey time reductions);

– searching out timetable enhancements (journey time reductions, frequency increases);

  – more generally, Network Rail has limited volume incentives;[16]

– balancing local needs (eg, in relation to asset management) against centrally determined standards;

– its corporate governance structure means limited pressure to respond to financial incentives, with implications for the effectiveness of economic regulation in reducing cost.

There has also been limited opportunity to seek benchmarking information from alternative approaches to maintaining, renewing and enhancing the infrastructure.

Investment coordination across the industry is another problematic area:

– enhancements and cross-industry projects are often hampered by incentive misalignments;

[16] The structure of access charges does not provide incentives for Network Rail to grow volumes; the specific volume incentive it faces only pays out (in the form of a RAB addition) when traffic is above target.

  - it is not clear that best value is being obtained from HLOS projects;
  - some TOCs currently have limited incentive to input into projects to ensure that they are least-cost, highest-benefit;
- cross-industry engagement in cross-industry projects tends to occur too late in the process;
- in the case of new rolling stock, there seems to be limited incentive to achieve the right balance between operator and Network Rail design preferences;
- Network Rail often finds itself at the mercy of statutory consultees, leading to project delays and added cost.

Stations reflect many of the more general problems in the industry. The existing arrangements are complicated, time-consuming and appear unsatisfactory to all stakeholders. Split responsibilities for station development appear to be a particular source of inefficiency.

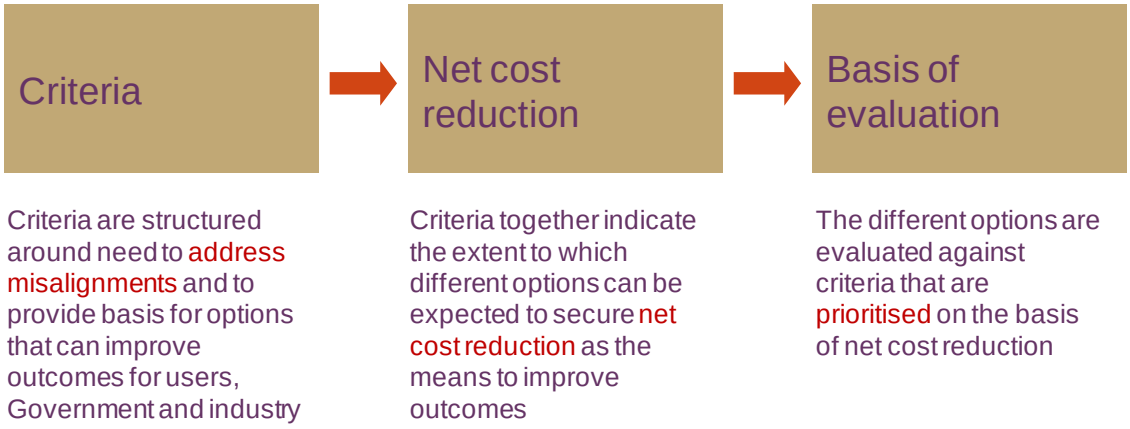
This section began by noting the disjoint between revenue and cost in the industry's P&L, and the need to ensure that revenues are earned from providing services to customers, rather than from fulfilling the terms of a contract, franchise agreement or regulatory settlement. The next section describes how options were generated to address these and other misalignments.

3 Generation of options and of evaluation criteria

3.1 Introduction

This section describes how options to reduce industry net cost were generated, and how they were evaluated against criteria that relate to this objective. The high-level approach undertaken in this section is summarised in Figure 3.1.

Figure 3.1 Toward option evaluation



Source: Oxera/Arup.

The remainder of this section is structured as follows:

– section 3.2 describes how options were generated to address the issues identified as part of the review of industry process case studies
– section 3.3 describes the range of options considered.

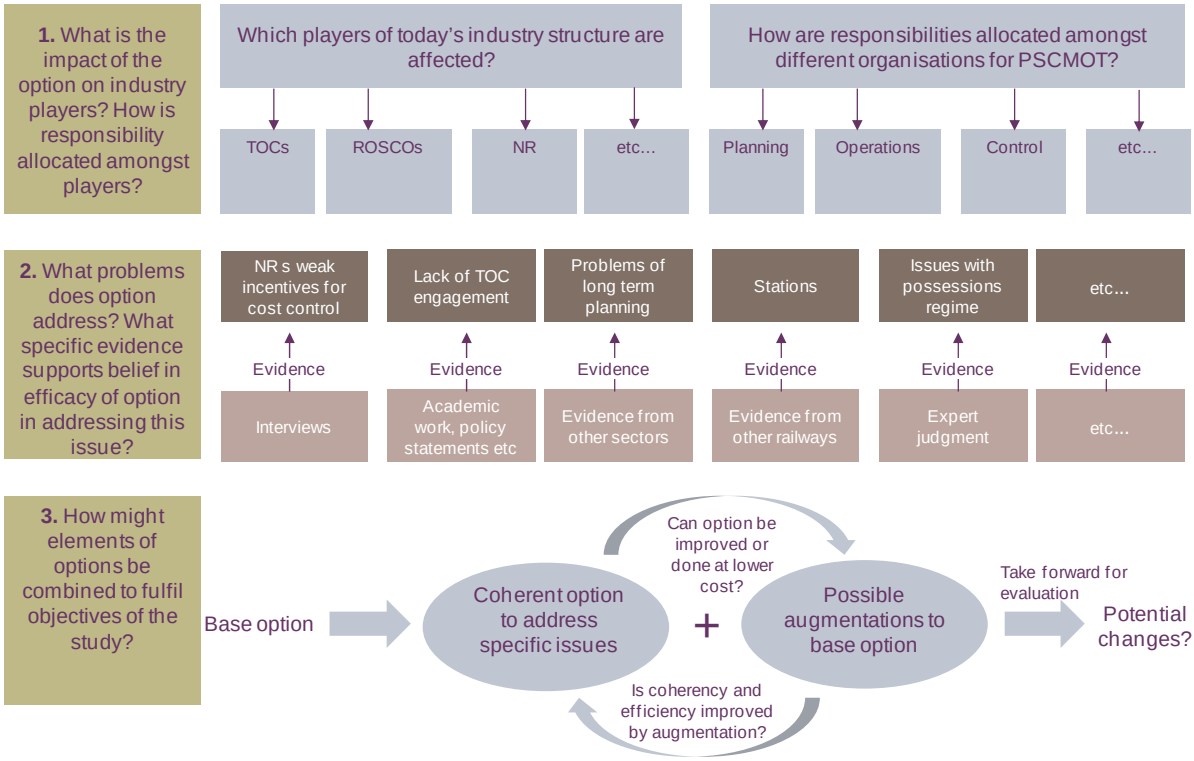
3.2 Option generation

A number of options were generated in order to identify how rail industry interfaces, incentives and structure can be improved to reduce industry net cost. This process is summarised in Figure 3.2.

The first stage in option description was to identify the impact it would have on different industry players, and how responsibilities for different aspects of the railway were distributed among these organisations. The second stage described the specific problems that the option would address by the changes proposed in the first stage, and used the available evidence to test hypotheses around potential efficacy.[17]

[17] This analysis did not explicitly examine safety issues. It is assumed that any option will not be taken forward in a manner that is detrimental to safety.

Figure 3.2 Process of option generation



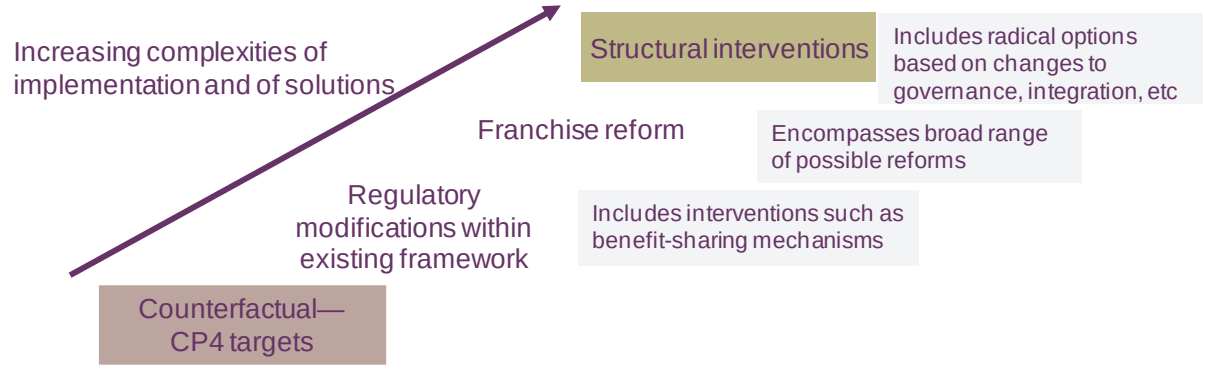
Note: PSCMOT: plan, specify, change, manage, operate, transform.
Source: Oxera/Arup.

The possibilities arising from combining elements of options are explored to examine how different approaches could work together to reduce industry net cost. Testing all possible combinations of options is clearly infeasible. Instead, the approach taken has been to consider how different possible options for change could be combined to take advantage of their different strengths. For example, options around franchise reform could exert a strong influence on the operations side of the industry, but could, potentially, be strengthened with complementary changes on infrastructure.

It is possible to characterise the options that emerged under this process according to a spectrum defined by the timescales for, and complexity of, implementation. The options ranged from incremental changes that existed within the regulatory envelope defined by the CP4 settlement, to significant structural changes that would fundamentally alter the landscape of the industry.

Figure 3.3 provides a high-level illustration of this spectrum of options.

Figure 3.3 Spectrum of regulatory interventions



Source: Oxera/Arup.

Figure 3.3 is useful insofar as it highlights the need to consider a range of options: radical structural changes may not necessarily deliver the best overall industry net cost reduction. Instead, options should be evaluated on their own merits, without a presumption that more complex or radical changes are always superior to more incremental approaches. Similarly, more radical options that appear to offer a credible basis for securing reductions in industry net cost should not be neglected in the context of the 'once-in-a-generation' opportunities presented by the VfM Review.

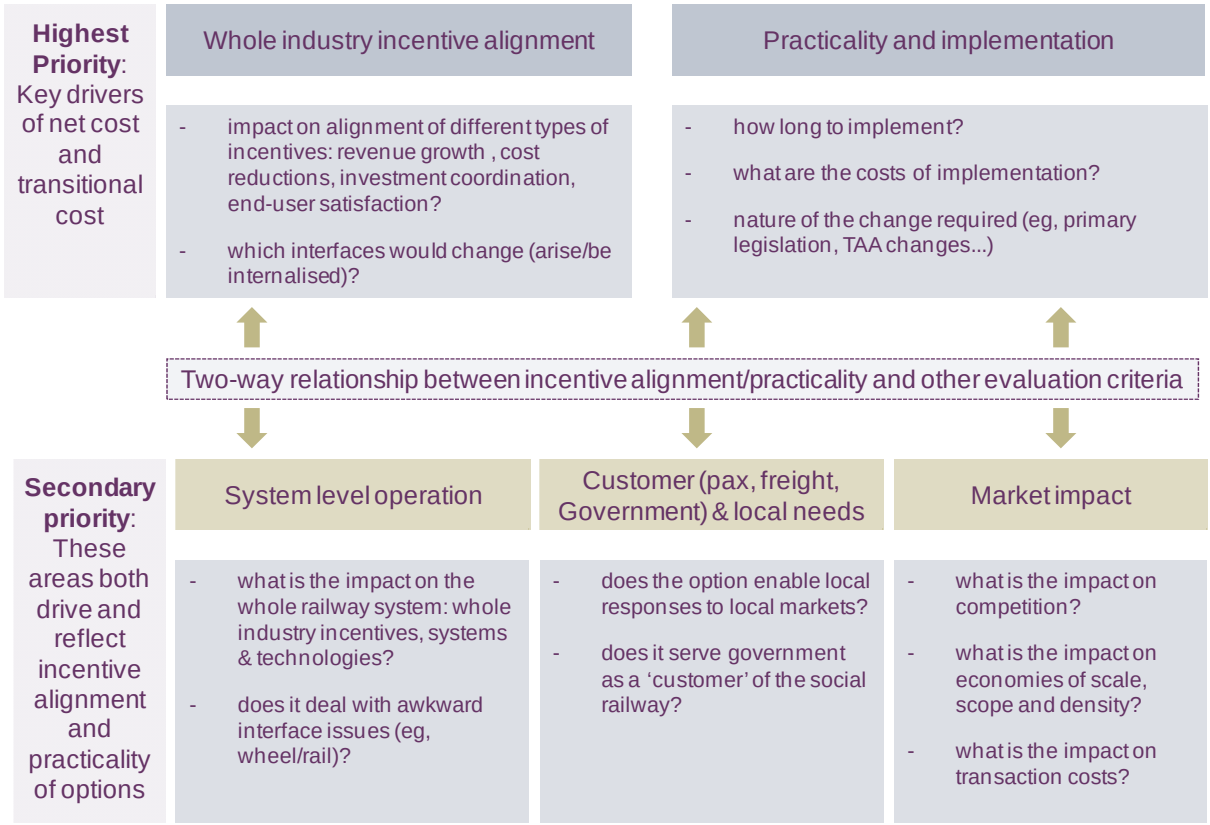
3.3 Evaluation criteria

The options were evaluated against a list of criteria developed and agreed with the VfM Review team. Five categories of criteria were selected:

– whole-industry incentive alignment;
– practicality and implementation;
– system-level operation;
– customer and local needs;
– market impact.

The specific issues addressed under each of these areas, and the priority attached to them, is summarised in Figure 3.4. This list of criteria was developed as a means of eliciting how individual options might be expected to perform in addressing the drivers of industry net cost. They were chosen so that the focus was not purely on 'incentives' or the practical costs of change, but on the wider impacts on the railway system, users, and markets. Put differently, they attempt to set out where final outcomes (prices and quantities of output) might be affected by 'intermediate' mechanisms, such as incentive alignment, practicality considerations, market impact, etc.

Figure 3.4 Evaluation criteria



Source: Oxera/Arup.

The priority of the different options is defined by the scope of this study. The highest priority is attached to whole-industry incentive alignment and the practicality and implementation issues associated with different options, since these two criteria together define the expected success of any option in reducing industry net cost through securing better incentive alignment, and the costs of making the transition to that new set of incentives. Put differently, incentive alignment is seen as a core driver of outcomes, and practicality as a core driver of implementation costs.

For these reasons, incentive alignment is particularly emphasised in the discussion summarising the review of the nine industry process case studies in section 2 and in Appendix 2—ie, as a means of identifying where change would be particularly valuable in securing improvements. Practicality and implementation is obviously an aspect of options for change rather than existing industry processes per se.

The areas of secondary priority are both drivers and consequences of the higher priority criteria. New sets of incentives cannot be imposed on the industry without regard to their consequences for the operation of the railway as a system, the impact on passengers, and the general market impact with respect to competition and cost. It is noted that there is a two-way relationship between the alignment and practicality criteria on the one hand, and between the other criteria on the other. Therefore, in addition to serving as a type of 'check' on the options, these other criteria affect, and are affected by, the incentive alignment and implementation costs of different options.

However, this does not necessarily imply that the scale or direction of impact will correspond precisely between primary and secondary criteria, since it is possible in some circumstances that an option for change that has a significant impact in one area would have a limited impact in another. For example, a franchise reform option (as discussed in section 4) could

have a significant impact on incentive alignments in relation to revenue growth, cost reduction and end-user satisfaction, but a limited impact on competition (since a franchising system is retained), on economies of scale, scope and density (assuming no changes in franchise size or on Network Rail), or on transaction costs.

This also demonstrates that individual criteria will clearly have a scale and direction of impact that is specific to the option being considered. The evaluation addresses this by considering each impact separately. It is important to note that the evidence on these individual impacts is not readily available in many cases, and that many of the criteria are qualitative in nature. This poses challenges for combining options into a single, overall quantitative measure of impact.

These two features of the option evaluation—the lack of available evidence in many cases and the qualitative nature of some of the relevant criteria—mean that the conclusions presented in section 4 should be read and interpreted in a manner that reflects these qualifications to the analysis.

Impacts are classified as being low (£10m–£100m), medium (£100m–£500m), and high (£500m+).

4 Option description and evaluation

This section describes and evaluates the 'base' list of options. The approach taken has been to define a number of internally consistent, coherent options that are expected, on the basis of available evidence, to address incentive misalignment and transaction cost issues, and thereby to reduce industry net cost. Where appropriate, and following the initial evaluation, elements of these options were combined and subsequently re-evaluated.

The list of base options evaluated is as follows.

- Incremental changes within the existing framework:
  - option '1a': changes that are likely to be feasible within the framework of the CP4 settlement;
  - option '1b': radical franchise reform.
- Horizontal separation.
- Vertical integration (VI):
  - VI at market level.
- Greater public sector involvement:
  - 'Railways Agency' model;
  - nationalisation.

The following describes these options in terms of intended objective, misalignments addressed, impact on players, likely transition processes and evaluation performance.

4.1 Option 1a—changes within the existing CP4 regulatory framework

4.1.1 Objectives

The objectives of this option are summarised in Box 4.1.

Box 4.1 Key objectives

Deliver change without the need for change in legislation, radical adjustments to franchise agreements, or an interim access charges review.

Strengthen and improve the alignment of incentives on Network Rail and (primarily franchised) operators.

4.1.2 Description

This option focuses on making limited, incremental changes to the status quo within the existing regulatory framework and industry structure. The intention is that the changes envisaged would be focused on 'activating' elements of the CP4 determination, and introducing successful elements of the existing structure into other processes—specifically, the following.

- Introducing risk capital into Network Rail, either along the lines of the CP4 determination (issuing debt outside the Financial Indemnity Mechanism, FIM), or via the introduction of quasi-equity.

– Activating JNAPs for possessions on a similar basis to JPIPs, and covering all operators (including freight and open access).

– Removing no-net-loss, no-net-gain (Clause 18.1/Schedule 9A) arrangements surrounding the efficiency benefit share procedures introduced for CP4.[18] These arrangements nullify any financial benefit that franchise operators receive from Network Rail outperforming the CP4 settlement, removing the incentive otherwise in place for them to work with Network Rail to enable it to outperform. By making local data available, extending the concept to a route level, together with exposure of TOCs to downside risk, would strengthen the incentives associated with the mechanism

– Disapplication or relaxation of fares regulation for operators that are able to present clear evidence that ex ante regulation for some products in the fares basket is no longer necessary as sufficient competition for such products exists.[19] The ORR would regulate fares on an 'economic' basis, evaluating the case for regulation/deregulation using standard market power tests (and other analytical techniques as appropriate). Overall fares policy would remain with government, but the ORR would decide which products would be subject to economic regulation. The use of an economic, rather than an administrative basis for fares regulation should lead to more efficient allocation of resources in the sector.

4.1.3 Which incentive misalignments would this option assist with?

– The option is designed to strengthen incentives on Network Rail to outperform the periodic review settlements through the use of a strengthened efficiency benefit share mechanism. If Network Rail were able to introduce risk capital, then this effect, particularly with respect to cost efficiency, would be strengthened further.

– The JNAP mechanism, and other aspects of the option, would bridge gaps between operators and Network Rail, imposing clear incentives on Network Rail to improve network availability.

– The disapplication of fares regulation for products that no longer justify protection would enable TOCs to manage demand more effectively in areas where pricing is currently not an option in this context. There would be a need to ensure that the incentives on operators to undertake this are not limited by lengthy franchise change procedures being triggered. There would also be a need to ensure that financial benefits to operators are not clawed back in full by the DfT, as this would remove the incentive for operators to request the removal of regulation.

4.1.4 Impact on players

Network Rail/infrastructure managers (IMs)

– The precise impact depends on the nature of risk capital introduced. However, introducing debt outside the government guarantee (and within this category, junior debt)[20] into Network Rail's capital structure is likely to have a limited impact unless strong covenants are attached.[21] Our interviews have suggested that current arrangements (Network Rail's debt-only financing and the government guarantee)

[18] It is recognised that exposing TOCs to infrastructure risk may be a significant 'ask' in the context of an option intended to be incremental in nature, and it may be more appropriate to expose TOCs only to a proportion of infrastructure cost risks, such as variable track access charges or a proportion of OMR.
[19] This encompasses all fares subject to some form of regulation.
[20] 'Junior' debt would rank beneath 'senior' debt in terms of payment priority, and the bonds can be designed such that defaulting (missing repayments) on junior debt does not trigger default on the remaining debt items. In this context, junior debt can be seen as 'quasi-equity', in that junior debtholders, like shareholders (in the case of dividend payments), bear the risk that their repayments might cease or be reduced for a certain period of time.
[21] This conclusion reflects work undertaken for the ORR as part of the PR08 review process—see NERA (2006), 'Corporate Form, Financial Guarantees, and Efficiency Performance: Expectations and Evidence', December.

heighten external monitoring requirements on the company, which can involve considerable opportunity costs.

– Contractualising JNAPs in a similar way to JPIPs would provide a new set of customer reasonable requirements for Network Rail (under the Network Code) and enlist operator input in improving availability. They would provide additional focus on network availability to each operator, while still needing to deliver Network Rail's CP4 outputs (via network interventions). One question is whether existing PDI targets by operator (freight and passenger) consistent with the CP4 determination are sufficiently demanding (especially zero improvement/deterioration with respect to freight) to be used in contractualised JNAPs.

– Disaggregating efficiency benefit sharing (EBS) arrangements by route is also likely to strengthen incentives for outperformance, and to potentially necessitate a move towards route-based charging to reflect local ability to achieve efficiencies. The scale of the impact depends on Network Rail reaching its efficiency target, and the extent to which TOCs can benefit from outperformance. At a minimum, Network Rail would need to undertake considerable work to deliver the data necessary to populate the mechanism, although under its Transformation Programme, the organisation is already reporting at a route level.

Franchised TOCs

– TOCs would need to engage with Network Rail on the development of meaningful, contractualised JNAPs and associated changes to National Rail franchise agreements. They would also need to understand how to develop schemes with Network Rail that will facilitate it to become more efficient, enabling outperformance to be maximised, and the TOC to obtain its share. Rail franchise bidders would need to be able to undertake due diligence in respect of the potential for EBS revenue to be obtained over the life of the franchise. Finally, TOCs would need to develop competencies to be able to demonstrate that a product should be excluded from the fares basket, when they feel there is a case to be made, and the flexibility is commercially worthwhile.

Freight operators

– If the only change to EBS were the pure disapplication of Clause 18.1/Schedule 9A, this is likely to have limited impact on FOCs, except if TOCs start to respond and freight operators obtain benefits (or disbenefits). However, if EBS were made more meaningful at a route/operator level, then freight operators might start to engage—except if the solutions involved deteriorations in service quality. An EBS system for freight would require all operators to agree to a particular form of network-wide response. A national freight JNAP, or a more localised set of JNAPs also covering freight are expected to be delivered in due course (but without the contractualisation being considered in this option); although, exactly how this will work—given that Network Rail's CP4 target is for no change in freight disruption from possessions—is questionable.

ROSCOs

– Limited impact expected.

Funders

– Funders would expect reductions in net cost over time, associated with demand management and/or revenue increases arising from changes to fares regulation. They would also expect outputs to become more affordable to deliver (and passenger revenues to increase) via JNAPs and stronger incentives on Network Rail for outperforming ORR determinations.

The ORR

– Considerable work would be required for the ORR to disaggregate data to support a disaggregated EBS, unless this can reflect Network Rail's new route-based asset management plans. There might be a case for unwinding some aspects of access charges regulation (particularly monitoring functions) if Network Rail successfully introduces risk capital, with the presumption being that regulation can benefit from external monitoring by capital markets. A successful EBS (with operators taking on more monitoring and management responsibilities with respect to Network Rail) would also offer similar benefits. Under the proposed changes to fares regulation, the ORR would need to develop capability to deal with transition from ex ante to ex post fares regulation. It would also have a new customer reasonable requirement to monitor (with respect to JNAPs).

Industry supply chain

– Will need to respond to expected changes in Network Rail behaviour driven by the introduction of risk capital and the disaggregation of EBS.

4.1.5 Transition process

– Contractualising JNAPs would be achieved through changes to National Rail franchise agreements, and ensuring it is clear that compliance with the JNAPs process is a customer reasonable requirement of Network Rail, enforceable by the ORR (so identical to the JPIPs process). Careful monitoring of the PDI measure would need to be undertaken to ascertain whether it is sufficiently robust to be linked to the new arrangements.

– The disapplication of Clause 18.1/Schedule 9A in respect of EBS could be introduced through a reverse of what was put in place at PR08 to apply these terms to this mechanism in the first place.

– If the mechanism were to become more locally focused, then an appropriate geographical level of accounting information would need to be agreed on. This is likely to be at a Network Rail route level, given that its asset management planning now occurs at this level. The information would need to be produced and verified, and benefit sharing at a more local level would need to be agreed. Over time, if it transpires that some areas are less likely to see efficiencies than others, then this would need to be reflected.

– TOCs may see a relaxation of fares regulation as a positive development, but a suitable benefit-sharing mechanism would need to be agreed. The approach assumed for evaluation is that the DfT takes relatively small benefits initially, leaving the gains with the TOCs as a means of helping recovery after recession, with the big gains to funders coming in new franchises.

Timescales

– The contractualisation of JNAPs could be delivered in the next 12 months, although this depends on how franchise agreements are changed.

– Disapplication of Clause 18.1/Schedule 9A in respect of EBS could be achieved in a matter of months.

– Localising EBS would take longer (at least into the next financial year), with a need to produce local data. The introduction of risk capital into Network Rail (debt outside the government guarantee) is also likely to occur in the next financial year, according to Network Rail.

– Changes to fares regulation could be introduced on a new franchise basis, otherwise on January of the earliest possible year (potentially 2012).

The costs of the transition

– The costs of setting up this option are—by construction—expected to be reasonably limited, except in relation to Network Rail issuing debt outside the FIM (although this was anticipated in the CP4 settlement). The latter is expected to cost between £50m and £80m over the remainder of CP4 in additional interest costs—this is according to Network Rail, and depending on the extent to which maturing indemnified debt is refinanced using debt outside the FIM. Management time and some specialist support would be required to set up route-based EBS, revised fares regulation and contractualised JNAPs.

4.1.6 Evaluation

Table 4.1 Evaluation of Option 1a—changes within the existing CP4 regulatory framework

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	Slight positive	Low	Fares regulation changes; JNAPs also (availability increase)	Interviews
Cost efficiency	Positive	Low/medium	Network Rail risk capital; localised EBS	Interviews/previous studies on CLG incentives
Investment coordination	Slight positive	Low	Expect slight improvement from EBS and JNAPs	Interviews
End-user satisfaction	Slight positive	Low	Increase in network availability. However, the fares regulation change might be seen as a negative initially, although in fact it may lead to more efficient pricing in the future	Work on South West RUS (Route Utilisation Strategy)
Changes to interfaces	Limited change to interfaces	Low	Change to fares regulation requires more interaction between TOCs and the ORR; slight increase in interfaces to reflect JNAPs	Based on option construction
Practicality and implementation				
Costs of implementation	–	Low/medium	Depends on the approach adopted by Network Rail to introducing risk capital	Network Rail paper
Time taken to implement, and nature of change required	–	Low/medium	Some franchise change issues.	Network Rail paper and past franchise change events
System-level operation				
Impact on whole railway system	Slight positive	Low	Coordination benefits of JNAPs and localised EBS	

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Implications for awkward interfaces	Slight positive	Low	Limited impact on awkward interfaces other than what would be delivered via localised EBS	Based on option construction: incremental modifications to existing regulatory framework not expected to have an impact in this area
Customer and local needs				
Local responsiveness	Slight positive	Low	Localised EBS and the change to fares regulation are the only impacts	
Government as customer of the 'social' network	No impact	–	No meaningful change relative to status quo, except perhaps due to localised EBS	Based on option construction
Market impact				
Impact on competition	Neutral	Low	Competition between operators might be affected by new fares regulation but ultimately the playing field will still be level	Based on option construction
Impact on economies of scale, scope and density	Neutral	Limited	Mild risk due to localised EBS and any accompanying accounting separation	Based on option construction
Impact on transaction costs	Transaction costs to rise reflecting new fares regime	Low	Change to fares regulation requires more interaction between TOCs and ORR	

Source: Oxera/Arup.

4.1.7 Summary evaluation

Based on the evaluation presented above, a limited impact is expected of such an approach, especially with respect to Network Rail accountability. However, it does offer some prospect of reductions in net cost from demand management by TOCs, as well as higher farebox revenue. There is also expected to be some amount of Network Rail efficiency upside.

What would have to be believed for this option to be successful?

– That the introduction of risk capital by Network Rail will strengthen incentives to outperform the PR08 settlement; that this can be achieved quickly and cost-effectively, and would enable capital market monitoring to lessen the need for monitoring of Network Rail currently undertaken by the ORR and other external stakeholders.

– That contractualised JNAPs will have similar properties to the JPIPs, in terms of their ability to adjust behaviour.

– That TOCs and the ORR will be able to identify which fares products should no longer be regulated, according to precedent from competition economics.

– That EBS will deliver a meaningful change in behaviour, either immediately (through disapplication of Clause 18.1/Schedule 9A), or over time as efficiencies are picked up at a local level. However, this depends crucially on the calibration. At present, EBS only pays out 25% of any saving to TOCs where they have demonstrated involvement—so any mechanism would need to go beyond this level to increase its effectiveness. There is also the issue of how much it can be expected to deliver beyond the level of engagement offered by TOCs in the Network Rail price review process under the status quo.

What type of actions related to this option could be undertaken in the short term?

– Relaxation of fares regulation, with TOCs building the case to the ORR for products to be deregulated.

– Clause 18.1/Schedule 9A provisions with respect to disapplied CP4 EBS arrangements.

– Starting the process of introducing risk capital to Network Rail.

– Contractualisation of JNAPs.

– Starting the process of localising EBS (data disaggregation).

Evidence and scale of impact

The criteria used to evaluate different options reflect a number of qualitative considerations, which are difficult to quantify in a study of this sort, and to combine with those aspects of the evaluation that are more readily quantified. This complicates defining a scale of impact, but the following evidence contributed to the construction of a broad impact range for this option.

By construction, a number of the impacts associated with this option are of low impact, including impacts connected with practicality and implementation (since most of the changes could be implemented quite readily), market impact (limited impact on competition and economies of scale, etc) and other evaluation criteria.

The principal impacts of this option therefore arise from the changes to Network Rail financing, end-user satisfaction (passenger impact of changes to fares regulation), and benefit-sharing mechanisms.

On financing, the issues connected with CLG (Company Limited by Guarantee) status are well understood in the industry, and have been the subject of past studies (eg, Oxera (2006) for the ORR).[22] The question then arises as to the type of impact that a move to introduce risk capital would have on cost efficiency, and whether the efficiency savings and reduced opportunity costs would more than offset the additional interest costs.

Experience from the water sector reveals that Welsh Water, a water and sewerage company with CLG status, has generally done no better than average in Ofwat's assessments of cost efficiency, which may be attributable, in part, to its CLG status.[23] This suggests that risk capital would, in line with economic theory, improve cost performance. Interviews with stakeholders have also suggested that the impact of CLG status, and the government guarantee on debt repayments, leads managers at Network Rail to focus more on external monitoring than is strictly necessary.

End-user satisfaction is expected to increase slightly. The option is intended to improve network availability and overall cost, reduce crowding at peak times and to lead ultimately to a more 'economic basis' for pricing. Against this assessment is concern that higher fares may not actually be successful in spreading demand, a finding suggested in Network Rail's South West Main Line Route Utilisation Strategy,[24] March 2006. It is also noted that spreading demand requires capacity at shoulder times.[25] However, on balance, we expect the overall impact on end-users to be slightly positive.

Efficiency benefit-sharing mechanisms at the time of the PR08 settlement were expected to contribute to Network Rail's overall efficiency performance,[26] with Network Rail sharing 25% of outperformance with train operators.

The **overall scale of the impact is expected to be medium**. Net cost reductions of about £100m to £500m per annum are expected, depending on the nature of the changes implemented. These ongoing benefits would be driven largely by changes to fares regulation, and from strengthened incentives on Network Rail with respect to cost efficiency and network availability. Transition costs, by their nature, would be once-off, and are not expected to be large for this option.

The costs of setting up this option are—by construction—expected to be reasonably limited, except in relation to Network Rail issuing debt outside the FIM (although this was anticipated in the CP4 settlement).

4.2 Option 1b—franchise reform

The key objectives of franchise reform are summarised in Box 4.2

Box 4.2 Key objectives

Deliver change without the need for change in legislation.

Produce strong alignment of incentives between franchised operators and Network Rail to minimise net cost, and maximise the benefits of the industry to the economy.

Enable franchised operators to focus on passenger needs, by enabling franchise costs and revenues to respond to market changes.

[22] Oxera (2006), 'The role of incentives in the GB rail industry'.
[23] NERA (2006), 'Corporate Form, Financial Guarantees, and Efficiency Performance: Expectations and Evidence', December.
[24] Network Rail (2006), 'South West Main Line Route Utilisation Strategy', March.
[25] Department for Transport, Transport for London and Network Rail (2007), 'Demand management techniques – peak spreading', April.
[26] Office of Rail Regulation (2008), 'Periodic review 2008. Determination of Network Rail's outputs and funding for 2009-14', October.

4.2.1 Description

This option expects radical change to the franchise agreement, albeit within the scope of existing legislation—specifically, the following.

- Longer franchises—not naively expecting 'investment'—but instead to facilitate a longer-term view of the business that the TOC has inherited, and the introduction of new incentives on the franchisee.
    - Break points every five years to ensure that good performance is rewarded with continuation.
- Change key parameters in the franchise agreement, limiting revenue risk within predetermined bands that vary by market; exposure to performance risk; and enabling costs to vary within the franchise in response to market requirements (eg, more/fewer services).
    - The lower beta (financing costs) inherent in these changes could be used to enable more exposure to infrastructure cost outturns.
- Consistent with this, stronger incentives on Network Rail with respect to passenger volumes.[27]

This overall package—the details of which will need to be developed over time due to their inherent complexity—could either comprise of the creation of a combined (infrastructure and operations) P&L account at a route level, with all parties receiving a share of profits or losses; or individual incentives on each party. In the latter case, the TOC would take risk on a proportion of Network Rail's expenditure changes at a route level, while Network Rail would take risk on a route basis in relation to unit changes in volumes. Such a package is expected to improve investment coordination across the wheel–rail interface, providing improved inputs into the HLOS and PR processes.

It is recognised that the changing of risks to which TOCs and Network Rail are exposed, and the sharing of the cost reductions across infrastructure and operations, will require careful implementation. In particular, the net cost impacts on the industry as a whole from, for example, exposing TOCs to more risk and having this priced into franchise bids, would need to be reflected in any modelling and practical implementation.

- Network Rail to issue equity, with a view to strengthening incentives to outperform periodic review settlements.
- TOCs to lead investment at most stations (albeit in the context of an increase in Network Rail's managed stations portfolio to reflect locations with considerable heritage, investment or development requirements) within the context of 99-year full repairing leases.
- TOCs to lead investment in rolling stock. It is noted that, nominally speaking, TOCs can lead the procurement of new rolling stock under existing arrangements, but recent trends have seen funders taking this responsibility on. This option reverses this trend in the context of longer franchises.

[27] The details of a volume incentive mechanism have not been defined as part of this study, but could perhaps take the form of a percentage of the increase in each TOC's volume over its base revenue at the start of the franchise or franchise breakpoint, or could be split by route and be based around the percentage of passengers and volume over the base level at the start of a control period. These options, and alternatives, would require further analysis as to their implementation, operation and likely incentive power.

– Disapplication/relaxation of fares regulation, similar to option 1a, albeit with the additional expectation that regulation can evolve to permit operators to set fares according to narrow time intervals in the morning and evening peaks, with the support of smart ticketing solutions.

4.2.2 Which incentive misalignments would this option assist with?

– Designed to re-integrate the industry P&L through TOC exposure to Network Rail costs, and Network Rail exposure to operator revenues. Facilitated by lower beta franchised operators.

– Limits on user-responsiveness on the part of TOCs; instead they get a long-term business with more control over its performance through focus on the passenger.

– Lack of incentives for investment coordination.

– More user-focused investment in trains and stations to overcome existing disconnections.

– Limited TOC interest in their actual revenues per train due to revenue allocation approaches, and limited information.

4.2.3 Impact on players

Network Rail/IMs

– The main impact on Network Rail would be through changes to the allocation of risk, including the assumption of some GDP (beta) risk, while sharing with franchisees some infrastructure risk. This will change the risk profile of Network Rail, with implications for its financing costs, and will also lead to different forms of engagement with the operators, both at the franchise bid stage and during the franchise.

– Considerable changes in responsibility for stations, with an increase in the managed stations portfolio, but diminished responsibility for the remaining stations.

– If equity were introduced, along with the possibility of takeover (the 'market for corporate control'), then the changes could be quite substantial, with incentives for outperformance strengthened.

Franchised TOCs

– Significant changes to franchises imply that the capital structure and organisational profiles of TOCs will need to change in response. During a franchise bid, the operator will need to capture changes to the franchise agreement, including the impact on margins of changed train-level revenue and cost risk, the need for due diligence with respect to the infrastructure cost risk it is taking on, the longer time horizon and scope for changes at breakpoints, and greater responsibilities around stations and rolling stock procurement.[28]

– Once a franchisee has taken over a business under such arrangements, it will be expected to manage Network Rail as its supplier much more actively than at present. It will need to lead investment at the majority of its stations and in any new rolling stock it needs, and keep under constant review opportunities for demand management via smart ticketing arrangements.

– Some localisation of charges is expected.

[28] The detail of what precisely would happen at the breakpoint is not defined here; for example, resets could take place against revenue lines, or against the Network Rail cost line, or some other basis. This aspect of the option is not expected to increase risk to TOCs inasmuch as they will have certainty as to what risks are expected to be managed over particular periods.

Freight operators

– Limited impact expected on freight operators, although care will have to be taken to ensure that risk-sharing arrangements between Network Rail and franchised operators do not limit sufficient focus being placed on the needs of freight operators on a route.

ROSCOs

– The main impact on ROSCOs will arise through TOCs taking a different approach to investment in rolling stock than under the existing arrangements. ROSCOs will still need to purchase most new rolling stock, even in the context of longer franchises, although some operators may be willing to take the residual value risk themselves. ROSCOs will be expected to price break clauses into lease agreements, in order to enable TOCs to change capacity on routes in response to economic circumstances.

Funders

– There is expected to be limited impact on funders in the short term, while in the longer term with stronger efficiency incentives on Network Rail, better investment coordination and decision-making, there is expected to be a considerable reduction in net cost required to deliver similar industry outputs.

The ORR

– The ORR will need to reflect new incentive arrangements on Network Rail in its determinations, including localised volume risk sharing. A key task will relate to rolling back regulation in response to equity-based financing, with capital markets discipline expected to replace monitoring and incentive arrangements set up specifically to face Network Rail's CLG status. It is also expected that to facilitate risk sharing across the TOC–Network Rail interface, accounting separation at route level will be required with respect to Network Rail. The ORR will need to monitor that TOC–Network Rail risk sharing does not occur to the detriment of freight and open access operators.

Industry supply chain

– Limited direct impact.

4.2.4 Transition process

– The majority of change is expected to be achieved through a new franchise agreement. For existing franchises with more than two years to run, changes would be introduced via existing change mechanisms.

– For information purposes, TOC franchises' start and end dates are presented in the table below.

Table 4.2 TOC franchises' start and end dates

Franchise	Franchise start date	Franchise end date
East Midland Trains (EMT)	11-Nov-07	01-Apr-15
East Coast Mainline (ECML)	14-Nov-09	
Northern	12-Dec-04	12-Sep-13
TransPennine Express (TPE)	01-Feb-04	28-Feb-12
First Capital Connect	01-Apr-06	31-Mar-15
London Southeastern Railway	01-Apr-06	31-Mar-14
Southern Railway	20-Sep-09	22-Jul-17
Stagecoach South Western (SSW)	04-Feb-07	04-Feb-17
Chiltern	03-Mar-02	31-Dec-21
Arriva CrossCountry (AXC)	11-Nov-07	31-Mar-16
C2C	26-May-96	29-May-11
National Express East Anglia (NXEA)	01-Apr-04	31-Mar-11
First Greater Western	01-Apr-06	31-Mar-16
Arriva Trains Wales (ATW)	08-Dec-03	14-Oct-18
London Midland (LOM)	11-Nov-07	20-Sep-15
Virgin West Coast (VWC)	09-Mar-97	31-Mar-12

Source: DfT.

– According to Network Rail, in order to issue equity, it would either have to reach agreement with CLG members to introduce private sector shareholders, or be nationalised and then once again privatised.

– Change to Network Rail access charging via licence modifications to strengthen volume incentives on a route basis. An interim review is not expected to be necessary, unless Network Rail (and its investors) considers changes to increase risk materially.

Timescales

– Licence modifications for Network Rail could be introduced immediately.

Costs

– Changes would have limited upfront costs, although there would be a need for surveys at stations, and higher franchise bid costs (eg, more due diligence, or longer time needed to account for new features and/or greater complexity).There will be some impact on the revenue requirement from the introduction of equity.

4.2.5 Evaluation

Table 4.3 Evaluation of Option 1b—franchise reform

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	Positive	Medium, tempered by demand management	Better incentives deliver volume growth, except in peaks	Interviews/previous work
Cost efficiency	Positive	High	Stronger incentives on operators to influence Network Rail efficiency. Equity issuance also contributes. Incentives on TOCs to improve own cost-efficiency	Interviews/previous studies on CLG incentives
Investment coordination	Positive	Medium	Expect aligned incentives to enable better investment coordination through greater role for TOCs	Greater input of users in other sectors (eg water) seen as better relating investment with needs and willingness to pay
End-user satisfaction	Positive	High	Driven by more focused operators	Interviews
Changes to interfaces	Number of interfaces expected to fall	Medium	Better aligned incentives should reduce the need for 'man-marking'	Interviews
Practicality and implementation				
Costs of implementation	–	Low/medium	Some franchise change issues only around the pace of change; the cost of Network Rail equity issuance depends on form	Interviews
Time taken to implement, and nature of change required	–	Low/medium	Within framework of franchising, except for equity issuance	Based on option construction; Network Rail work
System-level operation				
Impact on whole railway system	Positive	Medium	Coordination across interfaces designed to improve	Based on option construction

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Implications for awkward interfaces	Positive	Medium	Addresses difficult interfaces at stations, in relation to rolling stock, and across the wheel–rail interface	Based on option construction
Customer and local needs				
Local responsiveness	Slight positive	Low	Option will not necessarily deliver this, although operator incentives to help	Based on option construction
Government as customer of the 'social' network	-	-	Option will not necessarily affect this	Based on option construction
Market impact				
Impact on competition	Neutral	Limited	Measures to be introduced should make franchises more attractive	
Impact on economies of scale, scope and density	Neutral	Limited	Limited impact expected	–
Impact on transaction costs	Transaction costs to fall	Medium	Expect better incentive alignment to deliver reductions	Interviews

Source: Oxera/Arup.

4.2.6 Summary evaluation

We expect this option to deliver a considerable impact on net cost through stronger incentives on Network Rail to reduce costs (via equity issuance and TOC involvement), and much stronger incentives on TOCs (and Network Rail) to focus on delivering revenue growth. Better investment decisions should also be taken by virtue of better-aligned incentives, and coordination across the wheel–rail interface.

What would have to be believed for this option to be successful?

– That providing TOCs with a long-term business, and a different form of risk profile, where they are able to take 'normal' decisions to grow revenues and minimise costs, will deliver considerable net cost savings.

– That more accountable operators will make good investment and day-to-day decisions around stations, service patterns and rolling stock.

– That incentivising TOCs in respect of Network Rail's costs will deliver greater efficiencies, more quickly. Similarly, Network Rail's stronger volume incentives will enable operators to maximise demand.

– That Network Rail introducing equity will strengthen cost-efficiency incentives.

– Operators will be better able to respond to demand shocks, and will require lower bid margins than under the status quo, except for the infrastructure cost risk they will assume.

What type of actions related to this option could be undertaken in the short term

– Specify the Greater Anglia franchise, along the lines of the design set out above.

– Encourage demand management via softer measures (eg, publish average load factors per train in timetables).

– Move responsibility for station management and rolling stock procurement to TOCs with franchises ending in more than two years from January 2011.

– Begin negotiations with TOCs with more than two years to go on franchises in respect of changing their franchise agreements along the lines described above.

Evaluation and scale of impact

The introduction of outside equity can be expected to improve pressures on Network Rail to improve its overall cost efficiency. As well as cost efficiency, Network Rail would also face incentives to grow revenue with a financing structure that includes equity. Network Rail's work notes that the timing of these steps would depend on the process it needs to undertake, with the earliest sale of equity taking place close to the end of CP4.

By 'bridging the gap' in the status quo between infrastructure management and operations, and realigning incentives across this interface, we expect this option to deliver considerable improvements in net cost. This would occur through further strengthening of cost-efficiency incentives on Network Rail, and clearer focus on revenues by TOCs and Network Rail. Better investment decisions—aligned to the needs of end-users—would also reduce opportunity costs.

From improved incentive alignment, we expect this option to deliver improved functioning of the railway as a system (across the wheel–rail interface), and improved responsiveness of operators to the needs of local markets.

The impact on competition is expected to be slightly positive. The new franchise design could make franchises more attractive business propositions, although certain aspects will attract different types of bidders. Much greater facilities management requirements with respect to stations might require existing operators to team up with specialists in this area, while the prospect of taking on infrastructure cost risk will be more attractive to some bidders than others.

We would expect transaction costs to reduce as incentives across interfaces are improved.

The **overall scale of the impact is expected to be high**. Net cost reductions in excess of £500m per annum are expected, depending on the nature of the changes implemented. These benefits would be driven largely by stronger incentives for cost efficiency, revenue growth, demand management, and improved investment coordination, as follows. Risk sharing between Network Rail and TOCs should strengthen the efficiency potential of the former organisation, as will equity issuance. TOCs should see benefits flowing from Network Rail volume incentives, increased revenues (eg, through faster removal of speed restrictions), better understanding of costs of interventions such as possessions, better risk allocation, more flexibility (costs respond to volumes) and lower probability of default.

Economic fares regulation, plus a longer-term move towards pricing by time slide, should also improve resource allocation, as will better rolling stock procurement and better management of stations; ATOC's 2010 document on franchise reform identifies possible savings of £250–£500m over five years.[29]

In terms of transition costs, these are likely to be limited apart from the re-surveying of stations (Network Rail spends £20m per annum on condition surveys for one-fifth of its stock) and the changes required at Network Rail to permit equity issuance.

4.3 Horizontal separation by region

The objectives of this option are described in Box 4.3.

Box 4.3 Key objectives

To develop comparative competition for the provision of rail infrastructure, enabling innovation in financing, standards, contracting and the delivery of cost-effective infrastructure.

To deliver stronger incentives on the infrastructure providers in respect of volume, network availability and capacity utilisation.

To provide more local focus in infrastructure delivery, while minimising operational interfaces.

4.3.1 Description

This option focuses on the structural separation of Network Rail into a series of regional IMs, with emphasis on minimising operational interfaces.

In order to minimise operational interfaces across boundaries, a number of different splits would need to be considered. The likely coverage of the different IMs, subject to further analysis, could be:

– Scotland;
– Merseyside;
– Greater Anglia;
– Wales;
– the former Southern region;

[29] Association of Train Operating Companies (2010), 'Franchise reform and better value for money in rail', March.

– the rest of England, which could, for example, be broken down into Northern, East Coast, West Coast, and Great Western.[30]

Despite attempts to minimise operational interfaces, there is still a need for an overall system operator (SO) function, focused on coordinating the operation of the national GB network. The SO would take on Network Rail's national functions relating to strategic and network planning (within the HLOS framework), and with similarities to Network Rail's approach to Route Utilisation Strategies (RUSs), would sponsor Network Rail company standards and industry systems (such as ITPS and Trust), and would allocate capacity in the timetabling process (in relation to both long-term access rights, and possession plans). It would be incentivised through economic regulation by the ORR to maximise network availability, performance and volumes (via capacity utilisation), and be financed by equity.

In effect, this would be a 'shallow' SO, and contrasts with the type of 'deep' SO function possessed by Railtrack.

IMs would retain local controls, signalling, etc, and operational (as well as engineering) safety responsibilities.

4.3.2 Which incentive misalignments would this option assist with?

– A lack of efficiency incentives in the status quo would be improved due to comparative competition, and more innovative solutions should emerge, so long as these are not hampered by the need for national standards.

– A lack of local focus and responsiveness.

– Enables a proliferation of financing approaches, which should move away from 'meet but not beat' incentives with respect to the periodic review settlement seen in the status quo (Network Rail's CLG status). Similarly, the option enables multiple approaches to contracting for maintenance, renewal and enhancements, covering a variety of in- and outsourcing arrangements, depending on the circumstances of each IM, and management preferences.

– A lack of precise incentives in respect of network availability (going beyond the PDI), and the ability to trade off capacity utilisation against performance, on which the SO function should improve.

– Responsibility for national standards would be vested in the SO.

4.3.3 Impact on players

Network Rail/IMs

– Network Rail in the transition will be heavily focused on ensuring a smooth changeover, and will likely incur considerable cost as this process evolves. Risk of loss of management focus (and, hence, on delivering efficiencies) over this period.

– The new IMs will need to finance themselves for the first periodic review, and focus on delivering outputs safely within this envelope. Their interfaces with the SO will be crucial, particularly with respect to the maintenance function, and how this interacts safely with the operations function—especially around response to incidents. Consideration will then turn to obtaining efficiencies and innovating.

[30] The Isle of Wight would remain as is.

Franchised TOCs

– Interface with the infrastructure would be similar to now—focused mainly on the SO in terms of timetable development, and the IMs for day-to-day operations. Existing route planning arrangements would be maintained across the players to facilitate more localised decision-making. This should be made easier with a less centralised IM function. Note that franchised TOCs with longer contracts are likely to have somewhat greater buyer power in this setup, which should increase efficiency pressures over and above the pure comparative competition arrangements; although, this power will be limited by the fact that the infrastructure supplier cannot be changed.

Freight operators

– Will be concerned about long-term access rights, and facing more IMs on long-distance freight routes, with the potential for multiple charging methodologies. FOCs will want the SO to make things as simple as possible (eg, the SO acting as one-stop-shop for the freight operator to go to when it has a customer wishing to develop a new flow). Will probably respond positively to the SO role, particularly if it is heavily incentivised with regard to availability, volumes and performance, and interfaces between the SO and IMs are handled well.

ROSCOs

– Will need to engage with multiple IMs in relation to new vehicle introductions, and the acceptance process might become fragmented if the new stock needs to cross boundaries. A single national process, as exists now, might be maintained, perhaps under the auspices of the SO. Under this option, there would be a need to avoid proliferation of standards (which would enable IMs to innovate) becoming a problem for suppliers such as ROSCOs.

Funders

– Continue to develop HLOSs and SoFAs as at present, and should expect horizontal separation to deliver greater cost efficiency in the medium term. Will need to have a well-developed relationship with the national SO as the industry planning body. A key decision will relate to local cost-reflectivity (in terms of financial support to alternative IMs); this depends to some extent on how the Network Rail RAB is split to facilitate horizontal separation.

The ORR

– Will need to develop clear information protocols to be able to maximise the benefits of comparators. Will be setting more price controls, and using comparative data (under separate ownership) for the first time with respect to OMRE. It will also have to develop price controls for the SO function. Effort will be required initially to develop additional licences (for IMs and the SO), and to split the RAB. It will need to place less reliance on international benchmarking, although it will still be a useful measure of absolute GB efficiency.

Industry supply chain

– Opportunity to work for different IMs—increases competition for larger number of contracts. More opportunities for contracts means that there is less likelihood of a feast–famine cycle in respect of maintenance and renewals contracts. IMs would have an incentive to drive out costs, not necessarily apparent under current structures that are imposed on contractors.

4.3.4 Transition process

There are two likely approaches, of which the latter is likely to be preferable.

1. Forcible separation of Network Rail following Competition Commission (CC) inquiry/Act of Parliament.

2. Setting regional price controls along the boundaries of the regions established above, Network Rail may decide to separate out its OMRE businesses along these lines (as occurred in case of the gas distribution networks, DNs). The ORR would also develop a SO price control at this stage. Over time, the new structure would establish itself.

Timescales

– Under 1) a CC inquiry would need to follow due process—approximately 36 months from ORR Enterprise Act investigation; this may push timings into CP5. An Act of Parliament would be likely to expedite matters, and lead to a more certain outcome than the CC process.

– Under 2) the process is likely to be faster (in terms of price control separation), and also may achieve the desired outcome if Network Rail decides to sell the SO and IM functions that have been separated in terms of economic regulation. However, this may not be achieved until the CP5 period.

Costs

– Transition costs, at least under option 2) might be quite modest, assuming Network Rail sells the separated businesses. Likely to be some loss of focus on delivering CP4 outputs efficiently in the interim period.

4.3.5 Evaluation

Table 4.4 Evaluation of horizontal separation by region

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	Positive, given incentives on SO and IMs	Low/medium	Currently no strong volume incentive on Network Rail; introduce to SO—separate organisation may facilitate better-targeted incentives	Energy sector SO models
Cost efficiency	Positive	Medium/high	Comparative competition and innovation, and new financing will support cost-efficiency improvements at Network Rail	Gas DN separation
Investment coordination	Neutral	Limited	Negative impact from separation evened out by SO incentives	Gas DN separation
End-user satisfaction	Positive, given strong incentives on SO	Low/medium	Increase in network availability and capacity utilisation	Energy sector SO models
Changes to interfaces	Interfaces to increase	Medium	If increase in interfaces is handled by SO only, and it makes timetable development more efficient if done locally while respecting system-wide issues	Interviews; models from other sectors
Practicality and implementation				
Costs of implementation	–	Medium	Depends on the approach adopted	Separation in GB electricity and gas sectors
Time taken to implement, and nature of change required	–	Medium	Depends on the approach adopted. Primary legislation might be avoidable	Separation in GB electricity and gas sectors
System-level operation				
Impact on whole railway system	Slight negative	Limited	SO provides national function although will have crucial role in avoiding negative consequences of fragmentation of infrastructure; track to rolling stock shift less clear	Interviews and energy sector SO experience

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Implications for awkward interfaces	Slight negative	Low/medium	Option arguably does little to address existing interfaces, and overall likely to lead to an increase in interfaces	Based on option construction
Customer and local needs				
Local responsiveness	Slight positive	Low	Regional IMs will be more locally focused	Based on option construction
Government as customer of the 'social' network	Neutral	Limited	No direct impact on government's position in this regard	Based on option construction
Market impact				
Impact on competition	Positive (with respect to infrastructure)	Medium	Increase in comparative competition; more varied work for contractors	Gas DN separation evaluation
Impact on economies of scale, scope and density	Neutral	Limited	Some economies of scale may be lost, but the impact is expected to be limited (eg, from having multiple IMs)	ITS Leeds at PR08, ORR modelling
Impact on transaction costs	Negative	Medium	Increased number of IMs, and interfaces between SO and IMs	Electricity and gas separation

Source: Oxera/Arup.

4.3.6 Summary evaluation

Overall, this option appears to offer the possibility of a positive impact.

What would have to be believed for this option to be successful?

– That the long-term timetabling (with the SO) split from day-to-day operations (with the IMs) could be worked out.

– That the transition from Network Rail to the new IMs/SO could be worked out smoothly, especially in relation to Network Rail debt, the recruitment of senior staff, employee relations, and the transfer of contracts.

– That effective incentives could be introduced on the SO with respect to performance, network availability and capacity utilisation.

– That comparative competition would deliver benefits of the order seen in gas DNs. It might be the case that the benefits are even greater than seen in the gas DNs context, given that benefits quoted in that case related only to controllable operating expenditure (OPEX) (whereas here the SO function and maintenance, renewals and enhancements are also available for consideration).

– That retaining the planning function within the SO (mini-Network Rail) could be efficient and effective, and that such a body could command industry respect in this function.

What type of actions related to this option could be undertaken in the short term?

– Develop key performance indicators (KPIs) based on regional boundaries.

– Develop shadow price controls based on regionalisation and the SO function.

Evaluation and scale of impact

This option is expected to have its main impacts on: revenue growth, cost efficiency, competition, transaction costs, and economies of scale. Available evidence on these issues is reviewed below; the experience of the energy sector is particularly informative in relation to competition, cost efficiency, and transaction costs, and is considered first.

The National Audit Office's (NAO) 2006 report on Ofgem's impact assessment of gas separation noted significant reductions in energy and water networks attributable to comparative regulation, ranging from 4.9% per annum in electricity transmission in England and Wales, to 5% in water and sewerage in England and Wales, and to 7.7% in electricity distribution in Great Britain.

Comparative regulation of Network Rail will be supported by the regionalisation. Experience with the separation of the gas DNs[31] is described in Appendix 3. This experience found benefits for regulatory purposes that were substantially higher for four comparators (£355m) than for one comparator (£145m). These benefits related largely to the anticipated impact on 'controllable OPEX', which is broadly the day-to-day costs of running the regulated business, including staff costs, IT and insurance.

In the rail context, the big 'prize' is in relation to Network Rail's renewals expenditure, which may be slightly more difficult to benchmark than some of the process-driven elements of the controllable OPEX in the rail context; although, comparative competition should have a positive impact on determining whether particular pieces of renewals expenditure actually need to be undertaken, and, perhaps, also its scope.

[31] The costs of gas distribution account for about 20% of the final customer bill. It is noted by way of comparison that total average annual CAPEX assumed at the 2007 price settlement was approximately £663m in 2005/06 prices. See National Audit Office (2006), 'The Office of Gas and Electricity Markets: Sale of gas networks by National Grid', February.

For example, BSL's 2008 study[32] found significant gaps between Network Rail and the European average for maintenance, which were even more pronounced for renewals. On renewals, 30% of the gap was down to labour costs, 16% on plant procurement practices, 24% from possessions taken over too short a piece of line, 20% due to poor utilisation of possessions, and the remainder (approximately 10%) attributed to 'transaction costs' between Network Rail and its contractors. Experience from other sectors does not directly describe exactly how a gap of this nature would be overcome by comparative competition, and it is plausible that some of these issues will remain after a horizontal separation.

Horizontal separation did imply costs for gas shippers (very roughly speaking, one might consider TOCs as a type of analogue for shippers). To avoid the need for shippers to interact with multiple DNs, a single industry interface was created. The development costs of the agency to ensure that this single industry interface was retained resulted in one-off costs to shippers of about £25m, with ongoing yearly costs of £7m.

A further rough and indicative type of comparison may be drawn with the gas sales, which required changes to the way in which capacity was allocated from the transmission network. These changes ('offtake reform') were introduced by Ofgem to protect the interests of users in the 'fragmented' industry structure that would arise following the sales. One can imagine that the creation of the type of shallow SO described above would serve to achieve a broadly similar end in the rail industry, attempting to protect the interests of users in a fragmented industry structure, and working to ensure that operators' interests were not prejudiced by potentially having to deal with several infrastructure companies.[33] The additional regulatory costs and costs associated with the new offtake arrangements were estimated at £100m. Feedback suggested that the costs to the gas industry had been understated, as Ofgem had not quantified end-user costs.

The NAO's 2006 review[34] of Ofgem's analysis found Ofgem's estimate of potential benefits from the gas DN horizontal separation—95% of which were due to more comparators—to be somewhat conservative. A subsequent report by Europe Economics[35] identified scope for efficiency gains for the gas DN sector as a whole of 1.9–3.7% per annum over and above the underlying growth in productivity in the economy as a whole; although, this does not account for a counterfactual of continued ownership/integration, and the scope for efficiency improvements in individual gas DNs will depend on their specific circumstances.

It is also worth bearing in mind that the cost efficiencies described above for energy and water companies were achieved, in almost all cases, by equity-financed companies. The option under consideration here has been constructed so as to allow different financing structures to emerge. In practice, a degree of 'shadow' horizontal separation would allow Network Rail to investigate different ownership structures, and possibly also the discretion to dispose of certain regions.

Different sales would also have implications for economies of scale and scope. National Grid suggested during the sale of the gas networks[36] that the sold firms could benefit from economies of scale and/or scope under new ownership. For example, being part of National Grid may have conferred scale advantages, but new owners may extract scope economies that were not possible under integrated ownership given the characteristics of the businesses that could conceivably have bought the networks.

[32] BSL (2008), 'Rail Infrastructure Cost Benchmarking – Brief LICB-gap analysis and cost driver assessment', April.
[33] It is worth noting that the offtake arrangements are different to the functions of the gas SO—the point being made here is that they might be seen as a set of reforms that is intended to achieve some of the aims of an SO in the rail industry.
[34] National Audit Office (2006), 'Sale of gas networks by National Grid', February.
[35] Europe Economies (2007), 'Top down benchmarking of UK gas distribution network operators', April.
[36] Letter from National Grid Transco to Kyran Hanks, Ofgem, September 26th 2003, cited in NAO (2006), op. cit.

This raises the question of whether industry net costs would rise or fall if multiple infrastructure companies are created. Research by the ORR, which builds on the work undertaken by ITS Leeds at PR08[37] suggests that economies of scale (or at least route kilometres) are largely exhausted at a scale significantly below the current size of Network Rail, suggesting costs would not rise if an appropriate regional structure were to be imposed.

The introduction of a SO is also seen as delivering benefits under this option.

In the GB electricity sector, the core role of the SO is management of the electricity system in real time—to match generation with demand in real time, to minimise the cost of transmission constraints in causing generating plant operation to depart from the wishes of generators and to minimise the other costs of day-to-day system operation.[38] It also manages the commercial relationship between transmission and network users, including the charging for and rationing of existing network capacity.

The deeper aspects of transmission are handled by the network operators (ie, National Grid (in its TO role) and the two Scottish transmission businesses. This has some similarities with the type of SO proposed in this option, where infrastructure companies retain responsibility for elements of system operation.

As far as incentivisation of the SO's day-to-day system management is concerned, this is mainly handled through an SO incentives mechanism, and this is set annually.[39] The form of the one-year SO incentives is that:

– a target level of external SO costs is set;
– the SO is rewarded to the extent that costs turn out lower, and penalised to the extent that costs turn out higher—with both upside and downside limited by a cap and collar respectively.

The role of the SO in the electricity industry, which is arguably the most fully articulated model of system operation in the GB utilities sector, is explained in further detail in Appendix 3.

More generally, the deeper aspects in the electricity transmission industry are driven by a variety of regulatory mechanisms. This is not to say that the approach adopted in electricity is necessarily optimal, but it does provide an example of how SO-type functions can be specifically targeted and incentivised with a view to delivering better outcomes than less well-targeted incentive mechanisms. The efficacy of such incentives in the rail context would depend on a variety of factors, including operational details, and the capital structure of the regional IMs.

Overall, this option is viewed as being of **medium impact**. Net cost reductions of £100m £500m are expected, depending on how changes are implemented. Transition costs might be on a modest scale, particularly if a phased and evolutionary approach were to be adopted (eg, shadow regional price controls), but there will be transaction costs associated with multiple IMs.

[37] Institute for Transport Studies (2008), 'International benchmarking of Network Rail's maintenance and renewals costs: An econometric study based on the LICB dataset (1996-2006)'.
[38] Unlike with most electricity systems round the world, the job of the NETSO is not to 'despatch' generating plant according to its view of what are the cheapest (ie, lowest short-run marginal cost) plant on the system. The GB wholesale electricity market works, in the first instance, on the basis of self-despatch by generators (ie, generators telling the NETSO how much they want to generate in each half-hour period)—it is only close to real time that the NETSO, in effect, takes over and buys 'increments' and 'decrements' of generation from generators (relative to the generators' own submitted plans) to balance the system.
[39] This SO incentives mechanism covers purely the 'external' costs of system operations—for example, the cost of transmission constraints, the cost of paying reserve generating plant. Revenue to cover the 'internal' operating costs of the SO is set at the normal periodic review.

4.4 Horizontal separation by market

Box 4.4 describes the key objectives of this option.

Box 4.4 Key objectives

To develop comparative competition for the provision of rail infrastructure, enabling innovation in financing, standards and the delivery of cost-effective infrastructure.

To deliver stronger incentives on the infrastructure providers in respect of volume, network availability and capacity utilisation.

To provide more local focus in infrastructure delivery, reflecting specifically the differences in passenger markets in the way that infrastructure is provided.

4.4.1 Description

This option focuses on the structural separation of Network Rail into a series of market-based IMs, with emphasis on matching the infrastructure businesses to passenger markets. A degree of franchise mergers is also envisaged to match the number of IMs.

In order to reflect passenger markets, the coverage of the different IMs could be:

– inter-urban, matched to core strategic routes (eg, Trans Pennine);
– London and the South East, matched to current London commuter franchises with no overlap to the above (Southeastern, South Western, Southern, c2c, Chiltern, Greater Anglia);[40]
– remaining 'regional' operations, matching to franchises such as Northern, Merseyrail, ScotRail, and ATW. We envisage a number of regional IMs, reflecting more localised funding streams.[41]

As noted in section 4.3, there could be close similarities between this model and the regional model, depending on how the IM is split among markets. Some of the possible differences between the two approaches—which, as noted, could plausibly become immaterial depending on the manner of implementation—are described in section 4.4.6.

Within such a model, there is a need for an overall SO function, focused on coordinating the operation of the national GB network. This would take the same shallow form as in the case of regional separation. Some of the differences between this model and the regional separation model are considered below as part of the evaluation.

4.4.2 Which incentive misalignments would this option assist with?

– A lack of efficiency incentives in the status quo would be improved due to comparative competition, and more innovative solutions should emerge, so long as these are not hampered by the need for national standards.

– A dearth of local responses to the needs of local markets, especially on the social railway.

– A perceived lack of management attention to railways with different cost structures and needs.

– Better alignment of infrastructure with operations (in terms of geography) should enable more coordinated behaviour across these interfaces.

[40] As this stage, we do not take a view on which franchises might be merged, but possibilities include C2C with National Express East Anglia, London Midland and WC; and East Coast and FCC.
[41] The Isle of Wight would remain as is.

– Enables a proliferation of financing approaches, which should move away from 'meet but not beat' incentives with respect to the periodic review settlement seen in the status quo (Network Rail's CLG status). Similarly, the option enables multiple approaches to contracting for maintenance, renewal and enhancements, covering a variety of in- and outsourcing arrangements, depending on the circumstances of each IM, and management preferences.

– A lack of precise incentives in respect of network availability, and the ability to trade off capacity utilisation against performance, which the SO function should improve on.

– Franchise mergers should, if implemented, reduce duplication of resources and lower bidding costs, although there would be some risk to on-rail competition.[42]

4.4.3 Impact on players

Network Rail/IMs

– Network Rail in the transition will be heavily focused on ensuring a smooth changeover, and will likely incur considerable cost as this process evolves. Risk of loss of management focus (and, hence, on delivering efficiencies) over this period.

– The new IMs will need to finance themselves for the first periodic review, and focus on delivering outputs safely within this envelope. Their interface with the SO will be crucial, particularly with respect to the maintenance function and how this interacts safely with the operations function—particularly around response to incidents. Consideration will then turn to obtaining efficiencies and innovation.

Franchised TOCs

– The interface with the infrastructure would be similar to now—focused mainly on the SO in terms of timetable development, and the IMs for day-to-day operations. Existing route planning arrangements would be maintained across the players to facilitate more localised decision-making. This should be made easier with a less centralised IM function. Note that (post-merger) franchised TOCs with longer contracts are likely to have more buyer power in this setup, which should increase efficiency pressures over and above the pure comparative competition arrangements.

Freight operators

– Will be concerned about long-term access rights, and facing more IMs on long-distance freight routes, with the potential for multiple charging methodologies. FOCs will want the SO to make things as simple as possible (eg, the SO acting as a one-stop-shop for the freight operator to go to when it has a customer wishing to develop a new flow). Will probably respond positively to the SO role, particularly if it is heavily incentivised with regard to availability, volumes and performance, and that interfaces between the SO and IMs are handled well.

ROSCOs

– Will need to engage with multiple IMs in relation to new vehicle introductions, and the acceptance process might become fragmented if the new stock needs to cross boundaries. A single national process, as exists now, might be maintained, perhaps under the auspices of the SO. Under this option, there would be a need to avoid proliferation of standards (which would enable IMs to innovate) becoming a headache for suppliers such as ROSCOs.

[42] It is noted that the degree of on-rail competition varies considerably across the network.

Funders

– Still develop HLOSs and SoFAs as at present, and should expect horizontal separation to deliver greater cost efficiency in the medium term. Will need to have a well-developed relationship with the national SO as the industry planning body. A key decision will relate to local cost-reflectivity (in terms of financial support to alternative IMs); this depends to some extent on how the Network Rail RAB is split to facilitate horizontal separation.

The ORR

– Will need to develop clear information protocols to be able to maximise the benefits of comparators. Will be setting more price controls, and using comparative data (under separate ownership) for the first time with respect to OMRE. It will also have to develop price controls for the SO function. Effort will be required initially to develop additional licences (for IMs and the SO), and to split the RAB. It will need to place less reliance on international benchmarking, although it will still be a useful measure of absolute GB efficiency.

Industry supply chain

– Opportunity to work for different IMs—increases competition for a larger number of contracts. More opportunities for contracts mean that there is less likelihood of a feast–famine cycle in respect of maintenance and renewals contracts.

4.4.4 Transition process

There are two likely approaches, of which the latter is likely to preferable:

1. Forcible separation of Network Rail following CC inquiry/Act of Parliament.

2. Setting regional price controls along the boundaries of the regions established above, with the expectation that Network Rail will decide to separate out its OMRE businesses along these lines (as occurred in the case of gas DNs). The ORR would also develop a SO price control at this stage. Over time, the new structure would establish itself.

Timescales

– Under 1) a CC inquiry would need to follow due process—approximately 36 months from ORR Enterprise Act investigation. An Act of Parliament would be likely to expedite matters, and to lead to a more certain outcome than the CC process.

– Under 2) the process is likely to be faster (in terms of price control separation), and also may achieve the desired outcome if Network Rail decides to sell the SO and IM functions that have been separated in terms of economic regulation.

Costs

– Transition costs, at least under option 2) might be quite modest, assuming that Network Rail sells the separated businesses. There would be likely to be some loss of focus on delivering CP4 outputs efficiently in the interim period.

4.4.5 Evaluation

Table 4.5 Evaluation of horizontal separation by market

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	Positive, given strong incentives on SO	Low/medium	Currently no strong volume incentive on Network Rail; introduce to SO	Energy sector SO models
Cost efficiency	Positive	Medium/high	Comparative competition, innovation, new financing and contracting approaches; evidence on TOC mergers less positive	Gas DN separation evaluation, ITS Leeds report for the ORR on combining franchises
Investment coordination	Slight positive	Low	Negative impact from separation outweighed by TOC/IM alignment	Gas DN separation evaluation
End-user satisfaction	Positive, given strong incentives on SO	Low/medium	Increase in network availability and capacity utilisation	Energy sector SO models
Changes to interfaces	Interfaces to increase slightly	Limited	If increase in interfaces is handled by SO only, and it makes timetable development more efficient	Interviews, and models from other sectors
Practicality and implementation				
Costs of implementation	–	Low/medium	Depends on the approach adopted	Gas DN separation evaluation
Time taken to implement, and nature of change required	–	Low/medium	Depends on the approach adopted	Gas DN separation evaluation
System-level operation				
Impact on whole railway system	Slight positive	Low	SO provides national function; track to rolling stock shift less clear	Interviews
Implications for awkward interfaces	No impact	Low	Limited impact on awkward interfaces, although wheel/rail might be improved via mapping of TOCs onto IMs	Based on option construction
Customer and local needs				
Local responsiveness	Positive	Medium	Market-based IMs to be more locally focused and market-oriented	Interviews

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Government as customer of the 'social' network	Positive	Medium	Allows for explicit focus on different types of market, eg competitive markets versus 'PSO' routes, and allows costs and revenues of latter to be analysed separately	Based on option construction
Market impact				
Impact on competition	Positive (with respect to infrastructure)	Medium/low	Increase in comparative competition; impacts on-rail competition though	Gas DN separation evaluation
Impact on economies of scale, scope and density	Neutral	Limited	Some economies of scale may be lost but the impact is expected to be limited. Franchise mergers unlikely to generate economies of scale	ITS Leeds at PR08, ORR modelling
Impact on transaction costs	Expected to fall	Limited	If increase in interfaces is handled by SO only, and it makes TT development more efficient	Interviews, experience in energy sector

Source: Oxera/Arup.

4.4.6 Summary evaluation

This option is evaluated as being positive, subject to transition issues. Improvements are expected over and above the regional version. These incremental improvements are expected to emerge from better alignment of TOCs to the IMs, from greater cost-reflectivity, and from a better matching of revenues with costs across the business.

What would have to be believed for this option to be successful?

– That the increased interfaces would be outweighed by the efficiency potential from separation.

– That the transition from Network Rail to the new IMs/SO could be worked out smoothly, especially in relation to Network Rail debt, the recruitment of senior staff and employee relations.

– That effective incentives could be introduced on the SO with respect to performance, network availability and capacity utilisation.

– That merging TOCs can provide benefits over and above the loss of on-rail competition between operators.

– That comparative competition would deliver benefits of the order seen in gas DNs, despite Network Rail's reorganisation to facilitate benchmarking between delivery units in the counterfactual (and despite the presence of subsidy). It might be the case that the benefits are even greater than seen in the gas DNs context, given alignment with markets and TOCs.

– That retaining the planning function within the SO (mini-Network Rail) could be efficient and effective, and that such a body could command industry respect in this function.

What type of actions related to this option could be undertaken in the short term?

– Start considering how to alter TOCs along the lines suggested.

– Examine the availability of relevant accounting information and KPIs within Network Rail that could be used to support this analysis, and any possible links with Network Rail's recently announced transformation programme.

Evaluation and scale of impact

This option would reflect many of the issues associated with the regions-based approach to separation. These include: the need to create and manage interfaces associated with a number of IMs; the possibility of comparative competition (although of a different form to that associated with the regional horizontal option); as well as the opportunities afforded by this option for an incremental approach that could be undertaken in stages as a means to test issues (eg, internal reorganisation of Network Rail to accord with the new market-/route-based definitions as a precursor to deeper changes).

It is useful to reflect on the main differences between the markets-based approach to horizontal separation, and the regions-based approach.

– Regional IMs will have a mixed bag of routes—ranging from main line to suburban to branch—and will therefore have a fairly generic approach to maintenance and renewals, without any consideration of differentiation of approach.

– Market-based IMs should be aligned to markets or type of route—eg, two or three (more than one for comparative benchmarking) for the strategic network (eg, the Highways Agency for the motorway and trunk road network)—this would give better alignment with the needs of these key routes which are generally multi-user and often electrified, and a number for the regional network with a real focus on differentiating part for a lower-cost

solution reflecting lower usage—similar to the approach used by BR Regional Railways—lighter rails, etc.

– These would also have better alignment with the government customer, with the two or three strategic IMs being more aligned to the DfT's National Networks team and the regional IMs focused on city regions, Scotland, Wales, and other local government customers. The SO could be established in the same way as under the regional horizontal separation option, although it would be particularly important for the IMs to retain responsibility for timetabling under this option.

It is expected that the benefits of this option would emerge slowly (the majority appearing one control period after implementation), although the process could be expedited through the introduction of shadow regulatory arrangements (accounting separation, incentivisation, etc), and the combining of franchises at replacement. Comparative competition between market-oriented 'Network Rails' would reveal the costs associated with different markets. The SO function delivers some efficiencies, as in the regional version.

Cost savings associated with franchise combination are studied in a confidential paper by Wheat and Smith (2010).[43] The paper finds that merging TOCs has led to efficiency improvements through the removal of management contracts, but that these benefits could have been achieved simply through re-franchising. Taking contractual changes into account, this suggests that merged TOCs can take advantage of economies of train density, but this may require some TOCs having overlapping services. No evidence of economies of scale from merged TOCs was found.

Overall, this option is viewed as being of **medium impact**. Net cost reductions of £100m to £500m are expected, depending on how changes are implemented. The net cost reductions associated with this option are at least as great as the benefits of comparative competition. Transition costs could be significant, but could be tested using a phased and incremental approach.

4.5 Vertical integration

Box 4.5 describes the key objectives of this option.

Box 4.5 Key objectives

Clearly realign incentives between wheel and rail, at least in the case of what are now franchised operators.

Improved planning and use of resources.

Management of the railway as an integrated system, particularly as technology shifts onto trains going forward.

4.5.1 Description

This option reintegrates the network at a market level. In this model, LSE, regional and inter-urban VI concessions (perhaps of 15–20 years' duration) are let by the DfT and regional governments, with open access and freight operators retaining existing access rights under enhanced non-discrimination obligations for the new companies. This option envisages up to seven or eight concessions (LSE, inter-urban, plus five–six regional, for which there would be a degree of alignment with regional planning functions).

The new concessionaires would lead industry planning and decision-making, obtaining funding, required outputs and fares policy from the DfT, Transport Scotland and other

[43] Wheat, P. and Smith, A (2010), 'Econometric evidence on train operating company size', Institute for Transport Studies.

funders on a similar five-year basis to today. The ORR's role at this point is to check that the business plans are consistent with the funding available.

A (very) shallow SO would offer system-level coordination only.

An alternative form of VI might be to reintegrate the entire industry at a national level. In this model, Network Rail would be merged with all franchised operators, operating on a 20-year private sector concession regulated by the ORR. All other aspects would be the same as for the market-level VI option. We have not evaluated this national VI option as it does not appear to offer any realistic chance of reducing industry net cost, or fulfilling other policy objectives, such as promoting competition. This is because it would constitute an extremely large, complex organisation which is expected to be difficult to regulate (not least because of the difficulties involved in establishing efficient costs and ascertaining the performance of different parts of the business), and could well compound a number of the problems of the existing regime (such as lack of responsiveness to local conditions).

4.5.2 Which incentive misalignments would this option assist with?

– The misaligned industry P&L is expected to be realigned to deliver improved investment coordination, resource allocation, and incentives to maximise industry benefits and minimise industry net cost.

– A lack of transparency surrounding the relative costs of providing different types of rail services, and misalignment with regional and other objectives.

– A lack of completely integrated planning horizons should be overcome under this approach, with input from funders at appropriate intervals.

– 'Meet but not beat' behaviour under the CLG model in the status quo.

4.5.3 Impact on players

Network Rail/IMs

Significant changes to Network Rail, which will be both reduced in size (according to markets, and with the removal of SO functions) but also augmented with additional functions as it becomes part of an integrated entity with operations extending across the value chain.

Franchised TOCs

Franchised TOCs will not exist as under today's arrangements. Passenger train operations will be absorbed into vertical structures.

Freight operators

Rules would be required to ensure that the VI players did not discriminate in their own interests against freight operators. Freight operators would have to interface with the SO, which would have to be sufficiently empowered to ensure that access across the VI businesses remained feasible for freight operators.

ROSCOs

The rolling stock market would remain as is, but would have to adapt to new industry structures. This may prove beneficial, with the rolling stock network compatibility issues that occur in the status quo being removed. Over time, the concessionaires might decide to procure their own rolling stock, or alternatively bids for concessions would include bank financing of rolling stock.

Funders

Funders would establish available funding, required outputs and fares policy on a similar five-year basis to today. Upfront funding might be required to deliver this option, given the need to buy out existing franchised operators.

The ORR

The ORR role will alter significantly. It will have important responsibilities with respect to ensuring non-discrimination by the VI players, and evaluating concessionaires' business plans for consistency with available funding. It may also widen the use of comparative benchmarking to compare the VI organisations.

Industry supply chain

VI players are likely to outsource a proportion of infrastructure work similar to that undertaken by Network Rail under existing arrangements. Multiple infrastructure companies are likely to want to test multiple in- and outsourcing arrangements.

4.5.4 Transition process

– The existing franchised operators will need to be bought out based on the NPV of the remainder of their franchises; an alternative would involve TOCs buying out relevant sections of the network from Network Rail (in practice, the TOCs might be too small to do this). In the latter case, clear information on the state of the assets would need to be provided.

– Alternatively, the transition process could be done incrementally as TOCs come up for franchise renewal, and OMR activities could be bundled into the new franchise. If implemented in one stage, the transition costs of this option would be significant.

– A new regulatory framework will need to be developed to reflect VI, and also the need for strong non-discrimination clauses. A substantial task will be ensuring consistent information flow to the ORR to make certain that it can carry out its tasks robustly. The need to let vertically integrated concessions under this option will require either renationalisation of Network Rail, followed by privatisation, or a complex renegotiation of existing financing and governance arrangements alongside (initially) separate price controls for Network Rail at the level of the preferred geography.

– Funding that currently goes to Network Rail might need to move to regional bodies to fund regional elements of the network, once clear costing data has been obtained.

Timescales

– Transition is likely to be time-consuming under this option, with implications for Network Rail's willingness and ability to meet efficiency targets within CP4. Disruption would also occur to franchised operators as they experience uncertainty about the remainder of their franchises. Overall, the process is likely to last into CP5, with benefits taking time to materialise.

Costs

– The costs of setting up this option are expected to be high, with decisions to be taken as to whether to parcel off Network Rail assets to existing operators, or to merge TOCs with Network Rail and then separate the resulting entity. Network Rail debt would have to be taken on by the new companies, or taken into the public sector as a precursor to sale of VI concessions, while its CLG structure would have to be untangled. Moreover, TOCs' expected margins over the remaining life of their franchises would need to be bought out (unless infrastructure were to be added onto existing franchises). These costs will be particularly high if implemented in one stage.

– Costs of the transition to this option could be reduced somewhat by undertaking a staged, evolutionary approach. This could be given effect by, for example, requiring Network Rail accounting separation along the lines of the ultimate form of integration, and introducing VI on a piecemeal basis on franchise renewal (eg, let as a 20-year concession).

4.5.5 Evaluation

Table 4.6 Evaluation of vertical integration

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	Positive	Medium	Depends on effective monopoly regulation. Risk of loss of specialisation	Interviews
Cost efficiency	Positive	Medium	Depends on effective monopoly regulation	Interviews, literature on rail and energy market models
Investment coordination	Positive	High	Expect this to be a strong rationale for this option	Interviews, literature on rail and energy market models
End-user satisfaction	Positive	Medium	Depends on effective monopoly regulation. Risk of loss of specialisation	Interviews
Changes to interfaces	Positive	Low (in overall terms)	Should reduce (but not eliminate) interfaces	Literature on rail and energy market models
Practicality and implementation				
Costs of implementation	–	Low to high	Have to buy out the TOCs, disaggregate Network Rail data and sell the concessions. Scale of impact depends on form of implementation	Based on option construction
Time taken to implement, and nature of change required	–	Low to high	Transition expected to last into CP5. Primary legislation may be required. Scale of impact depends on form of implementation	Based on option construction
System-level operation				
Impact on whole railway system	Positive	High	Key design feature of this option	Based on option construction
Implications for awkward interfaces	Positive	High	Key design feature of this option, although some 'internal' interfaces will require management	Based on option construction
Customer and local needs				
Local responsiveness	Slight positive	Low	Some degree of local focus from local VI players	Based on option construction

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Government as customer of the 'social' network	Slight positive	Low	Regional VI players will make costs more transparent, and funders better customers	Based on option construction
Market impact				
Impact on competition	Negative	Medium/High	Depends on effective safeguards against discrimination, but considerable concerns remain. Expect some benefits from competition between VI players	Literature on rail and energy market models
Impact on economies of scale, scope and density	Positive	Low	Some amount of scope economies	Other railways evidence
Impact on transaction costs	Slight positive	Low	Smaller number of transaction costs within VI entities, but some cross-market interfaces will remain to some degree. Interfaces with SO function	Based on option construction

Source: Oxera/Arup.

4.5.6 Summary evaluation

Assuming strong monopoly regulation, this option delivers against most objectives. However, evidence from other sectors and other railways points the other way—towards separation in order to drive competition, and that there are considerable risks associated with discrimination by VI players against downstream competitors. Transition duration and costs are likely to be substantial, with transition on multi-user routes likely to be most problematic. It is noted that some TOCs have limited expectations of earning margins, so walking away with compensation could be attractive.

What would have to be believed for this option to be successful?

– That VI is the best way to deliver incentive realignment and investment coordination.

– That transition costs are going to be heavily exceeded by the long-term benefits.

– That the evidence from other sectors on substantial benefits from separation does not apply in the case of rail. It depends on the extent to which competition for VI concessions might alleviate the apparent disbenefits of this option.

– That strong monopoly regulation can outweigh issues of discrimination (such as foreclosure effected through the shifting of costs onto the infrastructure business) and loss of competition associated with the option.

What type of actions related to this option could be undertaken in the short term?

– Consider early initiatives to test the VI approach, perhaps in Scotland and (potentially) Greater Anglia.

– Work to consider the most appropriate approach to transition—merge Network Rail with TOCs (buying out the operators), or allocate TOCs parts of the network.

Evaluation and scale of impact

A key issue with this option is the scale of transition costs. VI would involve fundamental changes at all levels of the industry, with significant implications for contracts, management structures, and organisational structures. The costs and complexities would be comparable with, if not necessarily on the same scale, as those associated with the original separation (and privatisation) of the rail industry in the 1990s.

Experience of industry restructuring in other sectors, particularly the energy sector, was associated with large costs (billions in the case of electricity in the UK), but the reforms are still identified as having had a net benefit overall. A key issue is therefore the extent to which the benefits of VI can overcome what could be significant transition costs (the size of these costs depending on the incrementality of the implementation). Benefits are expected to accrue from improved investment coordination, better transparency of the cost of providing the social railway, the removal of formal interfaces, operation of the railway as a system, and competition for concessions. Further analysis on why rail might be 'different' in this regard would help to further understand the net overall impact of VI options.[44]

However, in addition to transition costs, these benefits will need to be weighed against considerable costs associated with loss of competition, concerns over incentives for discrimination, and a potential loss of specialisation in passenger operations and customer service.

[44] Gomez-Ibanez makes the case as to why rail differs from network utility industries, such as electricity and telecommunications, by analysing, among other factors, the proportion of industry costs that are in potentially competitive activities, and the nature of the services provided. See Gomez-Ibanez, J. (2003), 'Regulating coordination: British Railroads' chapter in *Regulating Infrastructure: Monopoly, Contracts, and Discretion*, Harvard University Press.

Overall, this option is viewed as being of **medium impact**. Net cost reductions—driven by incentive realignment, but tempered by losses of specialisation and competition, and discrimination concerns—of over £100m per annum are expected, but only after what could well be considerable transition time and cost. The overall impact is made of a number of offsetting factors.

- Many of the benefits from franchise reform (option 1b) and from variants of horizontal separation are expected to be present in this option.
- In addition to these benefits, greater incentive alignment between infrastructure and operations would occur via better investment coordination, enhanced TOC revenues, and greater efficiency on the infrastructure side.
- There are costs associated with loss of competition. These relate to the following risks.
  - That competition for the VI concessions will diminish relative to franchise competition, since with a 15-year franchise/concession, the likelihood that the incumbent would hang onto it is considerable
  - There may be weak pressures to reduce cost, and/or to share any benefits with users and taxpayers;
  - That downstream competition will be eliminated in the case of passengers, and made more difficult in the case of freight. It is noted that considerable benefits from downstream competition are expected in the England and Wales water sector from vertical separation.
- While some of these benefits would remain (eg, through freight competition), and there would be competition at the margins between routes (eg, Birmingham to London), and (more importantly) some of the benefits would accrue to government through the sale of the concessions, significant costs might still be expected.
- There are also costs assumed with loss of focus on the end-user, unless strong regulation is brought to bear.

Transition costs would depend on the approach to approach to buying out TOCs. Overall, the scale of the net cost impact associated with this option depends particularly closely on effective monopoly regulation and on an effective change control mechanism.

4.6 Alternative forms of public sector involvement—'Railways Agency'

Box 4.6 sets out the key objectives of this option.

Box 4.6 Key objectives

Separation of infrastructure ownership (public sector) from management and operation (private sector), reflecting long-term nature of ownership and ensuring that HLOS development is consistent with this long-term perspective.

Introduce network contestability through the sale of long-term concessions for the management of the network.

Create opportunities for infrastructure maintenance and operations to be combined with train operations (sub-option).

4.6.1 Description

This option includes one possible form of public-sector ownership—namely the ownership of the network transferring into the public sector, which would then sell infrastructure concessions. Specifically:

– The ownership of the infrastructure network would transfer into the public sector.

– A 'Railways Agency' would be created as a government agency of the DfT, letting both infrastructure concessions and passenger franchises (the latter as the DfT does currently).

– Enhancement projects would be specified by the train operators on a DBOT basis (this would have to occur in conjunction with longer franchises). The Railways Agency would sponsor enhancements affecting multiple operators.

– Some combination of infrastructure maintenance and operation with train operation is envisaged as a sub-option.

– The ORR would ensure cost-effectiveness in let contracts, albeit on a 'lighter-touch' basis consistent with access charging Directive 2001/14.

– Long-term planning of the network would be undertaken by the Railways Agency with the HLOS and SoFA being the (only) responsibilities of the DfT and Transport Scotland.

4.6.2 Which incentive misalignments would this option assist with?

– Limited contestability for Network Rail's activities, and limited scope for alternative contracting approaches with respect to maintenance, renewals and enhancement.

– Debt-only funding of Network Rail; this model allows a multitude of financing approaches to be taken to each concession.

– Lack of incentives for investment coordination, and responsiveness of investment to end-user needs.

– Improvements in system operation might emerge under the sub-option of train operators being able to bid to perform maintenance and infrastructure operation.

4.6.3 Impact on players

Network Rail/IMs

The business of infrastructure management would change significantly. In effect, Network Rail would be subsumed into the Railways Agency, which would let railway infrastructure concessions.

Franchised TOCs

TOCs would be permitted to take some responsibility for the specification, design and delivery of enhancements, and also for infrastructure maintenance through the franchise agreement. They would interface with multiple infrastructure concessionaires.

Freight operators

These would continue largely as at present, although FOCs would have to interface with multiple infrastructure concessionaires. The concept of FOC-led enhancements, along the lines of that proposed with respect to TOCs (and over and above what takes place already in relation to freight terminals), is also conceivable in the context of this option.

ROSCOs

The number of interfaces confronting ROSCOs would increase due to multiple infrastructure concessions.

Funders

The role of the funders would be limited to the SoFA process, and providing high-level guidance to the Railways Agency in respect of desirable outputs, much as DfT currently interacts with the Highways Agency.

The ORR

The ORR would need to regulate the access charges of the new IMs. A key task would be to ensure that regulation adjusts to reflect a more competitive environment with respect to infrastructure, and would retain many of its existing responsibilities. Note that the Railways Agency could not discharge many of the ORR's existing functions (such as licensing, contracts regulation, etc) if it was a signatory to all of the relevant contracts.

Industry supply chain

Activities undertaken in-house by Network Rail will revert to the industry supply chain, with the Agency contracting out all activities to external parties.

4.6.4 Transition process

– Government to assume Network Rail debt; ownership would transfer to the public sector in a similar way as occurred with LCR. Thereafter, businesses would be created for sale as separate concessions (sequentially), including obtaining state aid clearance, developing separate price controls, and devising a clear delineation of responsibilities for asset stewardship. Ensure continuity of CP4 activities by Network Rail under public ownership.

– As franchises terminate, include in specifications an 'enhancements budget' that the successful operator would be free to apply to improve services in response to market needs.

– Design new regulatory framework for the concessions, ensuring maximum access to information for ORR to enable comparative competition. Expenditure drawn from franchise enhancement budgets would need to be monitored to ensure money is spent efficiently and effectively.

Timescales

– Completion of the transition of this option is expected to last into CP5, requiring taking Network Rail into public ownership, the preparation of the infrastructure concessions for sale, followed by a sequential sale of the businesses. Elements of the option could be incorporated more quickly, including offering franchisees the chance to take on infrastructure maintenance and operations contracts, and the use of enhancements budgets.

Costs

– As with other structural alternatives, there is cost associated with unravelling Network Rail's governance and debt financing. In this option, the additional cost of the concession sales needs to be factored in.

4.6.5 Evaluation

Table 4.7 Evaluation of 'Railways Agency'

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	Positive	Low/medium	TOC enhancement budgets should assist; new IMs could receive stronger and more targeted volume incentives	Interviews
Cost efficiency	Positive	Low/medium (negative in the short term)	Some incentives on new IMs to deliver efficiency once created	Gas DNs sale
Investment coordination	Positive	Low/medium	Some impact expected due to TOC-led enhancements	Interviews
End-user satisfaction	Positive	Low/medium	TOC enhancement budgets should assist; new IMs could receive stronger volume incentives	Interviews
Changes to interfaces	Increase	Medium	Operators interface with multiple IMs	Gas DNs sale
Practicality and implementation				
Costs of implementation	–	Medium	Cost of taking Network Rail into public ownership and subsequent concession sales	LCR sale
Time taken to implement, and nature of change required	–	High	LCR has demonstrated that this is achievable, but very time-consuming. Primary legislation may be required	LCR sale
System-level operation				
Impact on whole railway system	Positive	Low	Some detriment due to multiple IMs, but TOC-led enhancements should assist	Interviews
Implications for awkward interfaces	Neutral	–	Most interfaces remain	Based on option construction

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Customer and local needs				
Local responsiveness	Slight positive	Low	Regional concessions might assist, along with TOC-led enhancements	Based on option construction
Government as customer of the 'social' network	Slight positive	Low	Railways Agency may allow for more direct articulation of needs	Based on option construction
Market impact				
Impact on competition	Positive	Medium	Competition for concessions, and comparative competition. Potential risk of discrimination if joint infrastructure and train operation occurs in the sub-option	Gas DNs sale
Impact on economies of scale, scope and density	Slight negative	Low	Some loss of scale economies through breakup of Network Rail	ITS Leeds PR08
Impact on transaction costs	Transaction costs to rise reflecting new IMs	Low	Increase in interfaces results in greater transaction costs	–

Source: Oxera/Arup.

4.6.6 Summary evaluation

We expect this option to have a considerable impact on net cost through stronger cost-efficiency incentives on the IM concessionaires. There would, however, be a negative short-term impact, both due to the expense of taking Network Rail into public ownership pre-sale, and as incentives on Network Rail are likely to be weak ahead of break-up (and public ownership). Benefits are also expected to accrue from TOCs leading the majority of enhancements.

An important set of differences between this option and those based on VI and horizontal separation is that this option places franchising, concessioning and regulation of IMs in the public sector.

What would have to be believed for this option to be successful?

– That the cost of taking Network Rail into public ownership in the short term, the creation of a Railways Agency, and letting infrastructure concessions, are worth it in light of efficiency savings expected in future control periods from the existence of the concessions.

– That TOCs (and FOCs) are best placed to allocate and specify the majority of enhancements to the network.

– Sub option: that TOCs undertaking O&M work on the infrastructure would deliver efficiency savings and better system operation.

What type of actions related to this option could be undertaken in the short term?

– Reclassify Network Rail debt as on the government's balance sheet as the precursor to full, short-term public ownership.

– Transfer responsibility for allocating and specifying enhancements post CP4 to TOCs.

Evaluation and net impact

The contestability that this option brings—in terms of contestability for Network Rail's activities (operations, maintenance, renewals and enhancements)—is expected to offer considerable net cost savings via improved efficiency incentives. TOC-led enhancements are expected to improve investment coordination, passenger satisfaction and revenue growth, while the latter is also expected to be influenced by competition among IMs.

On the negative side, this option is expected to lead to increases in interfaces, as a result of multiple contracting approaches and multiple IMs. Incentives across these interfaces will need to be designed carefully, and transaction costs are expected to increase. More problematic are the considerable costs and duration of transition associated with this option.

Overall, this option is viewed as being of **medium impact**, potentially offering savings of £100m–£500m per annum in net cost, although many of its benefits are limited by the long transition period (during which time disruption and inefficiency are expected) and the relatively high cost of implementation. The transition costs will also include losses associated with the loss of CP4 efficiencies. Another driver of the difference in the estimated cost impact of this option vis-à-vis VI is the greater extent of public ownership under this model.

4.7 Greater public sector involvement—nationalisation

Box 4.7 below sets out the key objectives of this option.

Box 4.7 Key objectives

Full integration of the majority of rail businesses under state control.

Close involvement of the state in decision-making.

4.7.1 Description

There are a number of possible structures for a nationalisation option. For the purposes of evaluation, we have assumed a structure along the following lines.

- A single state-controlled OpCo (operating company) (or possibly two: Great Britain excluding Scotland, and Scotland only), operating all services that are currently franchised.
- A single state-controlled IM (or, again, possibly two), with broadly the same responsibilities as those that Network Rail has currently. Note that the OpCo and IM could be subsidiaries of a single state railways corporation as long as accounting separation is maintained, and there were contracts in place between the state and the OpCo in relation to the provision of public service obligations.
- Open access passenger operations continue as at present.
- Freight continues as an open access private sector operation as at present. The OpCo could move into the freight market subject to competition regulation by the ORR.
- Rolling stock remains initially with the ROSCOs; however, in the future, the OpCo could decide to buy its own rolling stock or to buy out the residual value on existing fleets.
- Stations would transfer in full to the OpCo business, including existing managed stations.

4.7.2 Which incentive misalignments would this option assist with?

- Disconnects in the industry planning process caused by multiple stakeholders, since this option would provide a single, unified process.
- Incentive misalignment across the wheel–rail interface, as this option brings together the industry P&L (operator revenues and infrastructure costs)—although the OpCo and IM are nominally separate, they have the same shareholder.
- This option provides for an alternative model for station ownership, management and operation—namely a full transfer of the station portfolio into the OpCo business.

4.7.3 Impact on players

Network Rail/IMs

Network Rail would transition across to being a state-owned body with broadly similar functions as today, although some functions (in particular station management) would transfer between IM and OpCo. The option remains for the Scottish IM to be separated out from the rest of the organisation, going beyond existing price control separation.

Franchised TOCs

Passenger services would cease to be franchised and would be operated by the state-owned operating companies. Existing TOCs would need to be bought out of their remaining franchise term. The new OpCo would still need to interface with the IM organisation, and would take on the station portfolio with an associated need to develop or buy in sufficient competencies to play this role.

Freight operators

Freight operators would remain in the private sector. Access to the network would be via regulated track access contracts with the state-owned infrastructure companies. There would be concern that this arrangement would lead to restrictions of access to the infrastructure for freight (and open access) operators, given incentives on the integrated organisation to prioritise its own traffic (particularly if the new OpCo were to start its own freight business).

ROSCOs

Initially ROSCOs could continue to provide rolling stock. However, the state-owned operators could choose over time to procure their own stock, leading to a diminishing role for ROSCOs.

Funders

The majority of the railway would be directly funded by government, the farebox and freight users. Taxpayers bear equity risk.

The ORR

The ORR would retain all its current roles but some in a much reduced form—eg, there would be many fewer track access contracts than at present.

Industry supply chain

Current suppliers to Network Rail would initially see relatively little change as they would still face a monopsony buyer. Current TOC suppliers would see their buyer base reduced and, potentially, a reduction in demand as the OpCo took advantages of any economies of scale and, potentially, took some activities in-house. Inevitably there would be changes in procurement policy moving forward, with the new businesses subject to public procurement law requirements.

4.7.4 Transition process

There would be two main transition processes required.

– Network Rail would be transferred into state control, with associated changes in licensing and contractual arrangements, and its debt being added to the public sector balance sheet.

– The OpCo would be established and take control of TOCs, buying out residual franchise value.

There may also be some transfer of responsibilities between OpCo and IM, most significantly for stations.

Timescales

– The transfer of Network Rail into public sector ownership involves debt being assumed by the public sector, and the winding up of the CLG structure. It is beyond the scope of this study to consider how long the latter element would take, but we assume for now the entire process would take up to three years. A similar duration is expected for buying out existing franchises.

Costs

– Transition costs are mainly driven by buying out expected TOC margins over the life of their franchises. There will inevitably be considerable costs associated with the creation of the new public sector businesses, driven not least by uncertainty caused by the process itself and its impact on operator incentives in particular.

4.7.5 Evaluation

Table 4.8 Evaluation of nationalisation

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
Incentive alignment				
Revenue growth	-	-	Would depend on state objectives, but some evidence exists that suggests franchising has had revenue benefits (eg, Arup report on 'International Review of Service Delivery') and work by Smith and Wheat	Arup (2010)[45] study on international service delivery, and Smith and Wheat (2009)[46]
Cost efficiency	Negative	Medium to high	Creation of public sector monopolies expected to lead to loss of competition (and X-inefficiency); public sector ownership also expected to contribute	Economic theories of X-inefficiency and efficiency incentives under public ownership
Investment coordination	Positive	Medium	Single organisation controlling investment	Based on option construction
End-user satisfaction	Neutral	None	British Rail had lower user satisfaction than privatised companies do now, but no a priori reason to assume that a new company would have similar issues	
Changes to interfaces	Positive	Medium	Removes a significant number of interfaces	Based on option construction
Practicality and implementation				
Costs of implementation	-	High	Would need to buy out TOCs and transfer Network Rail debt into the public sector	
Time taken to implement, and nature of change required	-	Medium	Main issue would be buying out TOCs and removal of Network Rail's CLG status. Primary legislation required	

[45] Arup (2010), 'International Review of Service Delivery', January.
[46] Smith, A. and Wheat, P. (2009), 'The effects of competition for the market on productivity and efficiency: Evidence from the passenger rail sector in Britain', ITS Leeds.

Criteria	Assessment	Scale of impact	Reasoning	Main sources of supporting evidence
System-level operation				
Impact on whole railway system	Positive	High	Structured as a single system	Based on option construction
Implications for awkward interfaces	Slight positive	Low	Most interfaces remain, although it is expected problem areas would be resolved more easily. Stations component beneficial in this regard	Based on option construction
Customer and local needs				
Local responsiveness	Negative	Low/medium	Creation of single operating company will be detrimental	Based on option construction
Government as customer of the 'social' network	Slight positive	Low	Changes role of Government from customer to shareholder; EU requirements necessitate explicit contracting for PSO services	Based on option construction
Market impact				
Impact on competition	Negative	High	Removal of competition between operators, both for franchises and on rail (between franchises)	Based on option construction
Impact on economies of scale, scope and density	Positive	Low	Economies of scope expected to emerge from the common public ownership of OpCo and the IM. Rolling stock also easier to allocate	Based on option construction
Impact on transaction costs	Positive	Medium	Reduces, but does not eliminate the impact of current interfaces	–

Source: Oxera/Arup.

4.7.6 Summary evaluation

While this option has its merits—mainly relating to common ownership of operations and infrastructure and the expected benefits for the operation of the railway as a system—it is clear that the loss of competition associated with merging operators, and inefficiency associated with public sector ownership would be detrimental. In addition, public sector ownership risks losing the funding certainty that the industry has enjoyed in recent years. Moreover, this option involves high implementation costs, and the overall evaluation is negative.

What would have to be believed for this option to be successful?

– That state ownership is the only way of achieving an effective structure with minimisation of misaligned incentives.

– That required levels of funding will be available, and will not be drastically affected by changes in the government's fiscal position.

– That any structure will allow for government policy to be reflected, and to be sufficiently flexible to adapt to changes in government and/or changes in policy priorities.

What type of actions related to this option could be undertaken in the short term?

– Reclassify Network Rail debt as a precursor to public sector ownership.

– Over time, develop legislation enabling the buying out of TOCs and the removal of Network Rail's CLG status.

Evaluation and net impact

The creation of a public sector monopoly covering operations and infrastructure brings with it considerable risks. These relate in particular to the loss of competition between operators, leading to X-inefficiency,[47] and public ownership making it more difficult to make efficiency savings. For example, Pollitt and Smith (2002)[48] identified a decline of efficiency of 1% per annum under British Rail.[49]

Moreover, the industry currently enjoys relative funding certainty resulting from the Railways Act 2005, whereas this option risks funding being more subject to Comprehensive Spending Reviews (and even intra-review alterations).

Overall, this option is viewed as being of **negative impact**. The study team considers the option to deliver net cost increases, and high transition costs.

4.8 Conclusions: options for change

The above evaluation of options has provided the following conclusions.

– The franchise reform option offers a medium to high positive impact on net cost. Its main focus is on bridging the incentives gap between infrastructure cost and revenues from end-users, and also benefits from equity financing of Network Rail.

– The horizontal separation (markets) option has a medium to high positive impact on net cost. Its main focus is on incentivising infrastructure cost efficiency, and revealing the

[47] X-inefficiency captures the notion that there is a positive relationship between external pressures on a firm and the level of effort exerted by its employees. See, for example, Church, J. and Ware, R. (2000), *Industrial Organization*, McGraw-Hill, p. 145.
[48] Pollitt, M. and Smith, A. (2002), 'The restructuring and privatisation of British Rail: was it really that bad?', *Fiscal Studies*, **4**, pp. 463–502.
[49] Couto and Graham, in a study of 27 rail companies (including British Rail) from 1972 to 1999, find evidence of significant cost increases caused by inefficiency. Cuoto, A. and Graham, D. (2009), 'The determinants of efficiency and productivity in European railways', *Applied Economics*, **41**, pp. 2827–51.

costs of providing rail services in different markets, enabling more market-oriented interventions. The SO function also offers incentives for better use of the network. It is noted that horizontal separation may render benefits without needing to encompass the entire rail network.

– The horizontal separation (regions) option has a medium impact on net cost. It also focuses on incentivising infrastructure cost efficiency and better use of the network, but offers lower benefits compared with the market-oriented approach to horizontal separation.

– VI has a medium impact on net cost following a costly transition period, enabling incentive realignment between operations and infrastructure, but at a cost of lost competition.

– The Railways Agency option has a medium impact on net cost following a costly transition period. Its focus is on providing contestability with respect to infrastructure interventions.

– The incremental change option has a low impact on net cost. Its focus is on change within the regulatory and legislative status quo, incurring moderate costs, but enabling only moderate benefits.

– Nationalisation has a negative impact on net cost due to the loss of competition and the creation of a public sector monopoly.

Overall, our conclusion is that two options—franchise reform and a market-oriented horizontal separation of Network Rail—offer the strongest potential for savings in net cost. For the sake of clarity, we set out below the assumptions we have used to draw these conclusions.

It should be noted that, despite the focus of the two options being different, there is overlap between the benefits associated with the two options, such that they should **not** be regarded as entirely additive. Recall that the franchise reform option expects equity issuance by Network Rail towards the end of CP4, whereas the horizontal separation option expects the IMs to choose between financing approaches (so some of the benefits under horizontal separation include savings associated with stronger incentives from equity financing that are also seen in the franchise reform option).

It is also noted that a 'horses-for-courses' approach could be the most appropriate means of proceeding. A single, encompassing solution to a network as complex as the GB rail network arguably would not be in the best interests of users and taxpayers, and may not deliver reductions in net cost. Local solutions, and ones that recognise the differences between particular types of market (eg, intercity versus rural, 'PSO networks') could go a considerable way to improving (relative to the status quo) the alignment of incentives across the industry's value chain.

Issues connected with stations could be addressed as part of overall reforms, or separately, given their somewhat distinct character. In practice, it seems that there would be benefits in reforming the contractual matrix at stations concurrently with other reforms that may themselves have an impact on station issues.

A roadmap for change

We are proposing a roadmap for implementation, starting from the short- to medium-term changes, before moving onto the more radical options for medium-term change for completion by CP5.

– Stations: we propose that the forthcoming franchise re-letting exercises offer opportunities to test the market in relation to alternative station management models.

These would focus in particular on 99-year full repairing leases for the franchisee, with an increase in Network Rail's managed stations portfolio (see the franchise reform option); and Network Rail's preferred model, whereby it focuses on station fabric, with TOCs taking responsibility for customer-facing assets.

– Fares regulation: as per option 1a, the suggested move towards economic regulation within the DfT and Transport Scotland fares policy can be started immediately, for implementation in 2011.

– Franchise reform: implemented via franchise change for TOCs with over two years to run, and the rest changed at re-letting. An evolution through the first few franchises would ease implementation, and enable the detail of risk-sharing arrangements to be developed. A licence change would be required to enable Network Rail revenue sharing.

– Horizontal separation: begin the process of separating price controls immediately. In terms of implementation, one option would be short-term nationalisation, followed by restructuring to split out SO and IMs, then concession. An alternative would be for Network Rail to carry this out itself via the sales of separated businesses, as happened in the gas DNs case. Our expectation is that this could be implemented for 2014, and available for examination during PR13, although further work to develop the detail would be required.

A1 Incentives and industry structure

A1.1 Incentives and net cost

The possibility that incentives can affect net cost was introduced in section 1 of the main body of the report. This section briefly describes three aspects of incentives that are particularly important in attempting to understand industry net cost.

Private or asymmetric information

- Individuals or organisations have private information—alternatively known as asymmetric information—when they know something of which others are unaware. For example, in the case of rail, Network Rail will have more information than its regulator on the costs required to undertake a particular piece of renewal.

- On the one hand, private information exacerbates the friction between regulators and firms—it increases the scope for hidden actions that firms can take to maximise their profits at the expense of social welfare, hence the need for incentives and incentive schemes capable of aligning diverging interests. On the other hand, private information can be harnessed by incentive schemes by regulators willing to offer financial rewards to firms prepared to supply private information.

- The more significant the extent of private information, the more relevant the scope for incentive schemes. This observation follows on from the role of incentives: their goal is to align firms' interests to those of society, without resorting to direct intervention.

The 'power' of an incentive

- The 'power' of an incentive for a commercial, profit-oriented organisation (such as a TOC) is defined as the link, provided by the incentive, between the firm's prices and its cost or profit performance.

- More specifically, the power of incentive schemes is proportionate to the amount of profit to which the firm is made residual claimant—those schemes that leave the firm a large share of profits as a result of behaving in accordance with societal preferences are known as high-powered incentive schemes, whereas those that leave the agent a smaller share are known as low-powered incentive schemes.

    - For example, a franchised TOC may be confronted with low-powered incentives when faced with particular types of investment opportunities at the end of its franchise term.

    - As another example, the CLG status of Network Rail could weaken the general power of incentives that it faces, given the absence of shareholder pressure for dividends.

The temporal dimension of an incentive

- The temporal dimension of incentives is also important. In particular, frequent renegotiation of agreements between regulators and regulated firms may increase the scope for moral hazard[50] and adverse selection.[51]

[50] The agent takes one action that is not verifiable, but which affects the observable outcome.
[51] The agent holding private information benefits from misrepresenting it before the agreement is signed.

– Investment is, by its nature, intertemporal—involving costs and benefits occurring over a number of years—and has particular considerations in relation to incentives.

  – First, bargaining between two parties, each having private information about its willingness to enter into a contract, will be inefficient.[52] For example, consider a regulatory negotiation between a company and its regulator as part of a price review: the company will have an incentive to inflate its costs to extract larger returns from the investment; the regulator will have an incentive to keep the investment as low-cost as possible for the benefit of customers.

  – Second, the outcome of the bargaining may be that the investment is insufficient, or insufficiently specialised. Once the investment has been sunk, investing firms become the weaker party in the future negotiations with the regulator in relation to the returns on that investment. Foreseeing this, the investing firm might shun large or too specialised investments, in order to leave the door open to alternative options, thus improving its bargaining position.

  – For example, Network Rail may be reluctant to invest in capital programmes if it is uncertain regarding the rate at which the ORR will allow the assets to be remunerated at future price controls in the absence of regulatory commitment to allowed rates of return.

It is important to recognise the fundamental constraints that define how incentive (re)alignment might be realised.

– First, it is clear to policymakers that society benefits from cost-oriented prices and lower costs. The aim of policy in general is, therefore, to ensure that prevailing prices are cost-reflective, and that costs are as low as possible (given other relevant policy objectives, such as the maintenance of quality standards).

– Second, in the presence of private information, it is impossible to cope fully and simultaneously with both issues (ie, cost-reflective prices and prices that are as low as possible). In fact, one solution to above-cost prices is to expropriate firms' entire profits. It is also clear that letting firms appropriate all profits is, in most regulatory systems, the main driver of cost minimisation by firms.

This indicates the problems faced by regulators in attempting to balance a number of policy objectives with the fundamental concern of preventing monopoly abuses. The challenge of incentive design is to address relevant policy objectives in a context of private information, different levels of incentive power, and dynamic industry change.

A1.2 Industry structure and net cost

The fragmented nature of the rail industry means that there are a number of interfaces over which different organisations must negotiate and operate. There is an established body of theoretical and empirical literature concerned with industry structure, and debates around integration and various forms of separation. The literature suggests that industry structure can affect incentives and transaction costs in a number of ways, as follows.

– **Scale economies**: Vertical separation could lead to a reduction in the extent of scale economies by removing the ability of an integrated firm to exploit its size and achieve minimum efficient scale across its different functions—eg, Network Rail procurement of high-output track renewal trains could be subject to economies of scale.

[52] Laffont, J.J. and Tirole, J. (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge MA: The MIT Press, p. 99.

- **Scope economies**: Economies of scope in a vertically integrated firm may arise from a number of sources, including: enhancing the flow of information between different divisions within the company, and reducing transaction costs in the internal provision of inputs—eg, the wheel–rail interface.

- **Double marginalisation**: Vertically separated organisations both charge a mark-up above cost for their products, which reduces overall demand in the market.

  - In the case of the rail industry, this might suggest that a vertically integrated player would account for overall demand when setting charges to recover the costs of infrastructure and operations, rather than examine each in isolation.

- **Potential for investment hold-up**: Under separation, a firm supplying an input (such as a rail network) to another business faces the risk that this downstream business (ie, the company using this input) will not purchase the input after the investment has been made, which may lead to under-investment by the upstream firm—this is a potential issue in the case of the rolling stock market.

- **Coordination**: Coordination problems between different elements of the value chain may cause inefficiencies from losses in scope economies, as well as hold-up problems in investment and various transaction costs—this could be an issue between ROSCOs, TOCs and Network Rail (see the case study on track-friendly trains).

- **Impact on incentives to invest**: All of the above suggests that incentives to invest will be affected by industry structure, since changes in the extent to which different parties expect to be able to appropriate returns will alter the balance of investment along the value chain.

Some of these themes are explored in the literature review below, which reviews experiences with different models of market structure in rail markets.

A1.3 Assessments of market structure in rail markets

This section reviews literature analysing experiences with separation in European rail markets.

Vertical and horizontal separation

Cantos, Pastor and Serrano (2010)[53] analyse the effects of the structural reforms of European railways on efficiency, productivity and technical development. They find that the reforms overall were beneficial, in particular for industries where vertical separation was combined with new entrants in the freight sector. This highlights that the success of the reforms is dependent on how successful they are at introducing competition.

The paper uses non-parametric mathematical programming techniques on a sample of European railways over the period 1985–2005, and uses a productivity measure focusing not only on productivity, but also on the relation between productivity and efficiency. They find that countries that restructured their railways both vertically and horizontally observed the greatest improvements in efficiency levels and productivity indices, and had greater technical progress and productivity growth.

The authors decompose their findings to determine which structural change has the most dominant effect on outcomes, using dummy variables to reflect the different structural changes. The results show that vertical separation and the entry of new freight operators are

[53] Cantos, P., Pastor J.M. and Serrano, L. (2010), 'Vertical and Horizontal Separation in the European Railway Sector and its Effect on Productivity', *Journal of Transport Economics and Policy,* **44**:2*,* May.

associated with higher levels of efficiency and productivity; whereas the system of franchised passenger services does not have a significant effect on efficiency or productivity.

The failure of the franchised passenger system to have a significant effect on efficiency may be due to problems with the tendering process, such as the occurrence of predatory bidding, low numbers of bidders or future uncertainty. The results show that countries that introduced a system allowing the entry of new operators, in addition to vertical separation, had the highest levels of efficiency and productivity; although countries that introduced vertical separation alongside a system such as competitive tendering did foster some technical progress. Further, the results suggest that it is more beneficial to introduce horizontal separation alongside vertical separation, rather than horizontal separation on its own.

However, the authors highlight that their results are based on relatively recent data, due to the recent nature of many of the reforms. As such, whether these findings are reflected in future data should be a focus of future research.

In contrast to these results, Driessen, Lijesen and Mulder (2006)[54] find that competition for the market (competitive tendering) encourages efficiency more than competition in the market (open access). They explain this finding by suggesting that in the case of open access, operators have fewer passengers each so they do not benefit from economies of scope, subsequently keeping up costs. Further, they do not come to a decisive conclusion on the impact of vertical separation, but controversially find that it is not a necessary condition to achieving an increase in productive efficiency. However, differences in data, variables and estimation techniques may partly explain these conflicting results.

Welfare effects of vertical separation in Dutch Railways

Lijesen et al (2005)[55] use a partial efficiency analysis to determine the effects of the Dutch railways re-structuring in the early 1990s on efficiency and productivity outcomes in the industry. Their results highlight the differences in the delivery of horizontal separation.

They find that the reforms had an insignificant effect on the productivity and efficiency of franchised passenger services. Overall, despite the reforms, passenger transport had difficulties in reaching historical performance levels. The limited effect of the reforms on efficiency may be as a result of the government policy to award NS, the successor of the incumbent rail operator, the concession to run the mainline services until 2015. This greatly curtailed competition in the market.

These results are supported by the findings of Cantos, Pastor and Serrano (2010), who show that passenger franchising has a statistically insignificant effect on various efficiency and productivity outcomes. In contrast to passenger services, the reforms led to increased efficiency and performance as a result of the introduction of competition in the freight market; the former monopolist operator, Railion, losing 20% of market share as a result of the entry of new operators.

They find that this increased competition has improved Railion's productivity. Labour productivity measured in physical units improved significantly post-reform as a result of increased competitive pressure on tariffs and profit margins.

Again, Cantos et al (2010) offer support for these findings. They find that the entry of new freight operators has significantly improved efficiency levels.

[54] Driessen, G., Lijesen, M. and Mulder, M. (2006), 'The impact of Competition on Productive Efficiency in European Railways', CPB Discussion paper, Number 71, September.
[55] Lijesen, M., Mulder , M. and Driessen, G. (2005), 'Welfare effect of vertical separation in the Dutch railways', Netherlands Bureau for Economic Policy Analysis, CPB, The Netherlands.

Separation and privatisation in Great Britain

Pollitt and Smith (2001)[56] use social cost–benefit analysis (CBA) to assess whether the re-structuring and privatisation of the British rail industry produced savings in operating costs in the period post-privatisation and up to the Hatfield incident and Railtrack's entry into administration.

They find that the reforms achieved significant efficiency savings compared with the counterfactual of continued vertically integrated public ownership—the industry achieved efficiency savings of 13% (2% per annum) and unit costs fell by 17% (2.7% per annum) over the period.

Overall, they calculate that the reforms generated efficiency savings of £800m over the period. This figure is far outweighed by the restructuring costs, but assuming continued savings into the future, they estimate that the industry reforms should generate efficiency savings of £1.1 billion net of restructuring costs. Further, they estimate that consumers have achieved benefits of £1.2 billion.

They find that output quality also increased slightly post-reform, as train performance (measured as delay per passenger train) improved significantly compared with the counterfactual, where performance may even have deteriorated. Overcrowding got worse, although this may have occurred under the counterfactual due to the rapid increase in passenger numbers. The reforms had an insignificant impact on safety. Although the quality of the track decreased post-privatisation, with an increase in the number of broken rails, there is little evidence to suggest that this would have been any different under the counterfactual.

A counterfactual of continued integrated public ownership was constructed to estimate the performance of the industry in the absence of the reforms. The authors reviewed economic data and historical cost data from British Rail's accounts, focusing on the five-year period prior to the reforms, 1988/89–1992/93 (allowing for a transition period after this date). They constructed a counterfactual cost profile based on the 1992/93 cost level, projected forward using an assumption of annual counterfactual cost savings (based on historical performance and data from other UK privatised industries). The main results are based on an assumption of annual counterfactual cost savings of 1%. The authors suggest that this may be an overestimation, as unit costs in fact went up by approximately 1% per annum in the five years prior to privatisation, and as such, the actual efficiency savings of the reforms may be even greater than stated. It is important to recognise, however, that the estimated savings are very sensitive to the counterfactual cost assumption.

Of course, the longevity of these findings is put into doubt given the Hatfield accident, and subsequent considerable cost increases in the sector.

Vertical integration and competition

Pittman[57] has studied the issue of VI in railways, and uses the example of freight operators to highlight how often policy makers overlook the potential benefits that a system of VI can have over separation in the railway industry. He suggests that vertical separation between freight operators and infrastructure in the railway industry has not been as successful as the separation in other network industries and suggests reasons why this might be.

First, the relatively high network fixed and sunk costs as a share of operators' delivered service costs makes providing the correct incentives to operators for maintenance and improvements difficult. Second, the existence of economies of scale and density at the 'competitive' operations level means that the average costs of an operator decline as its

[56] Pollitt, M.G. and Smith, A. (2001), 'The Restructuring and Privatisation of British Rail: Was it really that bad?', DAE Working Paper 0118.
[57] Pittman, R. (2005), 'Structural Separation to Create Competition? The Case of Freight Railways', *Review of Network Economics,* **4**:3, pp. 181–96.

output increases. This may result in a 'first mover' advantage and the market may tend towards an uncompetitive structure. Third, strong economies of VI will be lost from separation due to the strong interdependence of freight operators and infrastructure owners at the point of vertical separation: the wheel–rail interface. Such problems are avoided within a vertically integrated market structure. Pittman uses the broad example of American railway reforms, where VI has been largely successful.

Pittman (2007)[58] suggests that the benefits of alternatives to VI rarely outweigh the losses from vertical separation. Pittman (2009)[59] further highlights the potential for competition to exist among vertically integrated providers, as this maintains whatever economies of VI that exist.

Economies of scope

Wetzel and Growitsch (2006)[60]analyse the effects of institutional separation and of economies of scope. They find evidence of efficiency advantages and economies of scope across many integrated European railways. Wetzel (2008)[61] finds no significant impact of separation on technical efficiency. It offers no conclusive result on whether the loss of economies of scope from separation outweighs the efficiency gains from increased competition.

Cantos Sanchez (2001)[62] analyses the relationship between infrastructure and operating costs in a number of European railways prior to vertical separation in the industries. The results indicate that there are important vertical relationships between infrastructure and operations. The paper highlights the potential high costs involved under a vertical separation; it could result in high inefficiencies from loss of economies of scope and loss of coordination effects. The paper finds that the costs deriving from passenger transport are independent of freight transport and are substitutable with respect to infrastructure costs. The paper does not attempt to estimate empirically the actual effect of separation on such costs.

Lijesen, Mulder and Driessen (2005)[63] find weak evidence for positive economies of scope in the Dutch rail industry between tracks and railway operation. This suggests that the resulting costs of separation would be limited as there would be insignificant diseconomies of scope. However, their results are limited as their data does not include capital costs, and they are unable to draw a clear distinction between economies of scope and economies of density in their findings.

Factors affecting operating company performance

Merkert, Smith and Nash (2009)[64] take advantage of the heterogeneity of structural organisation across European rail industries to carry out an efficiency analysis on passenger and freight train operators. The paper uses a cross-sectional sample of 43 Swedish, German and British TOCs for a single fiscal year.

Vertical separation is not found to have any significant effect on technical or cost efficiency. However, it is found to have a negative effect on the allocative efficiency of production staff and production material. Further, the results do not suggest that the positive effects of vertical separation in reducing operating costs compensate for the decrease in allocative

[58] Pittman, R. (2007), 'Options for re-structuring the state-owned monopoly railway', *Research in Transportation Economics,* **20**:1, pp. 179–98.
[59] Pittman, R. (2009), 'Railway Mergers and Railway Alliances: Competition Issues and Lessons for Other Network Industries', see http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1410132.
[60] Wetzel, H. and Growitsch, C. (2006), 'Economies of Scope in European Railways: An Efficiency Analysis', Working Paper No.29.
[61] Wetzel, H. (2008), 'European Railway Deregulation: The Influence of Regulatory and Environmental Conditions on Efficiency', Working Paper No.86.
[62] Cantos Sanchez, P. (2001), 'Vertical relationships for the European railway industry', *Transport Policy*, **8***,* pp. 77–83.
[63] Lijesen, Mulder and Driessen (2005), op. cit.
[64] Merkert, R., Smith, A. and Nash, C. (2009), 'The Effects of Institutional, Environmental and Transactional Factors on Train Operating Company Performance—a Tobit Regression Approach based on DEA Scores'.

efficiency. These results are supported by Merkert, Smith and Nash (2010),[65] who also find that vertical separation has no significant effect on technical efficiency.

One effect of vertical separation can be higher transaction costs. Both papers find that higher transaction costs significantly reduce technical, allocative and cost efficiency. In particular, their conservative estimate shows that increasing transaction costs by 100% (measured as the cost of staff with transaction occupations per train-km) reduces technical efficiency by 11.3%.[66] However, they emphasise that vertical separation is not the only source of transaction costs, and as such, it is not justifiable to conclude that as vertical separation increases transaction costs, it also reduces efficiency.

Merkert, Smith and Nash (2009) also provide findings on market structure. The results suggest that too much competition reduces technical and cost efficiency (through rapidly increasing transaction costs); however, at lower levels, it has a significant positive effect. They find that the most efficient operators face substantial competition. They also find that market concentration has a significant positive effect on technical and cost efficiency. However, they find that very large train operators become scale-inefficient with decreasing returns to scale.

Structuring of reforms

Friebel, Ivaldi and Vibes (2004)[67] highlight that the way in which reforms are structured and delivered will determine the extent of their effect. They use panel data from EU countries to evaluate the effects of the reforms on efficiency. Their key finding is that the effects of the reforms depend on sequencing; the introduction of multiple reforms in a 'package' has neutral effects, whereas sequential reforms improve efficiency.

They explain this by noting that sequential reforms allow a government to learn about the desirability of further reforms at intermediate stages and make amendments if necessary. However, the study has some limitations in that the data available meant that the authors could not control for heterogeneity across the reforms, or in their implementation. Second, they did not take account of the degree of subsidisation across European countries which may have implications for their efficiency measure.

Franchise Structure

Affuso and Newbury (2002)[68] suggest that unbundling and competition for franchises combined with commercial objectives can provide strong incentives towards better investment behaviour. They find that shorter contracts generate higher rates of investment as TOCs invest towards the end of their contract. This may be because TOCs facing re-procurement respond to threats of non-renewal of franchises from the regulator. Providing this result is robust, it suggests that higher degrees of competition through a frequent auction process due to short duration franchises may result in improved investment performance. The authors note that the results should be considered preliminary findings and there exists room for further research.

Their analysis only focuses on a period of three to five years from the start of operators' contracts, and is somewhat out of date. They find that TOCs tend to invest towards the end of their contract. The results only assess the behaviour of short-duration franchisees and it may be the case that the results of the investment behaviour cannot be interpreted as a marginal effect based on the length of duration.

[65] Merkert, R. Smith, A. and Nash, C. (2010), 'Benchmarking of train operating firms - a transaction cost efficiency analysis', *Transportation Planning and Technology,* **33**:1, February, pp. 35–53.
[66] Ibid., p. 50.
[67] Friebel, G. Ivaldi, M. and Vibes, C. (2004), 'Railway (De)Regulation: A European Efficiency Comparison', *Centre for Economic Policy Research,* Discussion Paper No.4139, March.
[68] Affuso, L. and Newbury, D. (2002), 'The impact of structural and contractual arrangements on a vertically separated railway', *The Economic and Social Review,* **33**:1, Spring, pp. 83–92.

Further, the results do not distinguish between the magnitude of an investment, only whether an investment was made or not; they do so on the basis that most investments are 'lumpy and jumpy' and the main decision is whether to invest or not, rather than how much. This however, limits the results as it considers a short-duration franchisee investment in a small transferable item towards the end of the contract to satisfy the regulator, equal to a larger specific and less-transferable investment by a long-duration franchisee, which may offer greater benefits. Despite this, however, the study may offer some insight into franchise design, conditional on franchises being of a short duration.

Summary

Due to the variety of data and analytical approaches applied, the picture that emerges from the above review is somewhat confusing. However, we have attempted to draw out common themes, albeit with the caveat that in the dynamic situation (in terms of liberalisation and industry structures) seen in rail markets across Europe over the past 20 years, evidence can often be out of date before it is even published.

– Evidence on economies of scope between infrastructure and operations is weak, suggesting that with well-aligned incentives across this interface, vertical separation should involve limited detriment.

– The introduction of competition is generally beneficial. However, this has generally been more successful in the freight market than in the passenger market, perhaps due to the difficulties of procuring PSO passenger rail services.

– Reforms should be undertaken sequentially, as opposed to in a package, to promote learning by government organisations.

This suggests no *prima facie* case for VI, and the importance of getting franchise specification, procurement and monitoring right for passenger services.

A2 Case studies: mapping of industry processes

This appendix describes key industry process case studies, with a view to establishing and analysing the interfaces and transaction costs that operate at the interfaces associated with these processes. Industry processes were discussed with stakeholders, and the following nine were selected for analysis:

– performance (delays, cancellations) improvement;
– possessions;
– station management, operation and upkeep;
– the specification, procurement and contract enforcement with respect to rail franchises;
– development of the timetable;
– the procurement and ongoing provision of rolling stock;
– long-term planning;
– the delivery of renewals and enhancements, including cross-industry projects;
– the allocation of capacity.

This work involved detailed research into legislation, regulatory documents and other relevant publications, supported by interviews with key industry players involved in delivering these processes within the current industry arrangements. Findings are presented below for each of these processes.

A2.1 Performance improvement

This section describes the systems and interfaces that are in place with respect to operational performance in the GB rail network. Measures of operational performance take several different forms and it is important to differentiate between:

– performance targets set by government (through the ORR) for Network Rail and the rail industry generally;
– the day-to-day performance regime which monitors and attributes delays and cancellations to the party found responsible for causing the delay incident.

The key elements of operational performance cover:

– measures of train performance (delays and cancellations);
– delay attribution;
– the PPM;
– JPIPs, involving both Network Rail and the TOCs.

A2.1.1 Objectives of the regime

Measures of operational performance were developed as part of the original rail privatisation legislation in 1993 and involve several parties at different levels of the rail industry. Central government (through the DfT and the ORR) will, for example, specify the PPM, and this will be used to measure the operational performance of Network Rail and the TOCs.

At the core of operational performance measurement is a ‘delay attribution’ process, whereby the performance of both passenger and freight operators is monitored with delays and cancellations attributed to either the operator or Network Rail. Financial compensation will then flow between the respective organisations, based on who is responsible for these delays and cancellations.

For passenger services, the Schedule 8 performance regime covers the regulatory framework for delay attribution. For freight services, there are also guidelines in the

performance regime to assess which party is responsible for delays and the financial implications of this. There were two original purposes of the performance regime.

- Compensation: 1) to compensate passenger train operators for revenue lost in the event of lateness and cancellations attributed to Network Rail or other train operators; 2) to compensate freight train operators for estimated costs and revenue loss from lateness and cancellations attributed to Network Rail or other train operators.

- Incentivisation: to incentivise train operators and Network Rail to improve operational performance through: 1) operational decision-making, and 2) investment appraisal.

Initially, Schedule 8 was a performance incentive scheme. Over time, however, the industry moved away from this incentive regime towards a compensation regime.

Schedule 8 is now complemented by a number of other incentives to improve performance. These include JPIPs, as discussed below.

A2.1.2 Features of the regime

The main features of the performance regime are as follows.

Payment rates (passenger)

- Based on the modelled impact of lateness and cancellations on train operators' revenue (the marginal revenue effect).

- Data for impact of lateness/cancellations based on industry demand studies.

Payment rates (freight)

- Based on a broad estimate of the cost and revenue impact of lateness and cancellations, but little information is currently available.

Benchmarking

- Reflects an expected level of performance.

- Set using historical performance.

- If both Network Rail and train operators perform at the benchmark level, no payments are due—this means that no additional funding to the infrastructure manager is needed to fund the regime.

Franchise agreements

- The level of expected performance (and associated Schedule 8 payments) is incorporated in (the majority of) TOC franchise agreements with the DfT.

- At the franchise bidding stage, the expected level of performance going forward is calculated and based on past performance on the part of Network Rail and the TOC, as well as a view of how this will change over the life of the franchise.

JPIPs

- JPIPs are a more recent mechanism for driving performance improvements—these annual agreements between the TOC (there are no formal mechanisms for driving forward joint freight performance schemes) and Network Rail detail how each party will contribute to continuously improving performance. The formal definition of a JPIP is in the Network Code (condition LA4). JPIPs are enforced through two mechanisms:
  - the Network Rail element of the JPIP constitutes a customer reasonable requirement and is therefore enforced through Network Rail's licence obligations;
  - TOC participation is enforced within the National Rail Franchise Terms, paras 9.1(c) and 9.4(a).

– JPIPs are generally credited with a significant role in driving performance improvements.

A2.1.3 Interfaces

The key interfaces involve government, the ORR, Network Rail and the respective TOCs and FOCs. The ORR is the independent arbiter of overall operational performance and monitors the performance of all parties. On a day-to-day basis, performance is monitored by teams whose task it is to attribute causes of delay and cancellations to the responsible parties.

The interface between Network Rail and the TOCs is a direct one, in that delays attributed to the former, such as a major signalling failure, will have an adverse impact on train services on that particular route.

The average 'minutes lateness' per train forms the basis of the financial compensation between Network Rail and the operating companies.

Equally, delays caused by a TOC will have an impact on other operators on the route and the total amount of minutes lateness will be quantified so that appropriate financial compensation is paid.

A2.1.4 Issues affecting operational performance

The key issues are as follows.

Complexity

The process of identifying which party caused the train lateness or cancellation is a complex process.

Source of management information

The delay attribution process provides important information to management as to the underlying causes of delay and what needs to be improved or adjusted to rectify this—eg, specific defects to a train fleet where rectification work has a significant financial impact on the relevant operating company.

Extensive resourcing requirement

The delay attribution process involves significant resources—300 to 500 people employed by Network Rail and train operators. In addition, the Delay Attribution Guide comprises 90 pages.

Length of minimum delay

All delays of more than three minutes are attributed to the relevant party. Although research into changing this threshold to five minutes has been undertaken, it was not taken further as it would distort delay attribution by taking out entire categories of delay caused by one party—eg, slow despatch of trains at certain stations may produce relatively low individual levels of delay, but cumulatively, this category could be significant.

Overestimation of Schedule 8 performance impacts

Network Rail has suggested in interviews with the study team that the impacts of Schedule 8 on performance can be overestimated. Network Rail states, for example, that farebox impacts and greater operational transparency between Network Rail and operating companies have driven performance improvements in recent years. Similarly, as noted above, the introduction of JPIPs has helped to focus performance issues on to more material, practical issues.

A2.1.5 Net industry costs and incentives alignments

The following considers aspects of the regime that contribute to industry costs and benefits, beginning with some positive outcomes of the regime:

– Performance has improved considerably since 2003.

– The delay attribution and management system has provided important information to both Network Rail and the train operators. This has been beneficial for managing and reducing delays.

– Given that performance-related compensation levels are known, this has allowed franchise bidders to remove or significantly reduce any risk premium that might have otherwise been added as part of their franchise bids.

– Performance regime benefits are included when investment projects are appraised. If there are significant performance benefits, a project's likelihood of approval is enhanced.

– Since its introduction in 2005, the JPIP system has functioned well and has jointly incentivised Network Rail and the TOCs to collaborate and improve performance. The JPIPs are based on a two-way obligation of Network Rail and the TOC to improve performance.

– JPIPs between Network Rail and train operators are the primary means by which performance outputs are planned and monitored on a year-by-year basis. JPIPs provide the basis for agreement of any changes proposed during the course of the control period.

– The performance regime acts as an insurance policy for TOCs in instances where severe disruption takes place.

The following lists some of the costs associated with the regime

– The process of attribution for every delay above three minutes is time-consuming and costly; considerable resources are used in this process.

– Although the performance regime reflects the short-term financial impacts of delays, there is uncertainty as to whether it accurately reflects the impact of much longer individual incidents. These incidents include major infrastructure failures that result in long-term speed restrictions, which may have a long-term detrimental impact on an operating company's business if they reduce passenger or freight volumes.

– From discussions with FOC stakeholders, there is a view that 'loss of reputation' is not compensated within the performance regime.

– Network Rail's performance regime is asymmetric as its upside is capped, so that it cannot receive payments above a specified level. By contrast, there is no cap applied to the payments associated with particular adverse incidents attributed to Network Rail. This means that the incentives in either direction are unbalanced, with Network Rail effectively being asked to bear the 'downside' risk.

– There is also asymmetry within the performance regime in the sense that a small change in a certain activity performed by Network Rail—such as relatively short periods of signalling problems—can have a much larger performance impact on TOCs through knock-on delays.

A2.1.6 Misaligned incentives

Revenue unresponsive to performance improvements

In certain circumstances, revenue can be unresponsive to performance improvements. This is likely when performance is already very good and thus there are few incentives to improve performance further if no significant revenue gains can be made.

Inappropriate balance between performance and capacity

There is concern that there is a misalignment of incentives with respect to performance and capacity allocation. To demonstrate, the incentives to maximise optimal allocation of capacity may be outweighed, or otherwise out of line with performance impacts. We understand, however, that this issue is expected to be addressed as part of the next HLOS.

Clause 18.1/Schedule 9A impacts

Clause 18.1 within TOC franchise agreements covers performance issues and there is a concern here that this may 'deaden' incentives as TOCs will not face the implications of recalibration of the performance regime at the time of periodic reviews.

Franchise agreements—revenue-sharing issues

The 'cap-and-collar' financial regime contained in some franchise agreements can potentially disincentivise TOCs. Examples of this include the 80% revenue share agreements, whereby TOCs do not receive much additional revenue from better performance but would still need to pay Network Rail.

Network Change

One concern expressed during the interviews is that Network Change provisions are increasingly being used to provide compensation with respect to performance. In extremis, this could mean TOCs not proposing changes to the performance regime at the time of a periodic review (as any benefits to them would be removed under Clause 18.1/Schedule 9A), but instead claiming compensation under Part G of the Network Code, from which the operator concerned would be able to keep the benefit.

A2.1.7 Findings of the 2004–06 Schedule 8 Performance Review

It is important to note that the Schedule 8 performance regime was reviewed in 2004 and 2005 with the final revisions implemented from April 2006. The main findings suggested that the regime was the least complicated way of providing compensation to train operators, and if this compensation was not available, additional risk premia would be factored into franchise bids.

With respect to the costs of the delay attribution process, the 2006 findings stated that any changes to this process (to reduce costs and bureaucracy) would not be implemented as it was not apparent that they would lead a significant reduction in disputes and resource savings.

In some cases, the changes would increase complexity and potentially lead to the loss of valuable information about the causes of delay.

To summarise, the Performance Review was comprehensive and involved all affected parties (Network Rail, TOCs, FOCs, etc). The findings concluded that, given the current structure of the rail industry, the Schedule 8 performance regime did not require significant change. Any changes that were proposed (on the grounds of reducing costs, reducing bureaucracy and improving incentives) were largely rejected on the basis that they would not improve the regime.

A2.2 Possessions

Network Rail's effort to maintain, renew and enhance the railway requires it to temporarily restrict access to stretches of track to allow for the necessary engineering work to be undertaken—ie, to take possession of track.

Over the price control period from 2004 to 2008 (CP3), the regulator allowed Network Rail approximately £23.7 billion in maintenance, renewals and enhancement (MR&E)

expenditure.[69] This expenditure remains broadly the same during the current price control period, the ORR allowing a total MR&E expenditure of £23.4 billion between 2009 and 2014.

The figures indicate that MR&E effort constitutes a substantial amount of work to be undertaken and, hence, will involve a high number of possessions. Table A2.1 shows the total number of possessions, of disruptive possessions, and the total percentage of disruptive possessions that were planned for the 2008/09 year.

Table A2.1 Number of possessions and disruptive possessions in 2008/09

	Possessions	Disruptive possessions	% of disruptive possessions
England and Wales	84,258	23,970	28%
Scotland	18,493	610	3%
Great Britain	102,751	24,580	24%

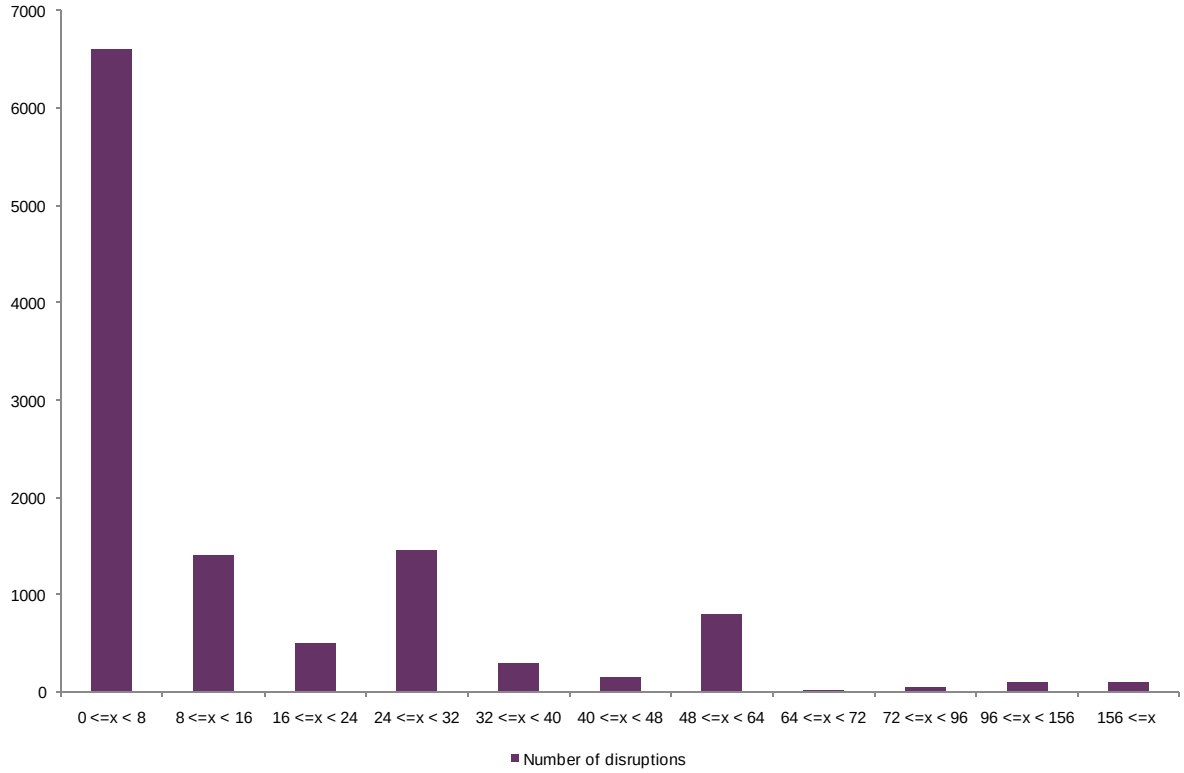
Source: Network Rail (2009), 'Annual Return 2009'.

Table A2.1 shows that, in 2008/09, about 102,000 possessions were taken by Network Rail, and about one-quarter of them were disruptive. Any possession that requires an alteration to a train schedule is defined as disruptive.

Most disruptive possessions tend to be brief, although the ORR notes that the extent of planned disruptions has increased in recent years, and that there has been greater reliance on long possessions.[70] In 2006/07, about 58% of disruptive possessions that were recorded in the Schedule 4 compensation scheme (S4CS)[71] lasted less than 8 hours and about 75% took less than 24 hours. Figure A2.1 shows the distribution of annual disruptive possessions for 2006/07.

[69] This figure is based on the final conclusion of the Access Charges Review 2003, and has been inflated to 2006/07 prices using the input price inflation (IPI) index.
[70] Office of Rail Regulation (2008), op. cit.
[71] The S4CS is the system used to calculate compensation for train operators.

Figure A2.1 Distribution of disruptive possessions by duration (hours) in 2006/07



Source: Faber Maunsell, (2007), 'Review of Possession Cost Compensation – Stage 2: Final Report', November.

A2.2.1 Objectives

Broadly, the possession regime should accomplish two objectives:

– balance the cost of the required engineering work with the negative impact arising from possessions on train operators and end-customers, while minimising transaction costs;
– allocate total industry costs of possessions to where they can be best managed.

Whereas information about the most cost-efficient work process for a particular piece of work might be more readily available to Network Rail when planning possessions,[72] assessing the cost of disruption of possessions is likely to be more difficult. Those costs mainly arise from lost revenues to train operators, the cost of providing replacement services or operating on diversionary routes for freight operators, and the inconvenience caused to passengers. Information about the economic value of these costs sits outside Network Rail, primarily with train operators.

Owing to the fact that such costs are split between Network Rail—bearing the cost of engineering work—and train operators and end-customers—being affected by the impact of disruptions—it is important that incentives are adequately aligned at the interfaces. A possessions regime that minimises net industry cost requires that Network Rail correctly reflects the cost of disruptions external to it in its planning process.

One way of internalising the external cost of disruptions is to set up appropriate compensation mechanisms that provide Network Rail with the pricing signals that correctly reflect the cost of disruptions. However, the design of such a mechanism should take into account the net costs associated with it, as the benefits of more accurate determination of the costs of disruptions might be outweighed by higher transaction costs. An appropriately

[72] In general, information about the most cost-efficient work processes could be revealed through tendering processes. However, this might not be available at the planning stage of the possession process.

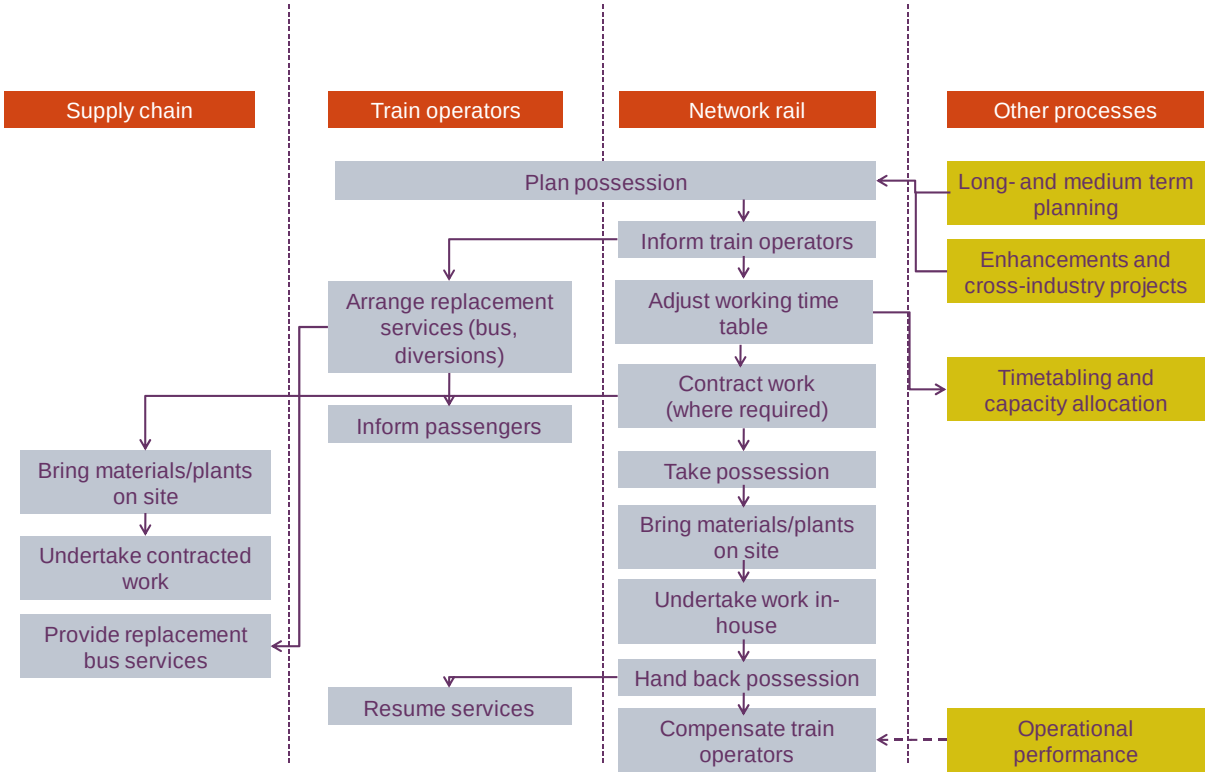
specified compensation scheme can also meet the objective of allocating the cost risk of possessions, both in terms of engineering cost as well as in terms of revenue impacts of disruptions to train operators and impacts on customers.

From the discussion on the performance regime, it is clear that it would be beneficial for the possessions regime to provide data for project appraisal purposes (eg, the possession required to enable the project will take so many hours, and will disrupt services at a cost of £X). However, it is Oxera/Arup's understanding that this is not the case currently, although work is under way to make the impact of possessions more predictable.

A2.2.2 Mapping the process onto its interfaces

Figure A2.2 outlines the possession process at a high level and maps it onto the industry structure, highlighting key interfaces. The figure also shows interfaces between possessions and other processes investigated in this study.

Figure A2.2 Interfaces in the possession process



Source: Oxera/Arup.

The main stages of the possession process are:

- possession planning;
- taking possessions and undertaking the work;
- handing back possessions and compensation.

Possession planning

At a high level, planning of possessions typically includes the following steps.

- **Planning and coordination**—at the first step of the planning stage Network Rail identifies the work to be undertaken on its infrastructure in a particular location and the likely amount of time for which use of track must be restricted to allow for this work. It also needs to coordinate possessions and overall work—eg, by considering:
  - whether to cluster different types of work at one location into one (potentially longer) possession;

  - whether to spread the work over a number of (shorter) sequences of possession;
  - how the possessions affect overall capacity;
  - the level of disruption resulting from the possession.

- **Informing TOCs, FOCs and open access operators**—having determined what work is necessary, and its impact on services, National Rail has to inform affected train operators in order to minimise negative impacts arising from the track restrictions.

- **Adjusting working timetable**—Network Rail needs to adjust the working timetable for affected routes, a process which also involves train operators. Timetable adjustments might include speed restrictions on adjacent tracks, scheduling of diversion routes, and associated changes of timings at stations or junctions along the new routes.

- **Contracting out work (if necessary)**—if the required engineering work is not entirely undertaken in-house by Network Rail, additional resources have to be contracted in from the supply chain. These services might include engineering transportation services to bring materials and plant on site.

- **Arranging replacement services**—in cases of complete closure of stretches of track, train operators might need to arrange for replacement services, such as buses. In addition, diversionary routes for passenger and freight services may need to be arranged.

- **Informing passengers**—in order to minimise disruptions, TOCs and Network Rail need to provide passengers with timely information about the restrictions resulting from possessions, and replacement services that have been put into place.

The structure of the possession planning process is set out in the Rules of the Route and Rules of the Plan. A detailed description of the process is provided in Box A2.1.

Box A2.1 The possession planning process in the Rules of the Route

The Rules of the Route are rules regulating the arrangements for access to the various parts of the main rail network, when affected by inspection, maintenance, renewal and other works. They are one of a pair of documents, along with the Rules of the Plan, which, when agreed on by the users of the relevant railway route, provide rights of access that, with certain rights of train operators, are described as 'firm rights' and enjoy priority in the timetabling process. The Rules of the Plan are rules regulating the standard timings between stations and junctions, together with other matters enabling trains to be scheduled into the working timetable for the various parts of the main rail network. The Rules of the Route are published annually for each timetable year.

At the start of the possession planning process, Network Rail starts gathering information about access and resource demands of major works to construct the Annual Access Plan and starts informal consultations regarding draft Rules of the Route (Draft V0) with principal stakeholders throughout the rail industry. This consultation process aims at achieving the optimal balance between access to the network for train operators and access for maintenance, renewals and enhancement work.

This is followed by formal consultation of the Principal Rules Proposal (Version 1 of the Rules of the Route) with train operators, leading to the issue by Network Rail of the details of possessions during the upcoming timetable year (version 2 of the Rules of the Route, called the Final Principal Rules). The Final Principal Rules details agreement for the first half of the timetable. Train operators have a right of appeal against the contents of the Final Principal Rules within 15 days. Any dispute would be referred to a dispute panel comprising representatives from Network Rail, freight and passenger operators.

A proposal (Version 3 of the Rules of the Route, called Final Principal Rule and Proposal for Subsidiary Change) for the second half of the timetable is published about 37 weeks before the timetable comes into force, and operators can comment on or challenge it within one month. The final Rules of the Route for the second half of the timetable is published about 22 weeks before the timetable. Again, operators can appeal the Rules within 15 days.

Detailed possession planning continues to be developed after the publication of the Principal Rules Proposal and requirements for additional, amended or cancelled disruptive and non-disruptive engineering access are compiled into a Draft Period Possession Plan (DPPP). The DPPP is reviewed with train and freight operators, with particular attention being applied to changes from the Principal Routes Proposal and to items disputed at this time. The agreed access arrangements from this DPPP meeting are consolidated into the Confirmed Period Possessions Plan (CPPP), which is issued 26 weeks before the timetable.

Train operator consultation

Train operator consultation runs parallel with the process of developing the Rules of the Route and finalising the Period Possession Plan. The stages are as follows.

- Period Possession Plan—every four weeks, a detailed programme of proposed engineering works covering a four-week period will be issued as the DPPP at least 28 weeks before the start of the four weeks concerned. This will contain details of proposed disruptive possessions, together with any non-disruptive possessions.
- Informed Traveller—the CPPP will be issued 26 weeks prior to the start of the relevant four-week period giving sufficient details of planned disruptive possessions to enable an assessment of train alterations to be made. Network Rail will advise details of disruptive possessions at this stage as part of the Informed Traveller process. This enables train operators to assess how they will deal with the disruption and keep travellers informed.
- Trains Meetings—in certain instances, prior to the conclusion of train operator consultation, Network Rail may arrange a Trains Meeting for all affected train operators. This is most likely where a particular possession, or series of possessions, will result in significant train alterations. Trains meetings will be held to:
  - outline the likely disruptive effects;
  - determine train alterations required;
  - agree an outline train plan.

  In addition to these formal processes, Network Rail will meet with any train operator, if requested to discuss any concerns they may have.

Change procedures

Changes to possessions given in the Rules of the Route are possible. These changes can arise, for example, from a re-assessment of engineering need.

Given the need to achieve deadlines provided in the Informed Traveller process, details of amended train services must be available 12 weeks before the date of operation. Therefore, Network Rail will consult with train operators regarding possessions and other capacity restrictions which are disruptive to agreed train paths in sufficient time to allow details of those disruptive possessions to be included in a CPPP, to be published 26 weeks prior to the start of each four-week period.

Where a need arises to amend the Rules of the Route in either the Draft or Confirmed Period Possession Plan to cater for urgent safety requirements or other emergency situations, all parties concerned will co-operate in accelerating the normal timescales in this procedure, commensurate with the urgency of the circumstances.

Source: Oxera, based on Network Rail (2010), 'Rules of the Route, Final Principal Rules and Final Subsidiary Rules, 9 July'.

Track possession

During track possession, the following steps are typically undertaken.

- **Taking possession**—before the work commences, the possession is taken. This typically involves putting the appropriate signalling into place and securing the work site.
- **Bringing material and plant on site**—this stage of the process involves bringing material and plants to the work site, which can be undertaken either by Network Rail, sub-contractors that are tasked to undertake the engineering work, or other contractors charged with the transportation services. This start-up process might require use of the

track where the work will be undertaken, use of an adjacent track, and/or site access via a road alongside the work site.

– **Undertaking the work**—the engineering work that caused the possession is either undertaken by Network Rail or contracted out to sub-contractors.

– **Handing-back of possession**—after completion of work, the track is handed back to the train operators. It can then either be fully used, or some restrictions such as reduced line speeds might persist if further work (and subsequent possessions) is required.

Compensation

The main mechanism for compensating train operators for disruptions caused by possessions is through the Schedule 4 regime, which is part of the track access agreements of franchised passenger and freight operators.

In the 2008 Periodic Review, the Network Code was changed in order to exclude Part G—compensation for disruption caused by restrictions of use in connection with the implementation of a proposed network change. Broadly speaking, compensation under Part G of the Network Code allowed for the possibility of compensation of train operators' actual costs arising from disruptive possessions in relation to network changes. This process involved detailed consultations between Network Rail and train operators.

Interfaces

The main industry interfaces in the possession process are between:

– **Network Rail and train operators (TOCs, FOCs, open access operators)**—in the initial planning stage Network Rail consults with train operators. The process is set out in the Rules of the Route. Where possessions are taken for network change, Part G of the Network Code provides further requirements for consultation. A further interface with train operators exists at the end of possessions: after the engineering work has been concluded, the possession is handed back to train operators.

– **Network Rail and the supply chain**—Network Rail interacts with the supply chain in cases where it contracts out engineering work. A further interface might exist if transportation of materials and plants is also undertaken by contractors.

– **Train operators and the supply chain**—Passenger train operators require services from bus operators if possessions demand setting up replacement bus services to bridge the closed track stretch.

The possession process also has interfaces with other processes investigated in this study. On the one hand, long- and medium-term network planning is to an extent driving engineering work such as electrification, which in turn will require possessions to be made.[73] On the other hand, network enhancement, work such as station redevelopment, might also require temporary track possession. These processes interact with the possession process, mainly in the possession planning stages.

A further interface with processes discussed in other sections of this study is related to the timetabling impact of possessions. When taking a possession Network Rail needs to adjust the working timetable on affected routes—eg, by adjusting line speeds or scheduling on diversionary routes.

Moreover, it must be taken into account that not all possessions are handed back as planned. When such unplanned overruns occur, compensation of affected train operators is

[73] Other important drivers of possession are maintenance and renewals work.

regulated by Schedule 8 of track access agreements, covered in the previous sub-section on performance.

A2.2.3 Existing incentives

The overarching requirement for the incentive regime for possessions is to strike an optimal balance between undertaking the required engineering work in a cost-efficient manner while minimising disruption.

The current incentive regime for possessions is based on two main pillars:[74]

– Schedule 4 of track access agreements;
– PDIs.

Schedule 4

Schedule 4 of track access agreements between train operators (franchised passenger and freight operators) and Network Rail sets out how train operators are compensated for the cost of disruption caused by planned possessions. It is also available to open access operators. Compensation is expected to be broadly reflective of the degree of disruption, while striking a balance between accuracy and simplicity to minimise transaction costs for minor cases.

For franchised passenger operators, Schedule 4 provides formulaic cost and revenue compensation for all possessions, with additional compensation available depending on the level and impact of disruption. A three-tiered structure of compensation in Schedule 4 was implemented in the 2008 Periodic Review. The characteristics of each tier are as follows.[75]

– **Type 3 possessions**: single possession greater than 120 hours (includes public holidays), will receive formulaic compensation as default, but with the possibility of actual revenue losses and costs (subject to a materiality threshold).

– **Type 2 possessions**: single possession greater than 60 hours, but equal to or less than 120 hours (excludes public holidays), will receive formulaic compensation as default but with the possibility of actual costs (subject to a materiality threshold and in respect of categories of direct costs only) mirroring existing Significant Restrictions of Use arrangements.

– **Type 1 possessions**: all other possessions will receive formulaic-based revenue and cost compensation.

To take into account the fact that the costs of disruptions from a series of subsequent possessions on the same stretch of track or in locations close to another might increase disproportionately to the number of disruptions, the 2008 Periodic Review introduced a separate compensation regime for repeated disruptions. For ‘sustained planned disruption’, compensation will be available on a similar basis to type 3 possessions, if specified materiality conditions are met.

Furthermore, the provisions of Schedule 4 include a rewarding mechanism for Network Rail for early notification of possessions to TOCs by discounting the marginal revenue effect that permits compensation in case of early notification.

Open access passenger operators are able to claim compensation for type 3 possessions and sustained planned disruptions. They can also claim compensation for type 1 and type 2 possessions if they pay an access charge supplement like franchised operators.

[74] Schedule 8 of track access agreements provides incentives relating to unplanned possessions (eg, possession overruns) and is covered in the previous sub-section.
[75] Office of Rail Regulation (2008), op. cit.

Freight operators also receive compensation for planned possessions under Schedule 4 of track access contracts. The 2008 Periodic Review introduced a three-tiered compensation structure similar to the one for franchised passenger operators.

Each tier of compensation aims at reflecting the impact that disruption has on freight services. Compensation might include flat-rate liquidated damages compensation for minor disruption and the possibility of additional compensation of actual costs and losses for the most disruptive possessions. Compensation is also contingent on whether notice of possession was given early or late.

In addition to Schedule 4 compensation, Schedule 8 provides compensation for disruption from unplanned possessions (including possession overruns) for passenger and freight operators.

Network Rail can broadly recover the Schedule 4 compensation expenditure during the price control period. In order to do so, forecast Schedule 4 expenditure is included in the allowed cost base in the determination of Network Rail's funding and outputs by the ORR. To recover the allowed Schedule 4 cost, Network Rail charges an access charge supplement to franchised passenger operators (and open access operators that opt in) to have the full Schedule 4 compensation. No supplement is levied on freight operators.

On aggregate and over the medium term, actual expenditure for Schedule 4 compensation has been slightly lower than the income from the access charge supplement over CP3, as Table A2.2 shows, although there might be more significant differences between compensation and supplement payments for individual train operators.

Table A2.2 Allowed and actual Schedule 4 expenditure (£m, in 2008/09 prices)

Year	Income from access charges supplement	Compensation to train operators	Outperformance/ (underperformance)
2008/09	98	74	24
2007/08	96	115	–20
2006/07	93	101	–7
2005/06	94	79	16
2004/05	86	67	18
2005/06–2008/09	**467**	**436**	**31**

Note: Values for 2004/05–2007/08 have been inflated using RPI.
Source: Network Rail (2009), 'Regulatory Financial Statements. Year ended 31 March 2009'.

4.8.1 Possession disruption indices

A second pillar of incentives to minimise total industry costs of possessions is related to the introduction during the 2008 Periodic Review of performance indices that measure disruptions and the prescription of pre-defined trajectories by the regulator.

To reduce disruptions from planned possessions, the regulator introduced new PDIs for passenger and freight traffic. Broadly, the passenger index (PDI-P) measures the impact of engineering possessions in terms of the economic value of the excess journey time that passengers experience, normalised by total train-km; and the freight index (PDI-F) measures the impact of the unavailability of track for freight use, weighted by the level of freight traffic operated over each section of track.[76]

These indices take a base value of 1.00 in 2007–08. In the future, they will show by what proportion the disruption experienced by passengers and by freight operators has increased

[76] Ibid.

or reduced relative to that in the base year. Table A2.3 shows the levels of these indices that Network Rail is required to deliver during the 2009–14 control period.

Table A2.3 Possession disruption index trajectories for passengers and freight (2007/08=1.00)

	2009/10	2010/11	2011/12	2012/13	2013/14
Passenger (PDI-P)	1.02	0.91	0.83	0.68	0.63
Freight (PDI-F)	1.00	1.00	1.00	1.00	1.00

Source: Network Rail (2009), 'Regulatory Financial Statements. Year ended 31 March 2009'.

These regulatory targets require a progressive reduction in disruption to passengers of 37% by the end of the control period in 2013/14, compared with the base year 2007/08; furthermore, there should be no increase in the level of disruption experienced by freight operators, as measured by the PDI-F index. In the event of failure to meet either of these targets, the ORR will assess whether this amounts to a breach of Network Rail's network licence.

In addition to these indices, the regulator has introduced supplementary KPIs. These KPIs will not form regulated outputs, but will be used to:[77]

– provide information on areas which are not fully reflected in the PDIs;
– help understand movements in the PDIs;
– act as a check against any perverse behaviour that might result from strategies designed to drive improvements against the PDIs.

In addition to these metric-based incentives, Network Rail and passenger operators have started to agree Joint Network Availability Plans (JNAPs). It is Oxera/Arup's understanding that these are now in place for all franchised operators, and that they are a way of reflecting Network Rail's national Network Availability Plan at a local level relevant to each TOC. However, unlike JPIPs, they do not have a contractual basis in either the Network Code or in the National Rail Franchise Terms.

A2.2.4 Net industry costs and incentive alignment

The analysis undertaken by the study team, including feedback received in stakeholder interviews, suggests that outcomes of the Schedule 4 process are broadly acceptable. However, this analysis also suggests areas for improvement relating to both the transaction costs involved in the process and the alignment of incentives between parties at the interfaces.

Transaction costs

The possession process involves considerable administrative burden or, in economic terms, transaction costs. These costs comprise the resources concerned with the process, both within Network Rail as well as at other stakeholders, primarily train operators. The process involves extended planning and consultation efforts at the interfaces between Network Rail and affected stakeholders, including train operators, but also, where relevant, PTEs/ITAs and local authorities, and between different departments within Network Rail. In addition, possession processes are affected in large renewals and enhancement projects, which often involve cross-industry disciplines.

Interviews with stakeholders have suggested that the previous possession planning system was seen to contain too many people, too many interfaces, and was not coordinated properly. However, it has also been suggested to the study team that recent changes, such as the reorganisation of the National Delivery Service (NDS), integrating main network

[77] Ibid.

planning functions, has improved the process. For example, the NDS reorganisation reduced staff levels by about 20%, which, in turn, reduced transaction costs. It has been suggested that the reorganisation of the possession planning functions has also had a positive impact on incentives to reduce the disruptive impact of possessions; this aspect will be discussed below.

However, there appears to be further scope for reducing transaction costs. The possession planning process is guided by a set of planning rules, which, for example, give indications on required possession duration. Possession plans collated by Network Rail's National Plan Integration team from its delivery departments are subject to a consultation process with train operators. A final possession plan is issued 26 weeks before the work is undertaken. It has been suggested to the study team that the time elapsed between the start of possession planning, and the end of the possession itself, can be as much as 2.5 years.

Network Rail has suggested to the study team that the efficiency of the planning process could be increased by streamlining the process and reducing the gap between the time that the final possession plan needs to be issued and the possessions being taken, currently 26 weeks. It has been pointed out that during this approximate six-month gap, work specifications can change, requiring alterations of the Rules of the Route along the requirements set out in the Rules' change procedure. This involves reiteration of the consultation process and has been described by Network Rail as increasing administrative costs, which could be reduced through a reform of the process.

During the interviews, questions have been raised regarding the effectiveness of the 12-week threshold of the notification discount factors of Schedule 4, as it has been suggested that most tickets are not sold more than 12 weeks in advance. However, this issue was picked up during Periodic Review 2008.

In addition, stakeholders have indicated the potential for increasing the efficiency of possessions through measures related to work processes on site, such as the time required to access sites, isolations and deadtimes, undertaking of preparatory work while lines are still open, or the sweeping of tracks before handing back a possession. This feedback is consistent with the messages in the work undertaken by Lloyds Register Rail for the 2008 Periodic Review, which compared possessions between GB rail and other countries.[78]

A2.2.5 Incentive alignment

Overall, the current arrangements of the possessions regime seem to be leading to broadly appropriate outputs; stakeholders have suggested that outcomes are—eventually—acceptable. However, the process might be considered laborious and there might be scope for streamlining it.

Although Schedule 4 compensation is broadly reflective of the degree of disruption and, therefore, possessions carry about the 'right' price signals, this incentive to minimise total industry costs by balancing the cost of engineering work and disruption is weakened by the regulatory allowance of Schedule 4 expenditure and the funding through access charge supplements. However, the motivation to outperform regulatory allowances for Schedule 4 expenditures provides an incentive to increase the efficiency of possessions.

Incentives for individual train operators to engage with Network Rail in the planning of specific possessions by which they will be affected are reduced because of an externality. The externality arises because the TOC payments to fund Network Rail's Schedule 4 expenditure are based on expected Schedule 4 costs by operator. Therefore, the payment is independent of any potential savings from reduced disruption of an individual possession over the remaining life of the price control period. However, if the specific TOC's engagement leads to longer-term efficiency gains in possession planning as a whole, the potential positive

[78] Lloyds Register Rail (2006), 'Possession benchmarking exercise', report prepared for the Office of Rail Regulation, September.

effect on Schedule 4 cost is shared with all train operators through lower regulatory allowances and hence contributions by operators in proportion to their individual supplement payment.

A direct incentive to Network Rail to increase the efficiency of the possession planning process is provided through the regulatory network availability targets. These targets set maximum levels of disruption from planned possessions as measured by the network PDIs for passenger and freight traffic. However, concerns have been raised in the stakeholder interviews in relation to the usefulness of this measure, and its ability to affect behaviour in Network Rail. Although it has been acknowledged that the indices work directionally—ie, planners have a broad understanding of whether specific possessions have a positive or negative impact on the PDIs—planners do not know the exact impact of possessions on the PDIs.

In addition to the issues raised in relation to the alignment of incentives created by Schedule 4 and the network availability targets, potential conflicting incentives of the possession regime with the aim of increasing overall network availability have been highlighted in the stakeholder interviews. In particular, that Schedule 4 might disincentivise increasing the times each day when day trains can run (ie, reducing white space) as it might reduce the time available for non-disruptive possessions—ie, possessions that are not subject to Schedule 4 compensation.

A2.3 Station management, operation and upkeep

Although stations are not a direct part of the vehicle-track interface, they are a very important part of the passenger experience of using the rail network. They are used by passengers to access rail services, as well as to purchase tickets, await services, and often to make use of retail/commercial outlets.

There were 2,535 stations in use on the GB rail network in 2009.[79] This number covers a wide range of stations in terms of scale and number of services using them. Some stations are staffed or part-staffed, and others are not.

For each station, a distinction needs to be drawn between the following parties:

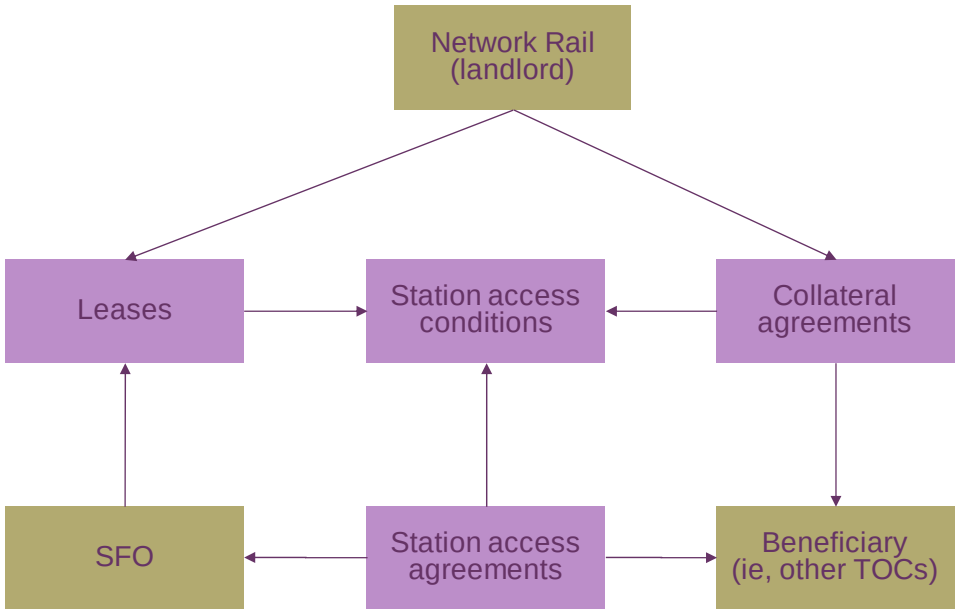
– the station landlord (normally Network Rail);
– the SFO;
– other TOCs that operate services that call at that station (also known as beneficiaries).

In addition, stations are part of the local community, creating interfaces with local authorities and other funders and stakeholders.

Some of the contractual interfaces between (rail industry) parties are illustrated in Figure A2.3.

[79] Green, C. and Hall, P. (2009), 'Better Rail Stations', November.

Figure A2.3 Stations interfaces



Source: Based on ORR, http://www.rail-reg.gov.uk/server/show/nav.224.

Network Rail is the landlord of most stations and is funded (by the DfT or the relevant government authority) to repair, maintain and renew the relevant parts of stations.[80] Network Rail must also carry out specified enhancements as part of its control period settlement and as part of other DfT schemes, such as Access for All. In addition to owning and maintaining most stations, Network Rail is responsible for managing several of the larger stations.[81]

The day-to-day operation of stations is usually leased out to TOCs via station leases. The lessee TOC then becomes the SFO with a licence to operate the station and normally takes responsibility for light maintenance.

The day-to-day operation of a station includes the provision of passenger information, cleaning, and the operation of dispatch teams, if required. The SFO can sub-lease property to retailers.

Other TOCs can access the station to a greater or lesser extent as their services require, provided that they obtain a station access agreement. At some stations they may only use one platform or part of the station, whereas elsewhere they may make use of the whole station. These other TOCs also pay a share of station long-term charge and qualifying expenditure to the SFO for operating the station.

A2.3.1 Objectives

The DfT specifies requirements from stations in franchise agreements and through programmes such as the National Stations Improvement Programme (NSIP) and Access for All programme.

However, in terms of delivery against these objectives, and as a whole, the implied objective of the current structure for stations is not entirely clear given the complex contractual nexus. The intention may be to divide responsibility for assets/maintenance and operation to those organisations that have the closest matching activities elsewhere in their business and are therefore best placed to carry them out.

[80] A small number of stations are owned by third parties such as St Pancras, Prestwick International and Warwick Parkway.
[81] Under the Station Licence granted to Network Rail Infrastructure Ltd., Schedule 2, Network Rail may manage Glasgow Central, Edinburgh Waverley, Leeds, Manchester Piccadilly, King's Cross, London Bridge, Birmingham New Street, Waterloo, Paddington, Liverpool Street, Charing Cross, Victoria, Euston, Liverpool Lime Street, Canon Street, Gatwick Airport and Fenchurch Street stations.

Network Rail acts as station landlord, partly for historical reasons, but also because it has a long-term role in the industry that does not exist for TOCs with their finite franchises. Many station assets can have long asset lives, perhaps as long as 100 years for something like a station roof. Therefore, Network Rail is arguably better incentivised to own and conduct heavy maintenance on station assets.

TOCs operate the stations because this aligns with their incentives for growing passenger revenues. By operating the station and performing light maintenance they can choose the optimal amount of effort to grow demand.

The various financial flows that occur between the parties aim to add some degree of price signals to each party.

The wealth of contracts also provide contractual protection for each party. For example, station access agreements ensure that beneficiaries are able to access the station.

The next sections set out in greater detail the nature of some of the interfaces that surround stations.

A2.3.2 Aspects of stations

Station ownership

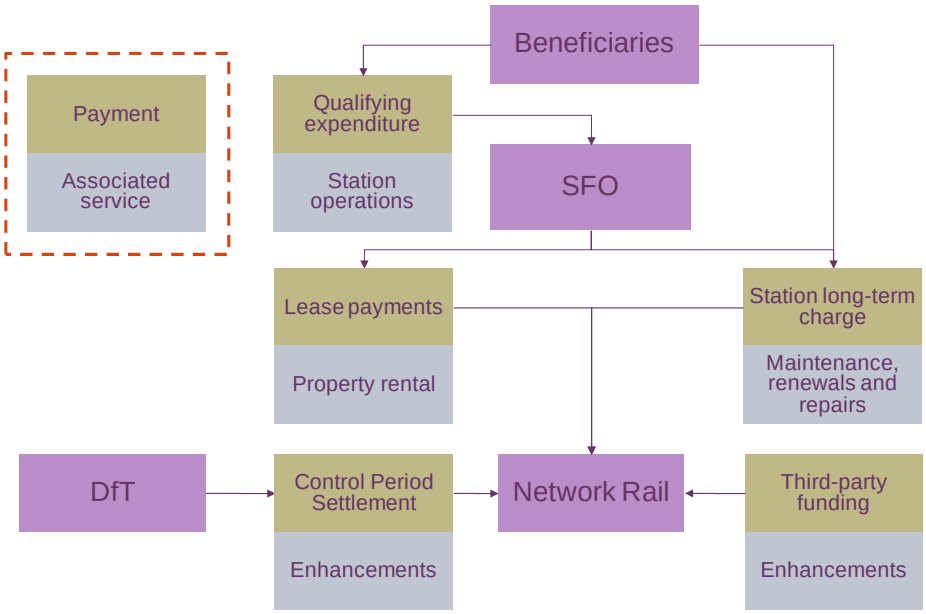
As mentioned above, stations are primarily owned by Network Rail, with some rare exceptions—such as Warwick Parkway, where a TOC owns the station. Non-Network Rail stations are usually newly constructed stations that were initiated by the TOC.

Decisions about the sale of property or the construction around a station are taken by Network Rail, but can often be influenced by a local authority or a TOC.

Station finances

Many of the interfaces surrounding stations involve payments of some kind. These payments all correspond to different tasks that are performed, such as enhancements or station operation. Figure A2.4 below depicts some of these financial flows and the associated services that the payments cover.

Figure A2.4 Stations—financial flows



Source: Oxera.

Leasing

When stations are leased from the landlord to the SFO, rent paid to the landlord has two components—property rent and long-term charges (LTCs).[82] Station access conditions, which include the LTC, are incorporated by reference into each station access agreement, and therefore the proportions of the LTC are applied to the beneficiaries as well as the SFO. Since station access agreements are subject to ORR approval, LTCs are therefore directly regulated.

LTCs should capture the costs of the maintenance, renewals and repairs conducted by Network Rail at a station. Prior to PR08, LTCs were implemented at the individual station level, even though the LTC at a station did not necessarily correspond to the expenditure at that station. In PR08, Network Rail's proposal to move to a LTC paid at the level of the franchisee was adopted in the final determination (although price lists at an individual station level that support this are also available).[83]

The SFO also pays property rent to the landlord. These rents are determined in the leases themselves and hence are not regulated by the ORR.

Qualifying expenditure

Qualifying expenditure (Qx) is the charge from the SFO to other users of the station. It covers the day-to-day running of the station and the SFO's maintenance, repairs and renewals obligations.

Qx is not a regulated charge as its value is not set out in any regulated document. However, the methodology for its calculation is regulated as it is set out in the station access conditions. Network Rail also charges Qx to users at the stations that it manages.

Qx charges are split proportionally among the TOCs using the station, depending on their number of departures as a proportion of the total station departures.

Station enhancements

Major station enhancements are specified and funded in Network Rail's control period settlement. However, for smaller enhancements that are either initiated by Network Rail itself, a SFO, or even a beneficiary, there is a formal station change procedure that must be undertaken. Station changes affect a wide number of parties, not least passengers, and therefore the procedure ensures that relevant consents are obtained.

Station change proposals

The station change procedure is incorporated in the station access conditions.[84] A station change proposal is promoted either by the SFO or by Network Rail. It requires the unanimous consent of other beneficiaries (ie, other TOCs using the station) and also consultation with the DfT and the ORR.

If the station change proposal modifies the access conditions to the station, then the ORR's approval is needed (be it general or specific approval; general approval allows the process to happen more quickly).

Network Rail proposals require significant information to be given to the SFO, followed by a station meeting at which the proposal is discussed with the affected parties. Ultimately, finalised Network Rail proposals must contain:

– details of the parts of the station affected;
– the nature and detailed specification of the development to which the proposal relates;
– estimated timetable for commencement;

[82] Standard Station Lease Document 2003
[83] ORR (2008), op. cit.
[84] 'National station access conditions 1996'.

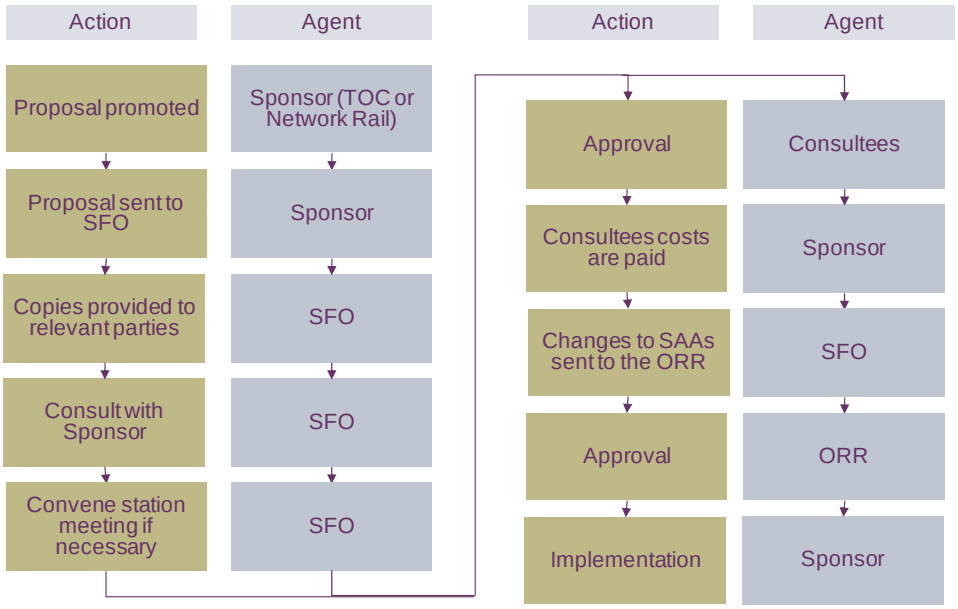
– proposals for obtaining planning consents;
– proposed working arrangements;
– proposed alternative arrangements;
– offers of compensation to the affected users;
– other features such as proposals to display information.

The conditions for acceptance of a proposal rely on either unanimous or majority acceptance where there is a failure of the requisite majority to give a Notice of Objection. If a proposal is accepted then any subsequent changes to the station access agreements must be submitted to the regulator.

The costs of proposals for change shall be apportioned between the relevant operators and Network Rail, unless it is a Network Rail proposal, in which case the costs shall be fully borne by Network Rail.

Figure A2.5 shows a highly stylised version of the station change procedure, providing some indication of its complexity and the number of interfaces that occur.

Figure A2.5 Stylised station change proposals



Note: Based on 'National station access conditions 1996 (England and Wales)', Part C.
Source: Oxera.

Other types of procedure at the station level are relevant. If there is a serious change that results in the closure of a station facility, then a network modification/closure process needs to be followed. Equally, if there is disposal of land then Network Rail's Network Licence Condition 26 needs to be applied.[85]

Station maintenance

In general, Network Rail as landlord is responsible for the structural parts of the station fabric, whereas the SFO is responsible for the parts of the station with which passengers may interact. In situations where either the SFO or the station landlord fails to adequately fulfil its responsibilities, the other party can issue a breach notice. If following this notice the work is not done within a reasonable time and to a reasonable standard, then the other party may undertake the work itself and charge the costs to the party causing the breach.

[85] Network Rail Network Licence.

Station operation

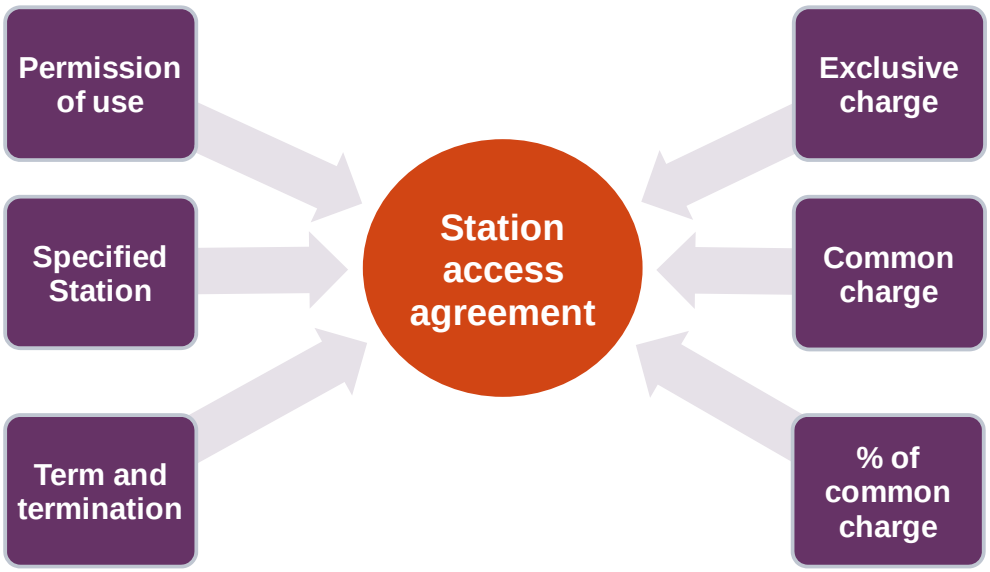
The operation of the station is very important from the passenger perspective, as they use stations for purposes other than as an access point to trains—ie, to receive information about services, purchase tickets and use other station facilities.

The SFO is normally responsible for station operation, except at stations managed by Network Rail. This means that an SFO may have to interact with retailers to try and maximise revenue earned at stations. It also has to interact with other TOCs by providing them with access to the station and dispatching their services.

Access agreements

In order to call at or use a station where it is not SFO, a TOC must have an access agreement. These access agreements must be approved by the ORR. Figure A2.6 provides an illustration of the contents of a station access agreement.

Figure A2.6 Station access agreement



Note: This is not an exhaustive representation of the contents of the station access agreement. The national station access conditions contain many important clauses for station access, as do the station-specific annexes.
Source: Oxera.

Station Access Conditions are currently being reformed, with respect to third-party rights among other aspects, with a view to achieving contractual simplification.

Station retailing

SFOs have incentives for improving the station environment for increasing patronage and for improving retail opportunities. Network Rail also often earns a share of the revenue from retailers and will face similar incentives.

Depots

Depots, like stations, are usually owned by Network Rail, but differ significantly from stations. Depots can have a number of operators using them, but tend not to require involvement from third parties in their development, which can allow more flexibility in making changes to them. Network Rail has indicated that these factors have tended to mean that there are fewer problems with the maintenance and operation of depots than there are with stations.[86]

[86] Interview with Network Rail on 24/06/10.

A2.3.3 Net industry cost and incentives

The large number of interfaces surrounding stations leads to much scope for a misalignment of incentives. Indeed, the existing arrangements are fraught with complexity and uncertainty surrounding responsibility.

- Many of the misalignments arise due to split responsibilities. Split responsibilities for station development (design, development, regulation and funding) lead to delays and, potentially, missed opportunities. The split responsibility for maintenance, renewals and repairs of assets may lead to game-playing with sub-optimal maintenance of certain assets—eg, a TOC may be incentivised to under-maintain an asset knowing it will subsequently be renewed by Network Rail.

- Network Rail could be considered to be too large an organisation to have visibility of where changes need to be made at all stations. Particularly because it does not routinely operate in certain stations in the same way that some TOCs do.

- Essentially, the current responsibilities lead to sub-optimal asset management (improvements may not be taken forward, and life-expired assets are maintained to avoid removal procedures) and asset knowledge.

- Network Rail (as landlord) has an incentive to help maximise retail takings, whereas a TOC may not if additional value gets passed through via lease payments.

- There are limited incentives for Network Rail or TOCs to make station changes due to bureaucracy costs. This can lead to significant cost inefficiencies—eg, not demolishing a redundant building, leading to unnecessary OPEX.

- There are limited incentives for either TOCs or Network Rail to gain detailed ‘knowledge’ of station assets—asset knowledge becomes dispersed under current arrangements.

- The requirements of Heritage Authorities may have a large associated cost. The Heritage Authorities themselves have limited incentives to consider the cost of development of stations when setting requirements for materials and structure to be used on stations.

- Responsibilities for station development are an important issue. As the ‘title’ for stations rests with Network Rail, any station developments opportunities (which have long lead times that can often span franchises) necessarily fall to it under the current (and potentially future) arrangements. Since development opportunities might also cover railway lands beyond the station environment, any changes in responsibility for stations will need to take this into account, or else create a new interface between Network Rail and the operators.

- One of the main issues is conflicts in responsibility for the different components of repair, renewal and maintenance. One of the main concerns was that where responsibility for an asset was split between the station landlord and the SFO conflicts could arise over:

  - when a repair should be done;

  - inefficiencies due to higher costs from assets deteriorating beyond the efficient level due to these conflicts;

  - problems for passengers and station users as a result of these issues.

- Assets with a direct impact on the quality of service experienced by passengers were given to the SFO, while assets which require long-term maintenance or high CAPEX were given to the station landlord.

There is a direct cost associated with monitoring companies across interfaces. For example, TOCs may monitor Network Rail to ensure it is genuinely performing its relevant duties, and vice versa. This could be an inefficient cost if there is unnecessary duplication involved.

There is a long-term opportunity (non-cash) cost of lost revenue/patronage on the rail network due to sub-standard stations. This could arise due to the limited incentives for long-term investment in stations by TOCs.

Network Rail has also suggested that possible cost savings could be made from the removal of unnecessary management tiers at stations. Hence at stations which have come under Network Rail management, it was of the opinion that it had reduced station costs.

A2.3.4 Conclusions

Stations are an important part of the rail network, yet the current system appears to be overly complex with a large number of parties involved and a number of potential misalignments.

The situation in Great Britain might be improved through focusing attention on management and repair to one party, subject to non-discrimination protections. Some of the potential solutions include:

- allocating TOCs a full repairing lease for the majority of stations,[87] which would require suitable 'end of franchise' protections to be built in so as to prevent lack of interest by the TOC in new development opportunities arising towards the end of the franchise;
- for stations where considerable engineering work is required for structural upkeep, an increase in Network Rail's portfolio of managed stations might be considered, and perhaps be combined with some refocusing of Network Rail's commercial incentives (and possibly re-creating a separate 'Major Stations' organisation);
    - criteria for deciding which stations should be added to the Major Stations list include transport network importance (eg, airport stations); passenger importance (eg, based on footfall); operational scale and complexity (eg, multiple operators and platforms); asset scale and complexity (eg, heritage status, number and complexity of asset types); imminent large projects; railway land ownership around stations; and development opportunities;
- moving to a more standard commercial property arrangement, whereby Network Rail as landlord takes responsibility for the fabric of the building, and TOCs take responsibility for customer-facing assets, akin to the model successfully applied in the Netherlands (where specifically the Network Rail equivalent takes full responsibility for rail-facing assets—platforms, canopies, shelters, etc—and the operator takes responsibility for the off-rail operational areas—retail, station building, redevelopment;
- local authorities/transport authorities may wish to assume more responsibility.

A2.4 Rolling stock

Rolling stock is the term for all vehicles that move on the railway, including locomotives, carriages (including multiple units), wagons and engineering equipment (tampers, etc). It is rolling stock that carries passengers and freight on the rail network.

Rolling stock is quite diverse both across and within countries. In Great Britain, certain technical differences can lead to incompatibility between rolling stock and parts of the network. One obvious distinction is that between diesel and electric-powered vehicles, but

[87] It is noted that Network Rail spends approximately £14m per annum on station surveys, covering 20% of station assets, and that such surveys will have to be undertaken to enable franchise bidders to be comfortable with the condition of assets they are taking on.

other incompatibilities exist from the size and shape (loading gauge) of trains, which can restrict their ability to use specific parts of the network, and through other technical incompatibilities (eg, interaction between electrification equipment and signalling). Track gauge differences are not a significant problem within Great Britain.

A2.4.1 Existing structures

Since privatisation, passenger rolling stock has been primarily owned by three companies, known as the ROSCOs (**Ro**lling **S**tock Leasing **Co**mpanies). These companies own the assets and then lease them to TOCs. New rolling stock is produced by manufacturers following an order from the DfT or TOCs and can then be bought and held by the ROSCOs, although direct leasing through manufacturers (or other leasing companies) is also possible.

Part of the reason that rolling stock is not generally owned by TOCs is because rolling stock asset lives are significantly longer than the length of a franchise, making it unlikely for a TOC to be willing to purchase rolling stock on its own balance sheet. TOCs generally have a relatively low amount of capital invested within them and hence can usually only sustain operating leases as their balance sheets are not otherwise strong enough.

The total available fleet is relatively fixed in size and capability, except when new stock is introduced, which can trigger a 'cascade'. A rolling stock cascade is the process of re-allocating stock across routes (either within a TOC or owning group, or more widely) following the introduction of new stock on one specific route. Franchise specifications lead to implicit or explicit prescriptions of certain types of rolling stock on specific routes.

Freight rolling stock is rather different to passenger rolling stock and is often specialised for the type of freight it carries. Some freight rolling stock is leased from ROSCOs, including specialist wagon-leasing companies. However, because freight operators are not franchised, it is also a realistic proposition for freight companies to own wagons and locomotives. Freight companies can do this because they can use the asset for its full lifetime.

A2.4.2 Rolling stock procurement

The procurement of new rolling stock may be triggered by a TOC or owning group, a FOC, by a ROSCO deciding to increase its portfolio, or by the DfT, Transport Scotland or other funder.

A key aspect of this process is the responsibility for bringing new rolling stock onto the network. On the freight side, freight operators determine what rolling stock they need, and take full responsibility for the process of bringing it onto the network, with the support of Network Rail, manufacturers and leasing companies. On the passenger side, the responsibility is less clear, falling sometimes to government, and sometimes to operators.

On the passenger side various models for procurement have been utilised. For example, the new rolling stock on the London Overground concession was ordered by TfL, then transferred to a special purpose vehicle before being sub-leased to LOROL for operation.

For the InterCity Express Programme (now on hold) and Thameslink, the DfT has initiated procurement but the intention is for funding to come from the private sector. Diesel Trains Ltd was set up by the DfT to buy new DMUs directly, with the intention to sell the fleet on to a financier.

Developments in the procurement of rolling stock are such that there may be an increase in the alternative methods to the traditional ROSCO acquisition of stock. ROSCOs are able to purchase rolling stock without government support and on the occasions where this has been done they have always been able to subsequently successfully lease the stock.[88]

[88] Competition Commission (2009), 'Rolling stock leasing market investigation', April.

A2.4.3 Rolling stock for franchises

Rolling stock generally has an asset life in the region of 30 years—a duration longer than that of current franchises. Rolling stock is generally constructed by manufacturers, bought and owned by ROSCOs and leased to operators. Each of these steps involves an individual interface.

Acceptance procedures also involve a number of interfaces, covering compatibility (safety and inter-working) and Vehicle Change (commercial risk) procedures.

Bidders may have limited incentives to negotiate with ROSCOs, since (at least under previous regulatory arrangements) the ROSCOs had to make non-discriminatory offers to each bidder, limiting the scope for alternative arrangements between bidders. There is some amount of duplication in terms of multiple bidders negotiating with ROSCOs, however this will primarily be a cost borne by the private sector bidders and is likely to be mitigated by the benefits of competition.

In the ITT for a franchise the requirements of the service level commitment (SLC) may either explicitly or implicitly require a specific type of stock to be used for certain services.

Franchise length has been a widely discussed topic. The general view is that franchises would have to be of a significantly longer duration than the current standard of seven years, and perhaps as long as 18 years in order for it to be feasible for a TOC to take some form of risk on rolling stock procurement, unless an acceptable means of managing residual value were to be found.

There are some examples of different models of ownership that have led to investment in trains, such as that of LOROL. In this situation, TfL placed the order with the manufacturer but did not fund the transaction. Instead, the contract was transferred to a special purpose vehicle which financed it. Subsequently the new stock was sub-leased to LOROL to operate it.

Leasing and pricing

TOCs often have a limited pool of rolling stock to negotiate for, because other stock is tied up in ongoing franchises, and in any case, there will only be a limited pool of suitable and compatible rolling stock for any particular route. In theory this could raise prices, however, ROSCOs claimed in the Competition Commission (CC) market inquiry that there were competitive constraints on pricing from the risk of other ROSCOs displacing their fleets.[89]

This area of rolling stock leasing is clearly an interface in the industry. Negotiations do occur between TOCs and ROSCOs, sometimes incorporating fundamental aspects of the lease agreement. There is a tendency for lease agreements to cover the entire duration of a franchise, as ROSCOs charge a premium for leases covering less than this (since they may have to find an alternative taker for their rolling stock). This can contribute to the fixed nature of franchisees' cost bases—as the option of reducing a fleet to respond to an economic downturn is typically not bought by TOCs at the point where lease contracts are entered into.

Maintenance

Lease types can come in several different forms—dry, 'soggy' and wet.

– Dry lease—where the maintenance of the rolling stock is performed by the TOC.

– Wet lease—where the ROSCO remains responsible for maintenance or for new trains.

– Soggy lease—where the TOC is responsible for light maintenance while the ROSCO remains responsible for heavy maintenance. These leases are negotiated privately.

[89] Ibid., para 26.

One aspect of the CC market inquiry was whether capital rentals increased if maintenance were moved away from the ROSCO to the TOC. In other words, do TOCs believe that they are more efficient at contracting maintenance? Therefore by tying maintenance to the operating lease and preventing TOCs from contracting or doing maintenance themselves, there could be an inefficiency.

The CC inquiry found that for two of the ROSCOs, the capital charge elements of lease were the same for dry and soggy leases. For the remaining ROSCO, capital rentals did sometimes increase when moving from soggy to dry lease, although this was not deemed a strategic attempt to deter entrants.

A2.4.4 Rolling stock cascade

The DfT can be involved in the specification of cascades. The rolling stock cascade procedure is the planned dispersion of rolling stock across the network following the introduction of new rolling stock in one specific part, which means redistribution of existing stock that can be used on a different part of the network (assuming that it is compatible).

The cascade process may affect the range of fleets available for use on a given franchise, and therefore decisions about the cascade affect other parts of the industry. Any uncertainty around the cascade can create planning difficulties for ROSCOs, and the details of the cascade also affect Network Rail—eg, if EMUs will be cascaded then this may affect the prioritisation of infill electrification.

Government is often unwilling to specify rolling stock cascades in detail at any one time (to some extent because it does not wish to guarantee to ROSCOs where their stock will be used going forward). Since ROSCOs have every interest in ensuring that operators use their stock, it would seem to make commercial sense for them to be taking cascade decisions rather than government, except perhaps in the case of procurement of large new fleets.

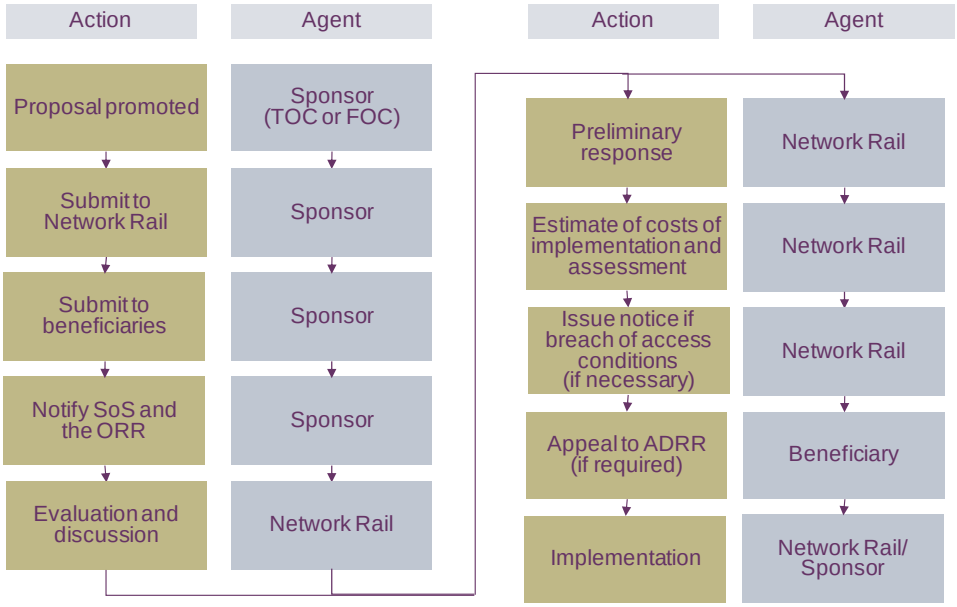
A2.4.5 Vehicle change and acceptance

Introducing new stock to the network requires compliance with the following processes.

- Vehicle acceptance—acceptance is a formal procedure that must be completed in order to introduce new rolling stock to the network:
  - a Vehicle Acceptance Body (VAB) that will apply RSSB's process for engineering acceptance of a vehicle such that it complies with the Railway Group Standards;
  - the focus of this process is on safety.
- Vehicle change—this process must be gone through in order to make any changes to vehicles that are permitted in the access contract:
  - condition F1 of the Network Code requires Network Rail to facilitate vehicle change—it is unclear what its commercial incentive to do so is;
  - the main aim of this process is to prevent adverse commercial effects on beneficiaries (through, for example, worse acceleration or breakdown performance arising from the new stock).

Figure A2.7 shows a visual representation of the vehicle change process and the range of steps and entities that are involved.

Figure A2.7 Vehicle change procedure



Source: Oxera, based on Network Code Part F.

Both of the processes mentioned above are lengthy. In one stakeholder interview, a FOC involved in these processes expressed the opinion that there is a significant degree of duplication between the two.

A2.4.6 Net industry cost and incentives

One of the main findings of the CC's market inquiry into the rolling stock market was that the operation of the franchise system limits opportunities for competition in rolling stock provision. There can therefore be limited incentives for a TOC to propose new rolling stock for a franchise because franchises are currently awarded by comparison with base case specification. There have even been occasions when the DfT has ruled out new rolling stock in franchise specification.

Vehicle acceptance procedures have been highlighted as having led to improved investment coordination. However, concerns have been raised at potential duplication in the process between compatibility (safety and interoperability) and vehicle change (commercial impact) procedures.

Incentives on franchised TOCs to obtain the best deal on rolling stock seem to be limited by a lack of bargaining power at franchise replacement, and during the life of the franchise. The DfT 'hard-wiring' of rolling stock fleets within franchise contracts contributes to this problem.

Negotiations between TOCs and ROSCOs are private sector negotiations and therefore this interface has no direct costs to the taxpayer. There are possible indirect costs from TOCs' poor incentives to negotiate, knowing that the same terms will be offered to all bidders.

There are some other possible indirect (non-cash) costs from sub-optimal specification of stock in franchise agreements. It is possible that TOCs would prefer the flexibility to choose stock themselves rather than meet the DfT's requirements and that they would be better placed to respond to market signals over the most suitable and desirable stock to use.

The cascade process can create uncertainty across the industry, and might become more market-driven than centrally planned.

A2.4.7 Conclusions

Many of the issues relating to rolling stock are a by-product of the current franchising system.

– One of the simple wins could be reforming the vehicle change and acceptance procedures to ensure that there is no duplication.

– Relaxation of the specification of rolling stock by the DfT may also generate benefits to the industry.

– Involving Network Rail in the design process (given that the stock will have an impact on its infrastructure) as a contractual (licence) requirement is another potential improvement.

There are several ways in which incentives might be improved in relation to rolling stock. These include:

– incentivising TOCs to reflect market conditions in their procurement and use of rolling stock throughout the franchise—this would involve a move away from centrally planned procurement and cascades, except in clear cases of market failure.

– expecting TOCs to pay a premium to ROSCOs to enable fleets to be increased and decreased during the life of the franchise in order to enable their lease costs to vary more with demand.

A2.5 Franchising

Franchise agreements are contracts that establish the role of the passenger operator in the industry. Many of the industry's interfaces arise as a result of these agreements. This review has identified these agreements as the cause of several incentive misalignments.

The importance of getting the franchising process right has been recognised by the Department for Transport (DfT), with the issuance of a consultation document on the reform of rail franchising.[90] The conclusions of that consultation will be considered alongside the Value for Money (VfM) Review final report before the Comprehensive Spending Review this October. Both of these sources will contain recommendations for the franchising system and there has been co-ordination across the teams to ensure that there is consistency in the messages arising.

The following sub-sections explain the current franchising system (and not any future changes) and its interfaces, before addressing the various alignments and misalignments of incentives.

A2.5.1 Current structure

Government authorities, primarily the DfT but also other devolved administrations, have the power to let passenger rail franchises (the right to operate rail services in a given area) to private sector companies by competitive tender. Each tender has a sole winner which enters into a formal franchise agreement with the relevant authority, signed by the TOC and the Secretary of State for Transport. The winning TOC then operates the passenger rail services on the franchise's routes for the specified length of time, typically seven to ten years. The government authority that lets the franchise is responsible for monitoring and enforcing the franchise agreement once awarded.

Currently there are 19 passenger rail franchises in Great Britain of differing durations, geographical scope and density. For example, the Essex Thameside franchise is a small franchise in the South East, mainly serving short-distance commuters, whereas the New Cross Country franchise covers the length of Great Britain and serves both short- and long-distance travellers.

[90] DfT (2010), 'Reforming rail franchising', July.

In exchange for the right to operate the services, franchisees collect the farebox revenue paid by passengers who travel on their services. The nature of the rail network and the franchise agreement offers a degree of protection from changes to other franchises. There is a degree of competition from open access (and other franchised operators), although there will always be at least one interoperable ticket available, which protects the operator from on-track competition in most cases. TOCs either receive a subsidy from the franchising authority or pay a premium, depending on the financial arrangements of their franchise bid.

Objectives

One of the objectives/justifications for the current franchising system is that private sector delivery ought conceptually to provide incentives for cost efficiency. The competitive tendering process should therefore achieve the minimum subsidy/maximum premium consistent with a deliverable franchise and acceptable risk profile for the DfT, thus minimising the taxpayer burden. The current existing risk-sharing mechanism also contributes to a reduction in the taxpayer burden, by protecting bidders from certain aspects of risk and thus lowering bid margins.

Franchising also allows the DfT to retain a degree of control over the services that are provided and therefore to achieve socially desirable outcomes that would not be provided by the private sector in isolation.

The following sections review the processes and interfaces involved in more detail.

A2.5.2 Tendering stage

The tendering process for franchises involves several stages and many parties, but is initiated by the DfT,[91] which runs the process.

Following DfT Board approval of the overall franchising programme, DfT project initiation begins with a Project Initiation Document (PID), which sets out the objectives for franchise specification.[92] The franchising process begins with the announcement of the tender and the assessment of pre-qualification by the DfT. The announcement of the tender must be issued in the Official Journal of the European Union. Up to this point, the main interactions are between the DfT and those that it consults during the specification stage.

Next, the pre-qualification stage aims to narrow the list of prospective bidders to around three to five who will be issued with an invitation to tender (ITT). At this stage, the DfT begins to interact with prospective bidders.

The ITT will specify minimum requirements for a bid to be compliant, both in terms of the contents of the bid itself and the proposed operation of services. The ITT can also include 'Priced Options', which can be any form of additional commitment (eg, extra frequency, different service quality). Each bidder must evaluate and submit an individual price for these options, although the DfT is not subsequently obliged to ask for the Priced Options to be implemented.

The tendering process itself involves many interfaces. Bidders need to negotiate with the rolling stock leasing companies (ROSCOs) regarding the rolling stock to be used on the franchise and the form of lease they will undertake. There will also be discussions with local authorities or PTEs regarding their requirements from the franchise and potential extra funding for additional components. Bidders will also need to hold discussions with Network Rail regarding operational issues or issues concerning depots and stations.

During the franchise process, a data room is made available by the DfT and accessed by bidders. It contains commercially sensitive information from the incumbent. There is a formal process for issuing questions to the DfT and responses being released to all participants.

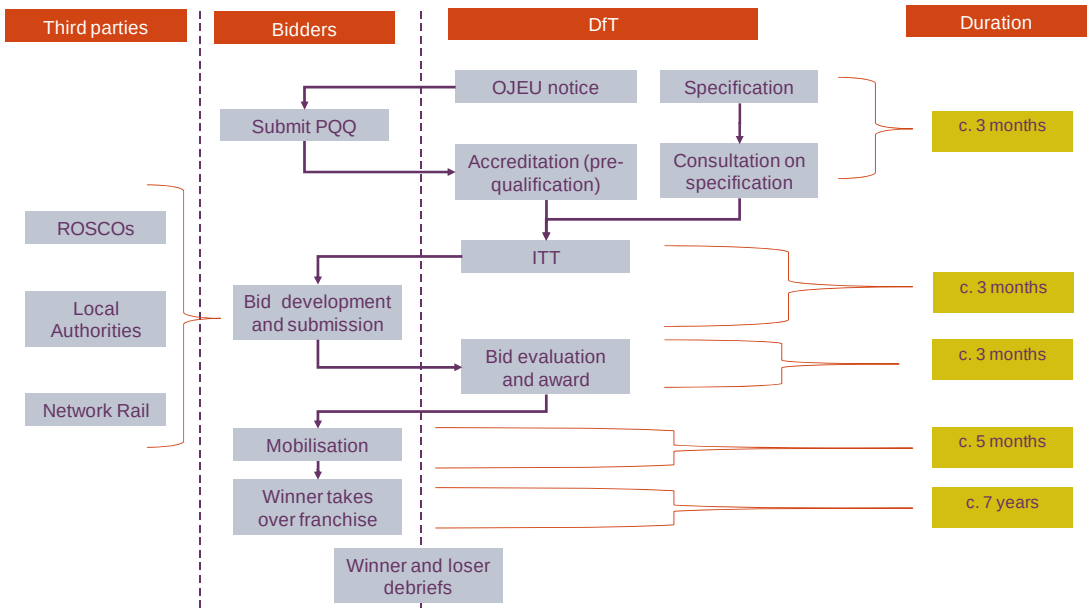
[91] Or another franchising authority.
[92] DfT (2010), 'Rail franchise process manual', January.

At this tendering stage there is a large duplication of interfaces between bidders and third parties and also with the DfT, but these are part of the competitive process as bidders seek to obtain competitive agreements for their inputs.

The assessment of franchise bids is also conducted by the DfT. Award is primarily based upon the risk-adjusted net present value (NPV) of the bid, providing that all requirements from the specification are met (including deliverability) and also that the DfT believes that the bid value is achievable. The risk adjustment to the NPV is applied by the DfT if the exogenous factors used by the bidder differ substantially (without justification) from the factors used in the DfT's comparator model. Where there is no material difference between the financial bid the award is determined based on the scores for the deliverability plan or other factors.[93]

Franchises are generally awarded on the base case specification, and only afterwards are variants to the base case considered and determined desirable or undesirable. A dedicated team of officials, the Contract Award Committee (CAC), decides which bidders should receive the ITT and who the ultimate winner of the franchise will be.

Figure A2.8 Example of tendering interfaces



Source: Oxera.

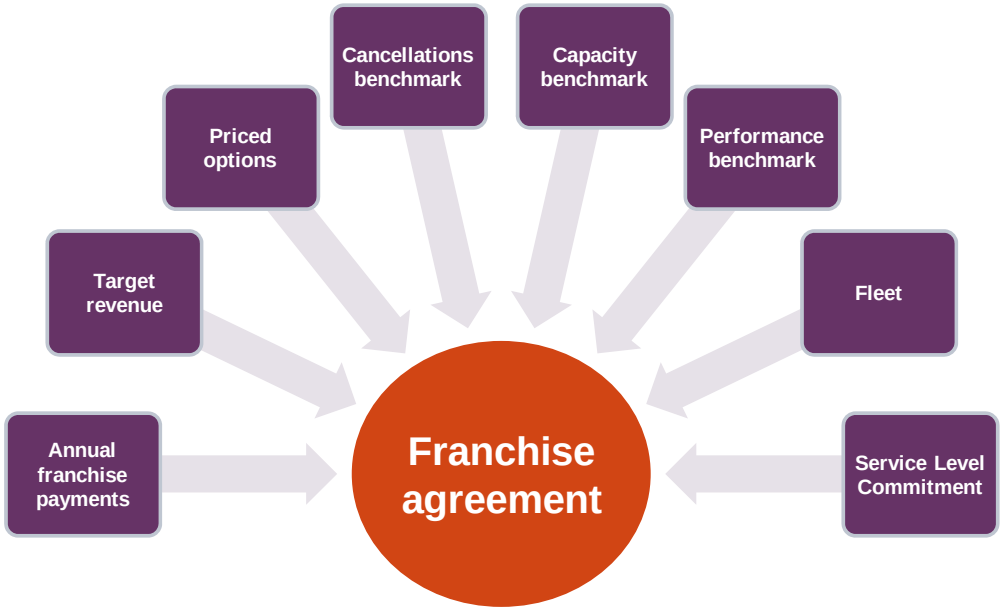
Following the award of a franchise through the tendering process, the TOC will mobilise to take over the business, having entered into a franchise agreement with the relevant authority.

A2.5.3 What does a franchise agreement contain?

Franchise agreements themselves contain a high degree of detail on the levels of service that are required from the operator, as specified by the relevant government authority. Figure A2.9 gives a stylised representation of some of the main components of a franchise agreement, setting out areas where the franchisee will interface with the franchising authority.

[93] DfT (2010), 'Franchise evaluation process charts', January.

Figure A2.9 Contents of a franchise agreement



Note: This is not an exhaustive representation of the contents of the franchise agreement. Some of these components are incorporated via separate documents or the National Rail Franchise Terms.
Source: Oxera.

The franchise agreements contain within them the financial details of the franchise, including the agreed level of franchise payment, which will have been used as the primary decision criterion for award of the franchise. The target revenue associated with these franchise payments is also included to inform the revenue share/support arrangements when they become active. These revenue share/support arrangements are currently broadly consistent across all TOCs. In addition, details of the performance and season ticket bonds will be included in the franchise agreement.

Several benchmarks which the franchisee must ensure its services meet are included in the franchise agreement. These include the cancellations benchmark, the capacity benchmark and the performance benchmark. If the operators' services do not meet the standards as specified in the agreement then the DfT (or equivalent authority) can issue a breach notice and require the franchisee to take action to rectify the situation. This action can involve requiring the operator to spend a specified amount of money on additional compensation for passengers, additional staff or extra capacity.

Most franchises contain breakpoints at which franchises can be terminated if they have under-performed. Alternatively, if the appropriate levels of service have been provided then the franchise can be extended for a pre-specified amount of time, usually a further three years. The decision about whether to extend a franchise is taken by the DfT, although there are examples where the franchisee can choose to terminate at the breakpoint.

The number and timing of the services that the TOC must operate are not usually specified in the form of an entire timetable by the DfT, but rather via minimum frequencies, maximum journey times or access requirements. These are set out in a SLC. TOCs can make commercial decisions on additional operations above these minimum levels of service.

All franchise agreements between the Secretary of State for Transport and a train operator incorporate by reference the National Rail Franchise Terms (currently the third edition).

There are several benefits of this relatively tight specification of services. First, by specifying services in this way the government can achieve socially desirable objectives, such as social inclusion, and prevention of severance of communities. In addition, there are several positive externalities from rail travel, such as decongestion and environmental benefits, which a

private operator may otherwise not take into account when determining its level of operations.

Opposing these benefits the cost of specification is that it prohibits innovation by the TOC and also limits its ability to respond to market signals—eg, by reducing under-utilised services.

A2.5.4 Franchise operation stage

Following the award of a franchise to an operator, the operator takes on a large number of interfaces through which it interacts in its day-to-day operations.

First, at the operational level there are interfaces between TOCs in terms of the revenue allocation process (undertaken via ATOC and the ORCATS system), and contracts such as station access agreements at the stations where the TOC is the SFO. Revenue allocation is complicated in the London Travelcard area, where there is a wider Travelcard allocation process involving the TOCs and TfL.

There are interfaces between the TOC and Network Rail, some of which are covered more specifically in the sections addressing Possessions and Performance. The formal contractual interface between a TOC and Network Rail is in the form of the Track Access Agreement, which sets out the conditions for access to the infrastructure, including the charges to be paid for access rights and the compensation arrangements for inability to temporarily access those rights. Network Rail also interacts with TOCs at an operational level through the operation of the networks signalling and through dispatch teams at some Network Rail-managed stations.

Importantly, changes to Track Access Agreements due to the Periodic Review process are nullified via the Franchise Agreement (on a 'no net loss, no net gain' basis) to protect TOCs against large swings in payments to or from Network Rail resulting from the process. However, this facet of the Franchise Agreement has considerable consequences for incentive alignment across the TOC/Network Rail interface, as infrastructure costs are, essentially, excluded from the commercial interest of the franchisee by such arrangements.

The extent to which there is an ongoing interface with a ROSCO depends in part on the precise form of lease that has been adopted. With a wet lease[94] there will be greater interaction with the ROSCO or its sub-contractors in order for maintenance to be carried out. On a dry lease the TOC will have separate interfaces of its choosing with its own maintenance firms.

There are numerous ongoing financial interfaces including the payment (receipt) of premium (subsidy) to (from) the DfT. This interface is complicated by the existence of risk-sharing mechanisms, designed to share some of the risk between government and the private sector operator. Some current franchise agreements contain cap-and-collar mechanisms for revenue sharing, with differing threshold rates at which the mechanisms apply and different shares.

Another interface comes from the DfT regulating certain fares that are set by TOCs (such as season tickets).

A2.5.5 Franchise change process

Franchise agreements are specified at the beginning of a franchise contract and are not routinely changed subsequently. Changes to franchises can be made either at the re-letting phase or, if they exist, by enforcing terms in the original franchise agreement. Otherwise the DfT must intervene and renegotiate the existing franchise agreement to make significant changes. Examples of franchise changes that can be specified by government include infrastructure enhancements, changes to rolling stock and changes to fares and ticketing.

[94] A lease where the ROSCO is responsible for maintenance.

Many of these forms of intervention require government to influence commercial contracts—eg, between TOCs and ROSCOs, which could harm the efficiency of private sector delivery.

The DfT is currently working on 14 interventions to provide additional capacity to the rail network.[95] The restrictiveness of franchise change procedures may act as a barrier to swift implementation of certain socially desirable outputs from the rail industry.

A2.5.6 Net industry cost and incentives

Tendering stage

As franchises are primarily awarded by the size of the bid premium/subsidy bidders are incentivised to bid aggressively, providing the DfT believes that their projections are realistically attainable and their plans for the franchise are deliverable with an acceptable risk profile—the DfT will not credit undeliverable bids. In theory this incentive allows the government to extract the maximum from the bidders and thus limit taxpayer support to the industry. However, there is also the possibility of a 'winner's curse' with winning bids systematically being 'too high', and hence leading to franchisees that struggle to deliver their commitments on franchise payment and ultimately end in default (eg, GNER, Connex). The strength of this incentive for over-bidding depends in part on the management incentives at each bidder, whether they just wish to win if the franchise will be profitable, or in order to increase the scale of the business, or for prestige reasons.

The cost of awarding franchises to bids that turn out to be unsustainable comes from the costs of having to re-franchise (even though there is a performance bond that the DfT can claim in the event of effective default by a franchisee). Lord Adonis, former Secretary of State for Transport, has been quoted as saying: 'The Government does not renegotiate franchises'.[96] Such a position can help to ensure that bidders do not excessively over-bid and that franchisees cannot simply abandon their commitments to franchises when things do not go according to plan.

Incentives to innovate against the ITT are limited due to tight specification of the core requirement. This is an area that is being consulted upon by the DfT in its franchise consultation.

There are also limited incentives to innovate significantly on rolling stock procurement. This is partly due to a previous requirement for ROSCOs to offer the same conditions to each bidder. Thus any tough negotiation on behalf of the TOC would not necessarily increase their chance of winning the franchise, as the same terms would have been offered to other bidders. Equally, franchise renewals tend to be staggered, leading to only a small pool of compatible existing rolling stock being available for use in the new franchise.

One element of the bidding phase is the assessment of external drivers of demand. These external drivers are a key element of determining the amount of revenue to bid. The DfT has its own assessment of the level of these external factors developed as part of its bid evaluation. During the bid evaluation, if the bidder has not provided comprehensive evidence that the DfT's benchmark should be adjusted, then the DfT's own factors will be applied to the bid for the purposes of award. However, following award, the successful bidder will be held to the original bid values when calculating the actual premium/subsidy payments. The bidder carries the risk of generating the additional revenue without receiving any benefit for it in the bid evaluation. There is also an incentive for bidders to get as close as possible to the DfT's comparator model, rather than aiming to provide the answer they genuinely believe to be correct.

[95] NAO (2010), 'Increasing Passenger Rail Capacity'.
[96] Statement of Lord Adonis to the House of Lords, July 1st 2009.

A further disadvantage of the tendering process is that any inescapable inefficiencies arising from the structure of the industry and affecting all bidders will be passed on directly to the DfT in reduced premium/higher subsidy.

Franchises are restrictive because TOCs are limited in their ability to run them as 'normal' businesses, such as by altering prices to match economic conditions or raising prices to fund investment, or removing services during a downturn.

The risk allocation between operations and infrastructure incentivises a disconnect in behaviour. The protection from changes in access charges means that TOC operations do not reflect any changes in underlying costs of the infrastructure that the services use.

Franchise operation stage

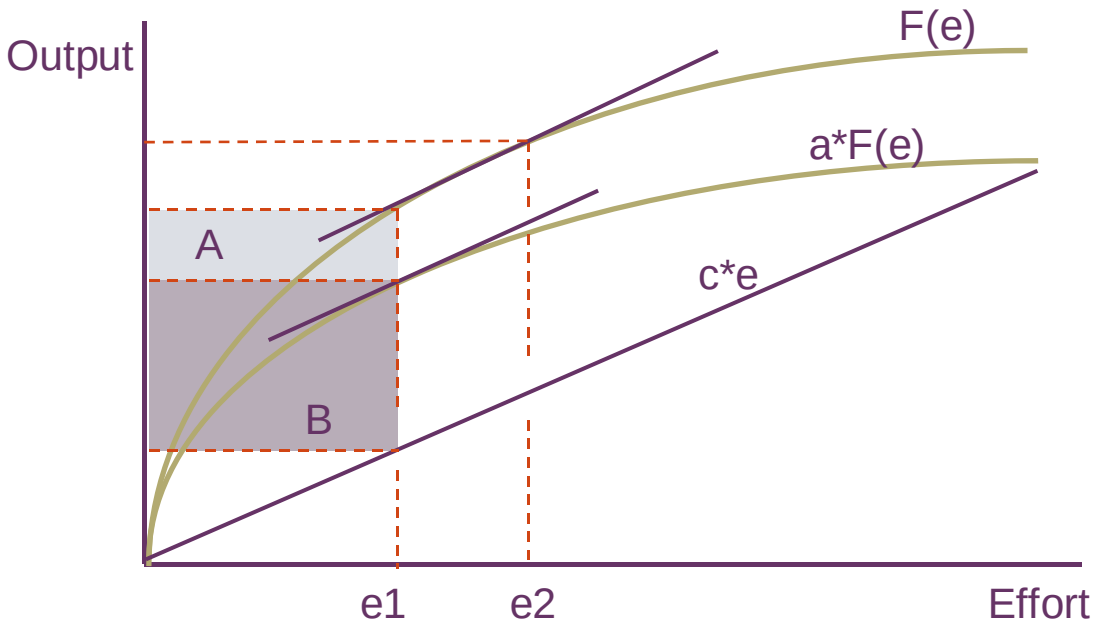
The main incentive for TOCs when they are operating services is to grow revenue and reduce their costs. However, high operational gearing means that the scope to cut costs is constrained. This is quite separate from any incentives on Network Rail to try and reduce costs.

Franchise agreements have a degree of risk sharing built into them that normally becomes active after four years. The 'cap' part of a revenue share arrangement limits incentives for revenue maximisation. There is no distinction made between revenue generated as a result of external drivers of demand growth, and revenue generated by the performance of the TOC itself.

The current cap-and-collar mechanism provides a degree of revenue protection for franchisees after approximately four years of the franchise. However, this form of risk-sharing mechanism limits the incentives on a TOC because in a revenue share context it does not receive the full percentage of the revenue generated by the franchise. If the revenue is below target (and under current arrangements, a considerable downturn in revenue relative to target can mean that the DfT takes on 80% of the revenue shortfall), the TOC can be assured that the revenue support mechanisms will protect it from major losses; conversely, if the financial performance is above target, the TOC knows that it will not receive the full benefit from its efforts to grow revenue.

In economic theory, an economic agent will only exert 'effort' up to the extent that the marginal cost of its 'effort' is equal to the marginal benefit/product. By engaging in revenue share arrangements, the marginal benefit accruing to a firm can be distorted. Under revenue share the marginal benefit to the firm may equal the marginal cost of effort at a lower level of effort than without the revenue share and this can lead to output being below the economically efficient level. Figure A2.10 shows a stylised economic analysis of the effect of revenue sharing that is usually applied to goods markets.

Figure A2.10 The impact of revenue share arrangements in a goods market



Note: F(e) represents the total production function of the firm. The firm receives only a proportion (a<1) of the output that it produces, the rest is shared with another entity. Hence the relevant production function for the firm's decisions is a*F(e). The firm chooses the level of 'effort' (e) to exert at a constant cost (c). Assuming a constant price, the firm therefore chooses to produce where the marginal cost of effort is equal to the marginal product it receives (ie, where the slope of the arc a*F(e) is equal to the slope of c*e). The producing firm receives an economic profit of area B and the firm that receives the share of output (1-a) receives a profit equal to A. This outcome is economically inefficient because the optimal level of the good to be produced would be at e2 where the slope of the total production function is equal to the marginal cost of effort.
Source: Oxera.

In the context of the rail industry, there are limitations on the TOC's choice of the level of inputs it may use as a result of its SLCs; although, the same economic principle is in operation—that if an agent receives only a proportion of its marginal product it can lead to a less than socially optimal level of effort being exerted. However, this analysis ignores any benefits resulting from the greater certainty that revenue sharing/support provides for TOCs and the government.

Perverse incentives can arise from such an arrangement. This includes TOCs being disinterested in revenue growth schemes (since the DfT would benefit more than the TOC in the 80% revenue share band), and even cost-conscious to the extent that it harms revenue-growth prospects (since it obtains all of the benefit from cost reductions, but only 20% of the additional revenue).

The relatively short duration of franchises also limits the investment incentives for TOCs.[97] There is an incentive to front-load investment projects in the franchise and later to allow assets, particularly stations, to fall into dilapidation towards the end of the franchise period. This is true of all assets with lives longer than the franchise, and to some extent true of operational performance given the lags in response in passenger demand to these operational factors.

Bid submissions are expensive for bidders; they have been estimated to cost in the range of £3m–£5m per bid.[98] However, these costs will fall to the private sector bidders. Whether

[97] However, it does not necessarily follow that TOCs will invest if they are awarded longer franchises. Instead, longer franchises would create the conditions in which the private sector would be more likely to invest in the long-term future of the business, subject to the terms of the franchise agreement.
[98] NAO (2010), op cit. citing ATOC.

losing bidders 'price in' the costs of previously unsuccessful bids to contracts is unclear, but this would be unlikely given that there are different bidders for each competition and attempting to price in previous bid costs would disadvantage any given bid.

However, if bid costs are not priced in, there is a question as to how they are funded. If bidders are parts of groups they may be subsidised by operations in other countries or in other sectors, but in theory this is only possible if these other markets are not perfectly competitive and hence the bidders have made economic profit they can afford to spend on ex post losing GB rail franchise bids.

Franchise procurement and bid evaluation at the DfT has a direct cost to the public sector, although this does not appear to be a significant cost.

A2.5.7 Conclusions

The process of franchising has brought a degree of private sector competition into the rail industry. However, franchise agreements are complex and lead to many interfaces, across which exist considerable possible misalignments of incentives.

The current franchise agreements are designed to transfer risk, especially infrastructure-related risk, away from operators. This has the effect of keeping bid margins low and thus minimising subsidy payments. However, this removal of risk from the TOCs has several adverse consequences, not least contributing to the separation of the industry P&L between TOCs and Network Rail.

Moreover, the current franchise financial 'model'—with considerable exposure to the macroeconomy for the first four years on both the revenue side, and the cost side (costs are predominantly fixed with respect to volume—means that bid margins should (conceptually at least) be very high. In financial economic terms, franchisees have high 'beta' risk, leading them to have high margin requirements to cover investor exposure to undiversifiable risk.

The cap-and-collar risk-sharing mechanism limits some of the incentives on operators. Revenue sharing limits the effort that the TOC will put into the franchise because it is not receiving the full value (marginal product) of its effort.

One option could be to apply a risk-sharing mechanism where TOCs receive the full marginal product associated with their effort (ie, separating revenue generated by exogenous and endogenous factors and allowing TOCs to receive 100% of endogenously generated revenue). While increasing complexity, this would increase the incentives on TOCs to invest the optimal amount of effort in their franchises.

There may be an incentive for a franchisee to 'run-down' the franchise when it nears the end of its life, as the prospect of not retaining the franchise following the re-tendering process may mean that any investments undertaken by TOCs would not be fully recouped.

A2.6 Development of the timetable

This section explains how the timetabling process is developed and the incentives facing the parties contributing to that process. This section also explores the interfaces with, for example, the open access operators and how they fit into the timetable development process.

A2.6.1 Current structure

The timetabling process is overseen by Network Rail; Part D of its Network Code sets out the procedures by which changes can be made to:

– the working timetable;
– Rules of the Route;
– Rules of the Plan.

Although changes can be made to the working timetable at any time, significant changes in the passenger timetable may be made only twice a year—at the dates referred to as the Principal Change Date (in December) and the Subsidiary Change Date (in May).

The development of a robust timetable demands dialogue between:

– Network Rail, TOCs and other organisations that are entitled to take part in the process;
– TOCs and their customers or customers' representative bodies.

More details on this process and associated interfaces are described below.

A2.6.2 Process and interfaces

At least two years before each Principal Timetable Change Date, Network Rail, in collaboration with operators, will commence preparation of a base timetable for that year, which will incorporate anticipated changes into a single unified timetable plan.

In terms of the interfaces between the parties, Network Rail has the role of managing the working timetable. It is responsible for accommodating within the timetable the contractual service specification of each train operator. This contractual specification is set out in the franchise agreement between each operator and the DfT.

The specification will normally allow a degree of flexibility to both Network Rail and the train operator, both in terms of the timing and other characteristics of the services.

A train operator's train slots are protected insofar as they are based on firm rights, which are not inconsistent with the applicable Rules of the Route and/or Rules of the Plan, provided that the firm rights have been asserted no later than the Priority Date.

Each year, at the start of the timetable development process, Network Rail is obliged to review the applicable Rules of the Route and Rules of the Plan and decide if any amendments should be made in respect of the period of the annual timetable commencing on the next Principal Change Date. In addition, each year, at the start of the process for development of the timetable changes applying from the Subsidiary Change Date, Network Rail is obliged to undertake a more limited review of the applicable Rules of the Route and Rules of the Plan.

Train operators are consulted on each review, and there is a right to refer disputes to the relevant dispute resolution panel within the Access Dispute Resolution Rules (ADRR) in accordance with a specified procedure.

The process and interfaces each year are as follows.

– At or before the start of the timetable development process, there will be dialogue between Network Rail and train operators regarding the base timetable content and any variations to those train services which the operator aspires to run in that timetable year.

– Each operator will notify Network Rail of any changes in the contractual rights (as set out in their access agreement with Network Rail) that they wish to exercise in support of these services, giving reasons for such changes.

– Operators will also notify Network Rail of any changes to the base timetable train slots which will be sought.

– Network Rail will convene an annual Timetabling Conference and invite all operators to attend to discuss openly the services which they seek to run in the following timetable year.

– The Timetabling Conference will include bilateral and multilateral dialogue between access parties and Network Rail.

– After the Conference, Network Rail will lead a joint industry process with operators to prepare and issue a draft timetable which will be developed from the base timetable, taking account of the consultation held before, during and after the Timetabling Conference.

– In respect of each timetable week, where Network Rail requires possessions or other restrictions of use in order to undertake engineering work, it will notify operators of the changes it proposes to make in the relevant week and whether it requires operators to submit revised bids for timetable slots for that week.

A2.6.3 Network Rail's role and responsibilities

As manager of the working timetable, Network Rail is required to make decisions about:

– whether to accept bids for new or different timetable slots;
– how to reconcile competing or conflicting bids;
– how to exercise any right it may have to flex a particular train operator's bids (to the extent that the train operator's service specification allows it).

Network Rail must therefore work to specified criteria when making decisions regarding proposed changes to the working timetable and to any applicable Rules of the Route and Rules of the Plan. These criteria are weighed and balanced by Network Rail in the light of the particular circumstances surrounding each decision. In certain circumstances, Network Rail must also consider whether it is reasonably practicable for proposed amendments to the working timetable to be developed and implemented in the time available.

There is a criterion that will enable Network Rail to preserve paths for subsequent spot bids where there is reasonable expectation that these will be required and utilised.

A2.6.4 Timetabling—summary of main organisations

To demonstrate the number of interfaces that occur to comply with the timetabling process, the following section summarises the main bodies involved.

– Network Rail: establishes and manages the process.

– Bidders: these are the TOCs and FOCs bidding to operate timetabled services.

– Network Rail accreditation staff/teams: needed to accredit timetable planners employed or engaged by train operators.

– Secretary of State for Transport/Scottish Ministers/Welsh Assembly government/Transport for London/Passenger Transport Executives/user representatives/any other parties with the right to be consulted: can be consulted by bidders regarding proposals for train services.

– ADRR Panel: if in dispute with Network Rail's timetabling decisions, bidders can make referrals to this panel.

– the ORR: for resolution of matters not solved satisfactorily by the ADRR Panel.

A2.6.5 Timetabling—incentives

Although the timetabling process is detailed within the Network Code, the number of interfaces, guidelines to follow, consultations and dispute resolutions make the system time-consuming and reliant on high levels of staff resources.

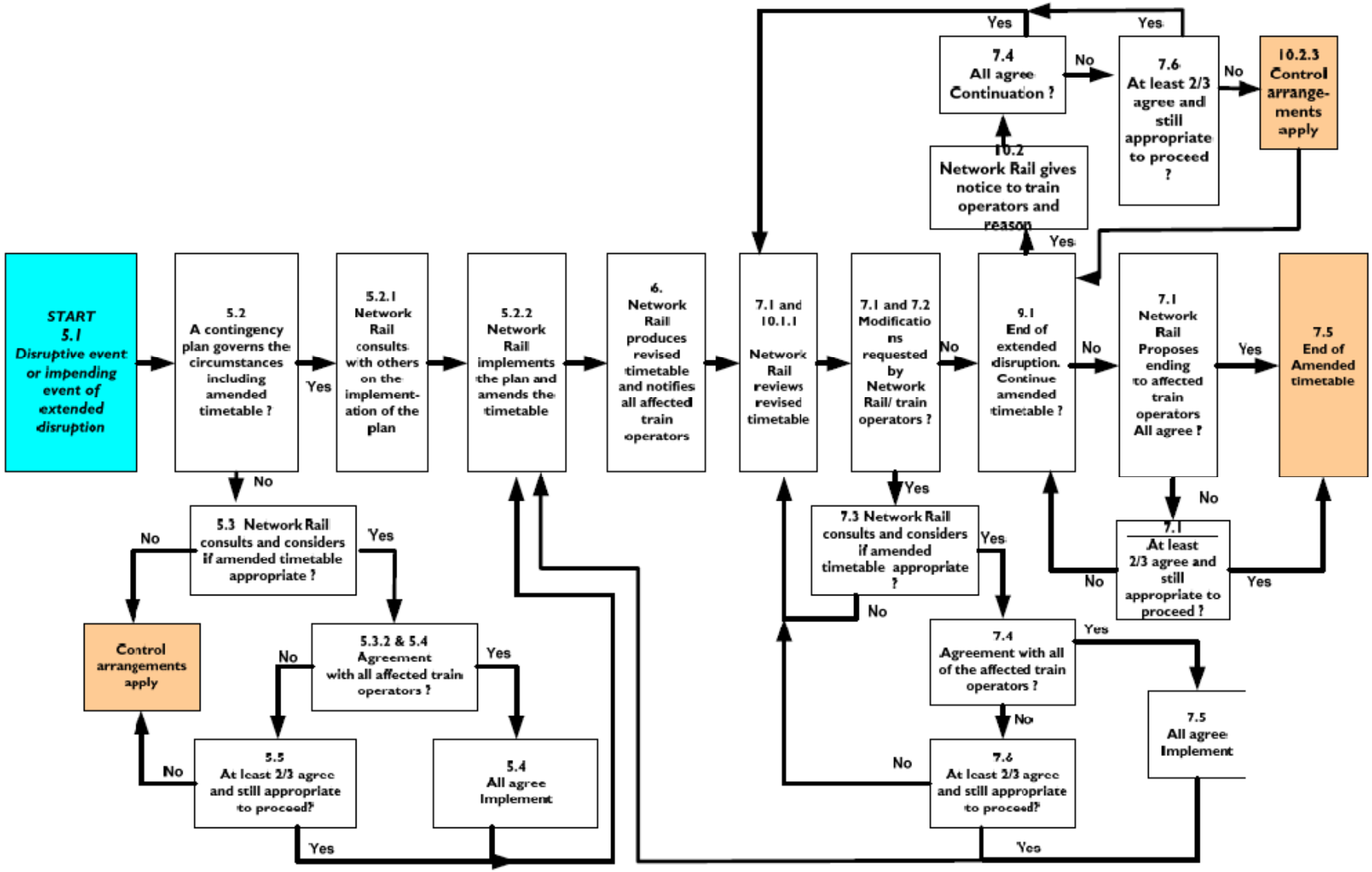
Network Rail is incentivised to complete the timetabling process within the timescales indicated in the Network Code. Despite this, there may be several misaligned incentives across the process as Network Rail's incentive to produce an operationally robust timetable

may cut across the incentives of each individual operator seeking to meet its own commercial objectives.

Emergency timetabling

If there is a disruptive incident or an impending event of extended disruption, Network Rail instigates a process involving all relevant train operators. The extent of the consultation process and interfaces between affected parties is illustrated in Figure A2.11 below.

Figure A2.11 Emergency timetabling procedure and interfaces



Source: Network Rail, Emergency timetabling in the event of disruption.

A2.6.6 Timetabling disputes resolution

In addition to the timetabling process described above, there are further processes and interfaces associated with timetabling disputes resolution.

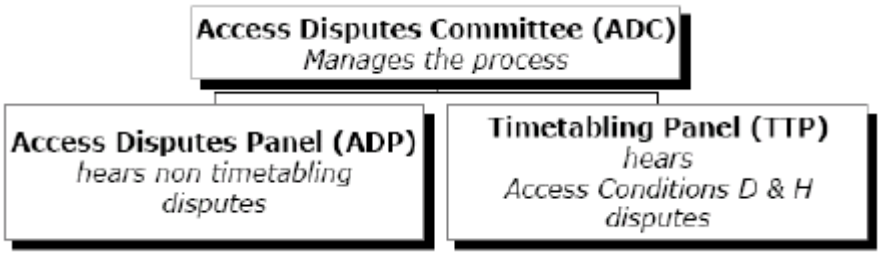
This process is set out by the ADRR, the purpose of which is to provide a clear, coherent, and effective structure for dealing with rail disputes arising under access agreements and access conditions/codes that must be dealt with under access-specific processes.

There are a large number of bodies, and hence interfaces, involved in this process:

– the Access Disputes Committee;
– the Access Disputes Panel (supervises overall process rather than deciding on the outcome);
– the Timetabling Panel.

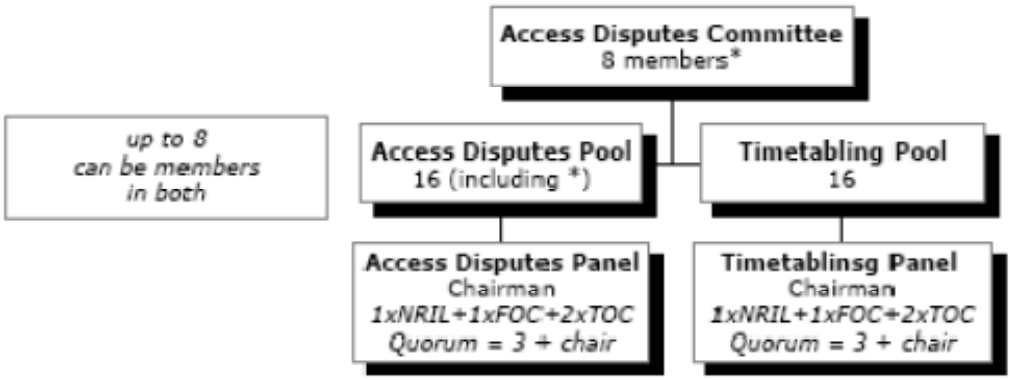
The structure of these bodies is shown below.

Figure A2.12 Disputes resolution—structure



Source: Annex to the Access Dispute Resolution Rules.

Figure A2.13 Disputes resolution—organisation of parties



Source: Annex to the Access Dispute Resolution Rules.

Box A2.2 Timetabling disputes: example

In December 2009, rail freight operator DB Schenker's Access Manager issued a notice of appeal to the ORR as it was dissatisfied with the decision of the Access Disputes Panel regarding two track access issues.

The overall process and timelines are summarised below.

– November 2009: Access Disputes Panel issues track access determination (hearing involves 12 representatives).
– December 2009: DB Schenker issues notice of appeal covering the Access Disputes Panel determination regarding two track access issues.
– December 2009: the ORR responds by issuing a letter to Network Rail, DB Schenker and Direct Rail Services (DRS).
– January 2010: DB Schenker issues a reply to the ORR's letter.
– January 2010: Network Rail representations.
– January 2010: DRS representations.
– January 2010: further ORR letter and update.
– February 2010: further DRS representations.
– February 2010: further Network Rail representations.
– February 2010: further DB Schenker representations.
– March 2010: final ORR determination.

For what was a relatively small track access dispute, the total number of parties involved was considerable, comprising:

– a Chairman;
– representative: non-franchised passenger TOC;
– representative: non-passenger operator;
– representative: Network Rail;
– representative: franchised passenger TOC;
– Network Rail: Customer Relationship Executive (DB Schenker);
– Network Rail: Customer Relationship Executive (DRS);
– Network Rail: Access Contract Policy Specialist;
– DB Schenker: Access Manager;
– DB Schenker: Business Manager;
– DB Schenker: Senior Solicitor;
– DRS: Commercial Development Manager;
– ORR: Track Access Executive;
– ORR: Deputy Director—Legal Services.

A2.6.7 Net industry cost and incentives

The following considers some positive aspects of the existing regime:

– It provides industry resolution of disputes in an adjudication style. In other words, the process functions as it is subject to an accepted form of adjudication between the parties.

– It reduces the learning curve, as the process is already well known in the industry.

– Timetabling planning generally works well.

– There is a robust set of rules, which are understood by all parties and have been successfully applied to previous disputes.

– It is free at the point of delivery as the costs of the process are covered by a levy on train operators.

– A body of case law has been built up, which will help all future access disputes.

In addition, there are a number of problematic features of the current regime:

– Panel member availability—given the number of parties involved with the panel-based resolution process, the unavailability of key members will delay the process.

– Practical constraints on the ability to hear large/complex disputes—the current process precludes consideration of larger disputes.

– The current system lacks flexibility.

The costs of the timetabling procedure can be placed into different categories:

– direct costs: timetabling staff costs (Network Rail, train operators, disputes panel, the ORR, etc);
– indirect costs: the costs of delays to the timetabling process (commercial implications of delays to train operators);
– opportunity costs: the costs associated with deploying highly trained staff on a drawn-out timetabling process when they could be more gainfully employed on other planning duties.

The costs of timetabling will differ significantly from route to route. For example, as the section on Capacity Allocation shows, routes such as the East Coast Main Line will have a far more complicated timetabling process compared with a route that is much shorter/less complex and where only one or a very limited number of train companies operate.

Some stakeholders expressed a view that a 'perpetual' timetable (eg, one that has the same timetabled train patterns from hour to hour) may be helpful in that there would be some modest staff reductions. However, the disadvantage would be that there would be no flex in the timetable to accommodate a new freight flow to a terminal.

A2.7 Long-term planning

There are three main processes of long- and medium-term planning in the UK rail industry:

– the HLOS—describing the government's output requirements from the rail industry in England and Wales in the medium term, and set out in its White Paper;[99]
– RUSs—providing medium- and long-term strategies for key routes developed by Network Rail;
– the Planning Ahead document—a joint document by Network Rail and the passenger and freight operators, setting out a long-term view of the industry.

In Scotland, Transport Scotland has devolved powers from the UK government, responsible for: managing the ScotRail franchise contract; funding Network Rail's delivery responsibilities in Scotland; delivery of rail projects, including infrastructure, rolling stock and timetable changes; and offering expert rail advice to inform policy choices.[100] As such, Transport Scotland is responsible for devising its own HLOS, separate to that of the DfT HLOS for England and Wales. An exception to this separation is safety requirements, as this power is not devolved to Scotland—Transport Scotland is required to ensure that safety requirements in the Scottish HLOS meet the minimum requirements in the DfT HLOS.

A2.7.1 Objectives

High-level Output Specification

The government White Paper, 'Delivering a Sustainable Railway',[101] sets out the government strategy for the railway industry for the long and medium term. Along with setting targets and

[99] Department for Transport (2007), 'Delivering a Sustainable Railway', July.
[100] See http://www.transportscotland.gov.uk/about-us/organisational-structure/rail-delivery.
[101] Department for Transport (2007), 'Delivering a Sustainable Railway', July.

improvements to meet the more immediate challenges, the government aims to set a strategy for the medium to longer term.

The government separates the medium- and long-term planning into three time horizons: to commit, to plan and to anticipate. In the current planning, the period up to 2014 is the commit phase, and includes the HLOS. Demand forecasts and other conditions can be expected to be sufficiently robust over this period, and as such, planning can be more specific and detailed.

The period 2015–24 is the plan phase over the medium term. As predictions of future demand growth are more uncertain, some flexibility is required to allow plans to adapt as forecasts become more reliable over time.

The final period, to anticipate, is beyond 2024, and as such, forecasts are attached with a high degree of uncertainty. Structural changes to the industry and the economy as a whole make future challenges hard to predict. Such structural changes include, for example, the effect on passenger demand of macroeconomic conditions, preferences towards inter-modal travel, changes to lifestyle and work attitudes, and technological developments.

Alongside the HLOS, the government lays out plans for infrastructure developments to meet current challenges and also to yield long-term benefits. Funding for these developments is separate to the SoFA.

Route Utilisation Strategies

RUSs outline the appropriate strategic direction of the railway industry, based on the future requirements of the network. The RUSs combine the medium- and long-term planning of the industry by incorporating the drivers of medium-term planning, such as the HLOS and franchise tendering, in addition to longer-term planning objectives for 19 strategic routes.

The planning horizon of the RUSs is over a 30-year period. This includes a detailed forecast of the first ten years, including contingency plans for forecast variations, as well as a more high-level plan containing scenarios to look forward a further 20 years beyond the initial ten-year period.

The development and maintenance of the 19 RUSs involves stakeholder engagement and consultation to ensure that they reflect the requirements of the whole industry. The RUSs are used as a driver for Network Rail's activity as part of its business plan.

The Planning Ahead document

The Planning Ahead document reflects the joint view of major industry stakeholders on the future of the rail industry; it is the result of the collaboration between Network Rail, TOCs and FOCs. It aims at planning ahead to CP5 and beyond, thinking about the industry in a wider context to inform decisions about future output requirements. It provides a high-level, strategic, long-term view of the industry to enable the industry to meet future demand and challenges.

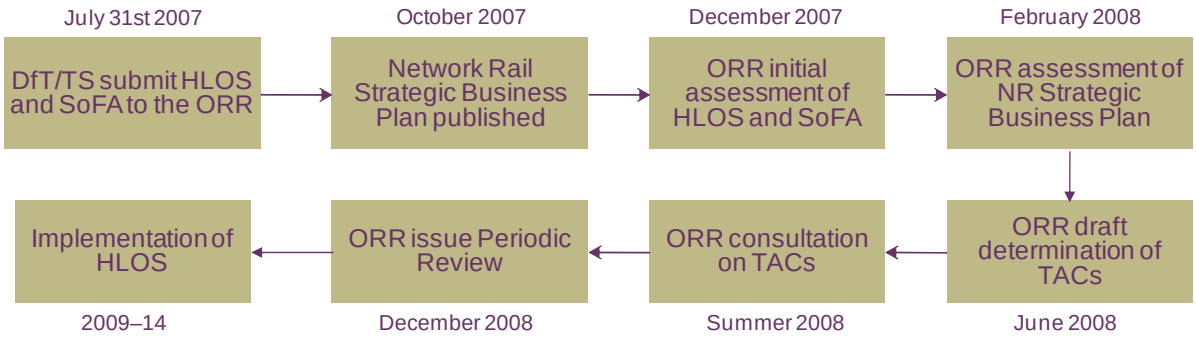
The Planning Ahead document is consultative, incorporating a wide range of views in the industry in order to portray a consensus account and vision. The document is intended to complement the more legislative documents—such as the ORR's periodic review, Network Rail's Business Plan, and the government's HLOS—providing viewpoints from across the industry, and a long-term strategic framework for the future.

Mapping the process onto its interfaces

The DfT sets its medium-term strategic outputs for the railway industry via a HLOS lasting for a five-year period, which it is required to send to the ORR. This is accompanied by a SoFA, which determines the industry-level government budget available to achieve this desired output.

Once the government has submitted the HLOS to the ORR, the regulator makes an assessment of whether these strategic outputs are achievable given the public funds available. Figure A2.14 provides a high-level overview of the current HLOS process.

Figure A2.14 HLOS process



Source: Oxera.

Delivery of the outputs that the DfT targets is the shared responsibility of Network Rail and the TOCs—Network Rail is responsible for delivering infrastructure outputs, monitored by the ORR; the TOCs are responsible for delivering operational outputs, under contract to the DfT.

Coordination is required at the interface between Network Rail and the TOCs, as in many cases, the outputs demanded in the HLOS can be achieved using a number of methods—either through infrastructure development or operational improvements. For example, passenger capacity could be increased either by TOCs increasing the number of carriages on a train, or by Network Rail delivering infrastructure enhancements.

The current HLOSs used three broad output metrics:

– safety;
– reliability;
– capacity.

By the 2014 HLOS, the government anticipates having gathered enough data and information to include an environmental metric.

Under the current HLOS, the following outputs are anticipated.

– The safety metric requires a 3% reduction in the risk or injury to passengers and employees. The responsibility of delivery falls on both Network Rail and the TOCs, while the Rail Safety and Standards Board (RSSB) monitors safety and reports back to the ORR.

– The reliability metric specifies an improvement in PPM from 88% to 92.6%, as well as a reduction of 25% in the number of trains delayed for over 30 minutes or cancelled. The industry has a well-established mechanism for working together, with Network Rail taking the lead in cooperation with the TOCs under Joint Performance Improvement Plans. This metric is very much output-focused and not prescriptive, allowing the industry to find the most effective, cost-efficient method of delivery.

– The capacity metric is an investment priority and requires developments to accommodate a 22.5% increase in passenger demand by 2014, in addition to a number of requirements relating to train load factors. This metric, while still offering Network Rail flexibility over delivery, is more prescriptive and has more intense government involvement throughout the delivery process.

The delivery of capacity requirements requires a high level of coordination between industry, led by Network Rail, and the government throughout the process. The main dimensions of this coordination include:

– Network Rail using the capacity options developed by the government to implement the requirements through discussion with TOCs and suppliers;
– Network Rail advising the government on the most appropriate and cost-effective way of achieving the increase in capacity through the regulatory process;
– the government negotiating amendments with the TOCs over changing their franchise agreements in order to meet capacity levels.

At a high level, the following interfaces are involved in relation to planning processes.

– **DfT and franchised TOCs**—government is required to negotiate franchise variations with TOCs when it is deemed necessary in order to deliver certain aspects of the HLOS. For example, in order for the government to deliver on its commitment to an additional 1,300 new carriages, made as part of the output target for the capacity metric in the HLOS, it must negotiate increases in the number of carriages that the TOCs provide for peak-time services and on trains with an above-target load factor.

– **Train operators and ROSCOs**—TOCs hire additional rolling stock from ROSCOs in response to changes to their franchise agreements in order to meet delivery improvements over the various planning horizons. The interface between TOCs and ROSCOs is dependent on decisions made by the DfT with regard to where the delivery of new rolling stock should be focused.

– **Network Rail and franchised TOCs**—coordination between Network Rail and the TOCs is required to achieve the coordinated and efficient delivery of the targets set out in the HLOS. However, this interface is subject to government involvement through the specifications of a TOC's franchise agreement. A further interface exists through the TAA, which sets out the charges that each TOC will pay to Network Rail, and which rail paths each TOC will receive access to in return. In addition, development of the RUSs and Planning Ahead document involves interaction and coordination between the two parties.

– **the DfT and the ORR**—the ORR assesses whether the output targets of the HLOS are achievable given the SoFA. In addition, the regulator outlines in the periodic review how it expects Network Rail to deliver the requirements outlined in the HLOS.

– **the ORR and Network Rail**—the interface between the ORR and Network Rail is through the implementation of the infrastructure output requirements of the HLOS. The ORR determines what costs Network Rail is allowed to incur in delivering the required outputs. These allowed costs are then offset against Network Rail's allowed income through access charges from operators, other sources of income, and expected efficiency savings from Network Rail. The ORR monitors and enforces Network Rail's compliance with the regulatory determination through the network licence. The network licence also requires Network Rail to establish and maintain RUSs.

A2.7.2 Existing incentives

Track access agreements

As has been seen above in the section on franchising, any changes to the TAA resulting from the Periodic Review (and, hence, the HLOS process) will be reflected in a 'no net loss, no net gain' condition in the Franchise Agreement. This means that TOCs have relatively little financial interest in the projects that Network Rail determines to be necessary to deliver the HLOS outputs, their scope and cost (as these are input into track access charges).

Franchise agreements

The key incentive in the franchise agreement with respect to planning is the Change function, which essentially leads to negotiations between the franchising authority and the TOC in respect of any material changes to the franchise agreement. As has been seen recently (and discussed in the NAO report on increasing passenger rail capacity), this can be a complex process, albeit one that interviewees have suggested works well in the end.

Network licence

The network licence is granted to Network Rail by the ORR. It sets out the conditions that Network Rail is required to meet in order to operate as the owner and operator of the national rail infrastructure.

The licence makes Network Rail accountable to the ORR, which is obliged to ensure that Network Rail is meeting its contractual obligations.

The requirements set by the government in the HLOS will be included in the network licence, implying that Network Rail is contractually obliged to deliver them. Failure to deliver can result in enforcement from the ORR, including enforcement action and monetary penalties.

A2.7.3 Net industry costs and incentive alignment

Overall, the stakeholder interviews conveyed the message that the long- and medium-term planning processes seem to work well. The processes are seen to have been quite productive over the past five to six years and most of the planning has been taken forward. Transaction costs are regarded as not being material.

One important benefit from the planning processes is seen to be to 'have the railway in one room', and it is the *process* that stakeholders go through that gives value because people gain understanding of the objectives and get involved.

Stakeholders consider that the RUSs broadly work well in terms of their impact on train operators, partly because franchises are generally shorter than the RUSs' planning horizon and therefore providing incentives to TOCs to engage in the process. However, concerns have been raised that longer franchises might make this process more difficult as they introduce TOCs as longer-term players whose (commercial) incentives do not necessarily coincide with the (societal) incentives of the RUSs.

In addition, it has been highlighted in the interviews that there might be an issue at the interface between network planning (RUSs) and the ORR's statutory duties, as laid down in section 4 of the Railways Act. Whereas the long- and medium-term network planning is based on the economic/societal value of projects as a whole (ie, WebTag assessment), the ORR's decisions are aligned with its statutory duties and hence is made on a different basis. The two decision frameworks of WebTag and Section 4 do not necessarily appear to be aligned for all decisions and, therefore, there might be potential conflicts if decisions are based on different planning rules.

A2.8 Renewals, enhancements, and the delivery of cross-industry projects

Scale of renewals and enhancement expenditure

Over the period 2009 to 2014, the planned expenditure by Network Rail on enhancement projects will increase significantly compared with previous years. Whereas the period from 2004 to 2009 (CP3) saw enhancement expenditure allowed by the regulator totalling around £2.7 billion,[102] the current control period from 2009 to 2014 (CP4) foresees enhancement

[102] Office of Rail Regulation (2003), 'Access charges review, Final conclusions', December.

expenditure of about £7.6 billion.[103] Moreover, this increase is particularly pronounced in the early years of CP4 as Table A2.4 shows.

Table A2.4 Renewals and enhancement expenditure from 2009 to 2014 (£m, in 2006/07 prices)

Year	2009/10	2010/11	2011/12	2012/13	2013/14	Total
Renewals expenditure	2,693	2,356	2,074	1,879	1,758	10,760
Enhancement expenditure	1,535	1,978	1,488	1,390	1,222	7,612
Renewals and enhancement	4,228	4,334	3,562	3,269	2,980	18,372

Source: Office of Rail Regulation (2008), 'Periodic review 2008. Determination of Network Rail's outputs and funding for 2009-14', October.

However, the increase in enhancement expenditure is partly offset by a reduction in renewals expenditure, which decreases from a total of approximately £14.6 billion in CP3 to about £10.8 billion during CP4, in 2006/07 prices.

Nonetheless, the increased enhancement programme can be considered to represent a significant challenge to Network Rail and the rail industry as a whole due to:[104]

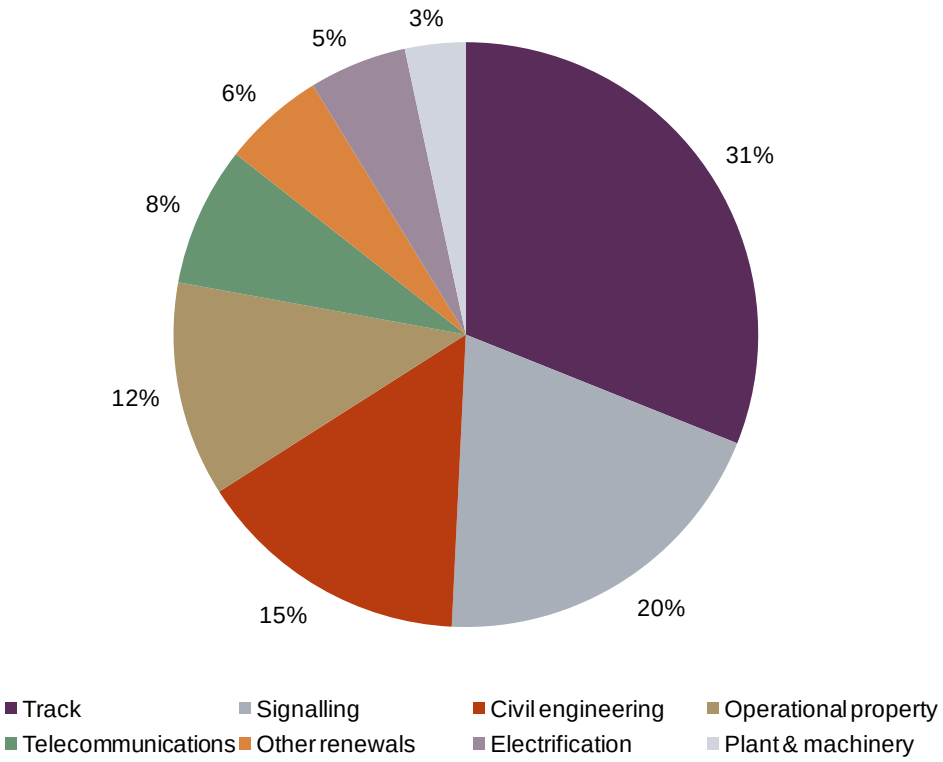
– the increased volume;
– enhancement projects being generally more complex than renewals, as schemes tend to be larger, more interdependent and multidisciplinary;
– many schemes needing external approvals and planning consent;
– multiple funding clients with which Network Rail must deal;
– a significant dependence by Network Rail on other parties—eg, rolling stock being procured by the DfT.

Drivers of renewals and enhancement

Renewals expenditure is needed to sustain the existing network. Expenditure is primarily driven by the wear and tear resulting from network usage—in particular, of tracks. Figure A2.15 below shows a breakdown of the main drivers of renewals expenditure over the period 2009 to 2014.

[103] Allowance for enhancement expenditure for CP3 and CP4 are given in 2006/07 prices. The figure for CP3 has been up-rated using input price inflation (IPI).
[104] See The Nichols Group (2008), 'Rapid review of Network Rail's capability to deliver its increased programme of enhancements', April.

Figure A2.15 Drivers of renewals expenditure (percentage of total planned renewals expenditure between 2009 and 2014)

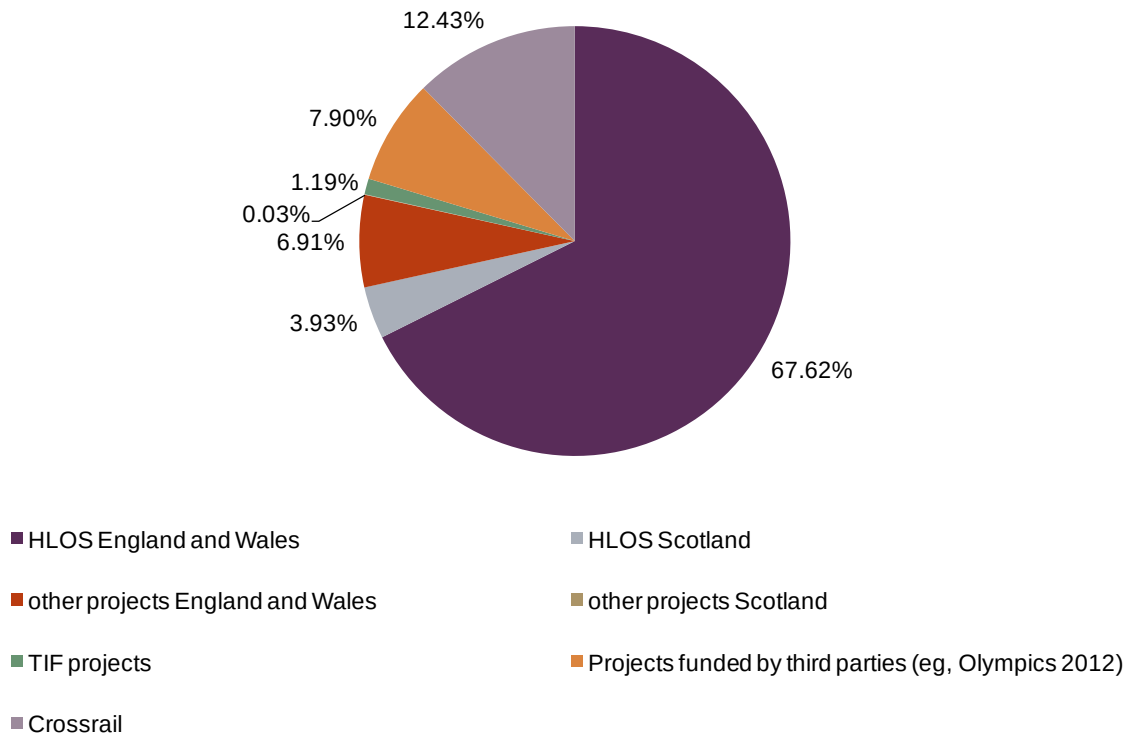


Note: Percentages are based on pre-efficiency allowances.
Source: Oxera calculations, based on Network Rail (2009), 'Regulatory Financial Statements. Year ended 31 March 2009'.

The majority of enhancement expenditure is related to the requirements of the HLOS. Key projects include the West Coast Route Modernisation, the Thameslink programme, and the Reading Station area development. Other important enhancement projects consist of Transport Innovation Fund (TIF) schemes, third-party funded schemes and Crossrail; the funding of these projects is outside the regulatory review process.

As Figure A2.16 shows, about 80% of planned enhancement expenditure is directly related to the output requirements of the two HLOSs.

Figure A2.16 Drivers of enhancement expenditure (percentage of total planned enhancement expenditure between 2009 and 2014)



Source: Oxera calculations, based on Network Rail (2009), 'Regulatory Financial Statements. Year ended 31 March 2009'.

A2.8.1 Objectives

The main objectives of Network Rail's renewals and enhancement expenditure can be split along the following two themes:

– meet the requirements of the HLOS;
– deliver other strategic, legislative or technical requirements.

These objectives can be further deconstructed into aims of:

– increasing safety;
– improving reliability;
– enhancing capacity;
– meeting objectives such as improvements in passenger experience and strategic network capability targets.

The safety, reliability and capacity objectives of the HLOS are operationalised through the HLOS metrics. The HLOS objectives are reflected in the regulatory price control review process and Network Rail's performance against them is monitored by the ORR.

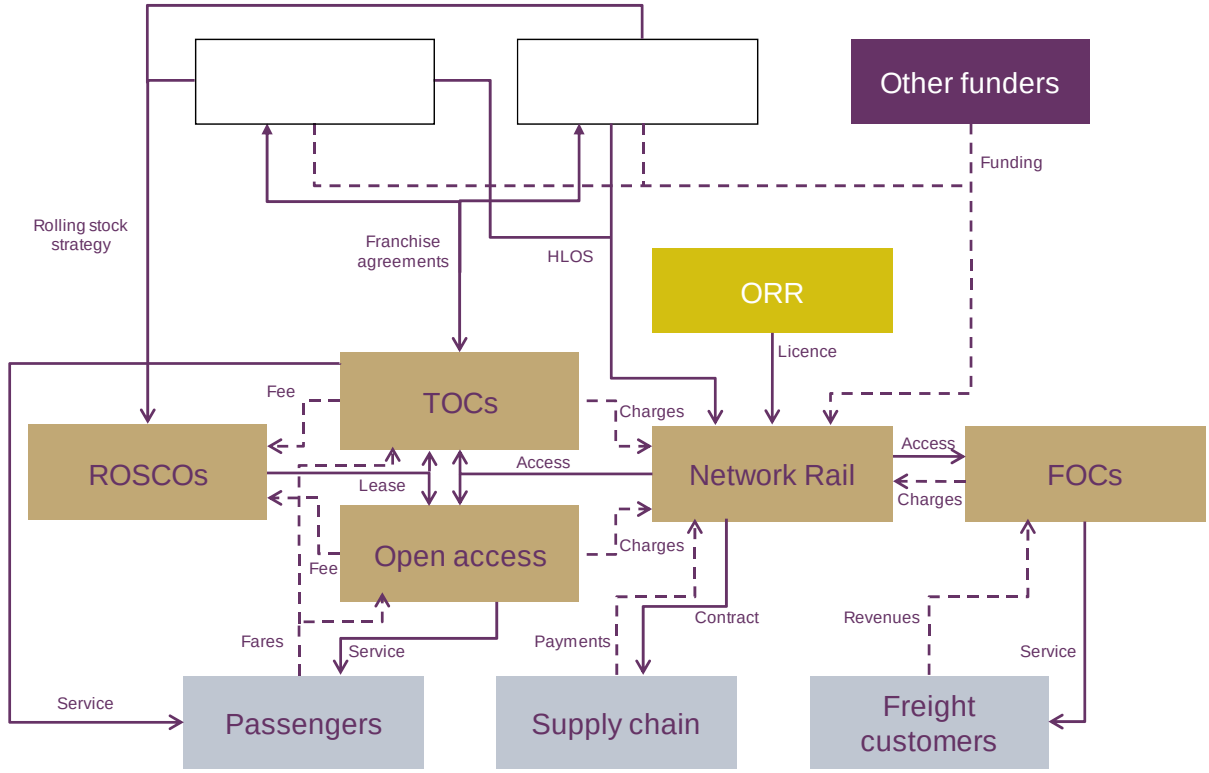
In addition to HLOS-driven works, additional enhancements might be required to deliver other legislative or technical objectives. For example, the Crossrail project is being undertaken with the strategic objective to improve commuter services in London and the South East and to relieve congestion of many Underground and rail lines in the region.[105]

[105] See http://www.crossrail.co.uk/the-railway/why-crossrail.

Other projects, such as the introduction of the European Rail Traffic Management System (ERTMS),[106] might be driven by legal and technical requirements. However, some of the projects initially driven by such requirements might eventually also be reflected in the HLOS; the ERTMS is an example of such a project.

The complexity of the GB rail industry translates into a large number of interfaces that can potentially be affected in enhancement and, particularly, cross-industry projects. Figure A2.17 sets out at a high level the main interfaces that might be involved in large cross-industry projects.

Figure A2.17 Interfaces for enhancement and cross-industry projects



Source: Oxera.

Large projects will commonly engage a number of the players in the figure, which might increase the potential number of interfaces. Moreover, additional parties, such as Network Rail, the ORR and the supply chain, are frequently involved in such projects.

An example of the number and complexity of interfaces is provided in Box A2.3, which describes the interfaces in the GSM-R project.

Box A2.3 Case study: GSM-R—interfaces

The ERTMS consists of two components:

– the European Train Control System (ETCS), a standard for in-cab train control;
– the Global System for Mobile Communications – Railway (GSM-R), an international wireless communications standard for railway communication.

The initial development of the GSM-R project was actually undertaken by British Rail before privatisation and, importantly, under a different industry structure. The decision to introduce GSM-R was partly driven by the need to implement European directives on interoperability (Directives 96/48EC and 2001/16/EC), but also by the recognition that the old systems were aging and hence

[106] The ERTMS is an EU initiative to enhance cross-border interoperability and signalling procurement by creating a single Europe-wide standard for train control and command systems.

would need increasing levels of maintenance and would finally become obsolete.

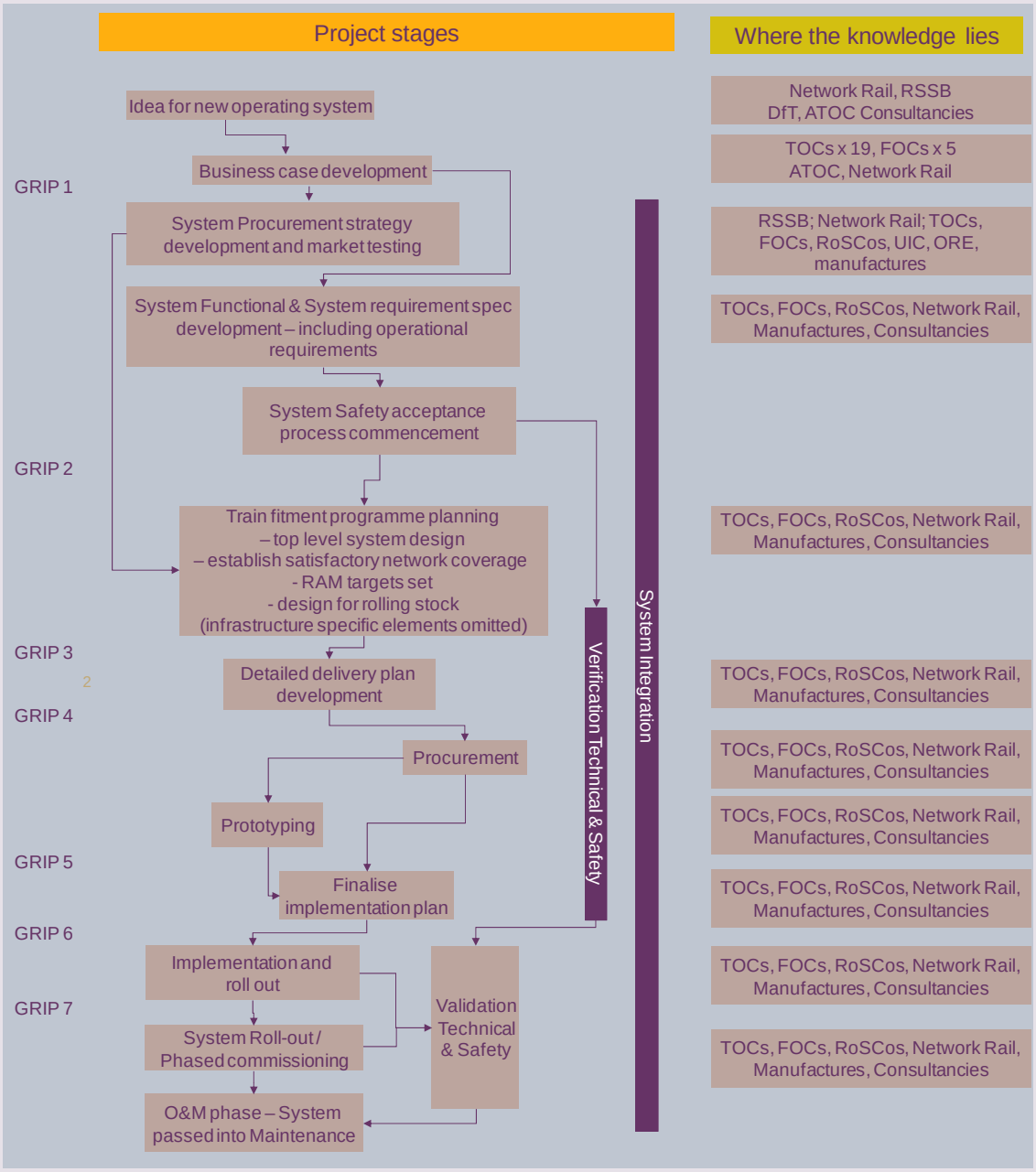
At privatisation, responsibility for the GSM-R project was transferred to Railtrack. However, the industry became considerably more fragmented and complex after privatisation as a result of separation and the increased number of stakeholders.

The project was then developed, first by Railtrack and then Network Rail, under the conditions of the Network Change process of the Network Code (Part G). The project has required consultation and coordination between a large number of stakeholders, including:

– Network Rail;
– the Rail Safety and Standards Board;
– train operators;
– rolling stock companies.

The mapping of the process and high-level stages of the GSM-R project onto the industry interfaces is shown in Figure 1.

Figure 1 GSM-R—project stages and industry interfaces



Source: Halcrow Group Limited (2010), 'GSM-R lessons learnt on industry collaboration: final report', July.

Large cross-industry projects might not only have interfaces with those stakeholders directly involved in the process, they might also affect related processes that are discussed in this study, such as franchise agreements or the rolling stock process. For example, the rolling stock programme is led by the DfT and Transport Scotland, but it has important implications for Network Rail's enhancement programme—eg, it is critical for Network Rail to understand these implications when defining depot layouts, stabling facilities, platform lengths and gauging detail.[107] Moreover, large cross-industry projects might have interdependencies with other projects, as the example of the ERTMS project shown in Box A2.4 below highlights.

Box A2.4 Case study: European Rail Traffic Management System—interdependencies with major projects

The project of introducing ERTMS to the UK to meet European directives is planned to be completed in 2038. In the development of the plan, a number of interdependencies with other major projects and programmes have been identified, including the following.

- The Channel Tunnel Rail Link—where linkages to the French ERTMS implementation plans need to be taken into account.
- Crossrail—with linkages to the Crossrail rolling stock programme.
- Thameslink—with links to the Thameslink rolling stock plans.
- Intercity Express Programme—where it is assumed that new trains will be fitted with ERTMS equipment; however, a dual fitting on parts of the network (Great Western Main Line) is also currently planned, although, we note that this project appears to be currently on hold.
- Electrification—where the RTMS programme affects the appraisal of electrification on specific routes as it has an impact on the cost of electrification.
- European Specification and Product Development—the tendering specifications are linked to specification evolution at the European level.

Source: Department of Transport (2007), 'ERTMS national implementation plan', September.

A2.8.2 Existing incentives

There are two main incentive mechanisms that affect large renewals and enhancement schemes and cross-industry projects:

- Network Code;
- bi-lateral agreements.

Network Code

The contractual structure of cross-industry projects is set by the Network Change process set out in the Network Code. Part G of the Code covers procedures that must be followed by parties wishing to make changes that are likely to have a material effect on the operation of the rail network or on the trains operated on the network (such as major enhancement or cross-industry programmes). The general principle is that before any change can be implemented it must be formally proposed and accepted by those it will affect. Where dispute arises, the Code allows for the resolution of the dispute in favour of the change being implemented.

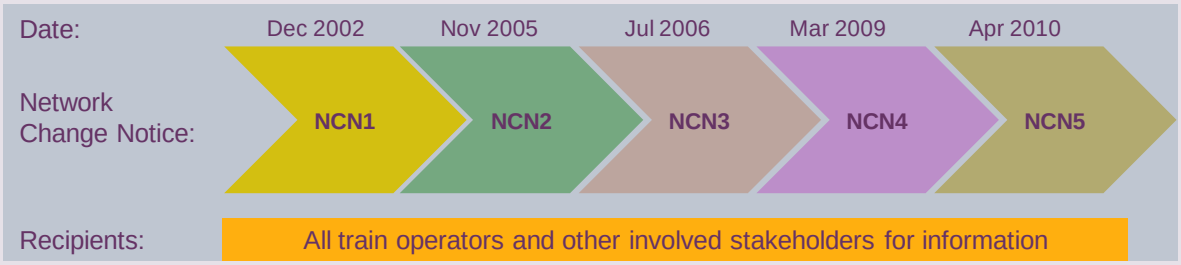
For large projects, which extend over a number of years, it might be necessary for more than one Network Change notice to be issued. For example, the GSM-R project involved publication of a number of notices (see Box A2.5).

[107] The Nichols Group (2008), op. cit.

Box A2.5 Case study: European Rail Traffic Management System—Network Change process

Overall, Network Rail has published five Network Change notices between 2002 and 2010 in relation to the GSM-R programme (see Figure 2 below).

Figure 2 GSM-R—Network Change notice



Source: Oxera, based on Network Rail (2009), 'Network Change notice NCN5', April.

In the latest notice, NCN5, Network Rail recognises that information is still incomplete, highlighting a number of areas where more information will be forthcoming

These include:

– arrangements to support long-term operation and maintenance;
– how the operational railway will respond to a railway emergency call;
– information relating to unresolved scope items;
– conclusions from the trials in Strathclyde;
– details on the final version of the national rollout programme.

Source: Network Rail (2010), 'Network Change Notice NCN5: The national rollout of the GSM-R system in Great Britain', April.

Funding and compensation

As part of the regulatory expenditure allowances, Network Rail is provided with funding of renewals and enhancement programmes. Network Rail is therefore subject to the incentive to outperform the regulator's allowance.

Under Part G of the Network Code, Network Rail provides compensation for the consequences of the implementation of a change, net of benefits. Broadly, the compensation shall be equal to the amount of the costs, direct losses and expenses, including loss of revenue, that are reasonably incurred by train operators. The compensation shall take account of any likely future benefits and the ability to recoup any costs, losses and expenses from third parties, including passengers and customers. The compensation is capped at 75% of all costs incurred by affected parties.

A2.8.3 Bi-lateral agreements

In addition to using Network Change, bi-lateral agreements can be negotiated between stakeholders. This might enable the project to move forward before the Network Change process has been completed, bringing time benefits. For example, in the case of the GSM-R communications project, it has been noted that working in this way has improved the working relationship between Network Rail and operators, with Network Rail willing to be flexible around the needs of the operators and in so doing, improving project delivery.[108]

[108] Halcrow Group Limited (2010), 'GSM-R lessons learnt on industry collaboration (CN00xx): Draft', June.

A2.8.4 Net industry costs and incentive alignment

Transaction costs

Large renewals and enhancement projects, in particular cross-industry projects, require extensive planning, consultation and coordination efforts. The complexity of such projects and of the GB rail industry implies that significant transaction costs might be involved.

This notion is reflected in statements that the study team received during the stakeholder interviews, when it was suggested that large financial savings could be achieved through better delivery of projects across all forms of enhancement.

The interviews suggest that project management costs are perceived as being very high as a result of a lack of precision in project specification at the start.

Further inefficiencies have been suggested, stemming from changing working methods and interpretations of standards during projects. These changes could often create additional costs, where the changes require alteration of the original plans.

Incentive alignment

The cross-industry nature of many large renewals and enhancement projects makes alignment of incentives at interfaces crucial. However, as a result of the complexity of the industry, there might be a risk of separation between:

– client roles defining requirements and providing funding (eg, the DfT);
– agreement of the appropriate level of funding and monitoring (the ORR);
– potentially the eventual ownership of assets (eg, ROSCOs), although in most cases, the enhanced assets will belong to Network Rail.

This potential separation of roles might in turn increase the risk of misalignment of incentives—eg, asset ownership and funding might be separated.

In addition to the risk of incentive misalignment, enhancement and cross-industry programmes are likely to be inherently more complex than equivalent renewals programmes for a number of factors. These include, for example, the requirement during development to demonstrate that they provide value for money. Cross-industry enhancement projects are also typically characterised by being multi-disciplinary, whereas renewals works are generally undertaken as single engineering disciplines.[109]

Complex renewals and enhancement programmes are characterised by many inter-linkages and, therefore, face a higher risk of being delayed. There are a number of sources of inter-linkages.[110]

– **Outputs**—several projects may be required to contribute to a successful overall output; although this is commonly the case in relation to the HLOS, the DfT broadly carries this output risk.

– **Specification**—the specification for one project may be uncertain until another project has reached a certain level of development; this applies in respect of certain fleet procurement projects and the HLOS.

– **Shared resources**—many of the enhancement and renewals programmes will call on the same resources as have already been described above and changes in phasing or demand for one can have knock-on implications for others that are planned at the same time or later.

[109] The Nichols Group (2008), op. cit.
[110] The Nichols Group (2008), op. cit.

– **Geography**—multiple projects may take place in the same vicinity, or on the same line or route, causing competition for access and logistics; although this applies to the enhancement programme, it has already been mitigated to some extent with combinations of smaller projects at busy nodes such as Reading.

Based on experience from recent cross-industry projects such as GSM-R, enablers for the delivery of complex projects have been defined:[111]

– parties interact according to clearly defined roles and responsibilities—there are fewer conflicts;
– parties actively collaborate via ideas and actions to improve project (industry) NPV outcomes;
– the process is underpinned by a rational flow of funds and risk allocation;
– the development process is quicker, more transparent and evaluates the best options; inter-generational equity is preserved by optimising between short-term disruption and long-term benefits;
– delivery is more efficient.

However, it has also been noted that there are limited incentives for industry input and collaboration resulting from:[112]

– the mismatch of benefits and costs over time;
– the indirect way in which specific project CAPEX is charged to operators over time via the 'averaged' track access charges;
– the lack of financial benefits arising in connection with schemes that enhance wider societal outcomes.

A2.9 Allocation of capacity

This section identifies trade-offs between franchise operators, open access operators and freight operators in the allocation of capacity, an important element of which is the opportunity cost associated with one operator, rather than another, being allocated a path.

A2.9.1 Basis of allocation

Track access contracts between Network Rail and the TOCs and FOCs cover:

– access rights to the network held by the TOC or FOC;
– conditions and obligations attached to those rights;
– charges for the exercise of those rights;
– compensation for not being able to exercise those rights on a temporary basis;
– the liability of the parties to each other in the event of breach of contract.

Section H of Network Rail's Network Code (dated November 26th 2009) sets out:

– the operation of train services on the network in accordance with the working timetable;
– the restoration of operations in accordance with the working timetable;
– the efficient and economical operation of the network and of trains operating on it;
– adherence to criteria published by the ORR.

This element of the Network Code is kept under regular review and covers such issues as: notification of disruptive events; contingency plans; clearance of track blockages and assistance to failed trains; emergency timetabling procedures; control arrangements; train regulation; seasonal-preparedness; and other matters necessary or expedient to achieving its objective.

[111] Jones, G. (2010), 'Improving development and delivery of cross-industry projects', February 12th.
[112] Ibid.

Box A2.6 provides an example of how capacity allocation can be particularly complex.

Box A2.6 Railway operations/capacity allocation: example—East Coast Main Line

This case study describes capacity allocation and capacity management from the perspective of capacity allocation and timetable development on the East Coast Main Line (ECML). Note that as the timeline below shows, this is a two-year process.

The key timelines in this process are summarised below.

February 29th 2008: the ORR issues a letter to identify all operators' aspirations for additional access rights on the ECML.

– February 29th 2008: Network Rail (Network Rail) publishes ECML RUS.
– March 4th–July 28th 2008: industry responses (18 different companies/stakeholders, including the DfT).
– June 25th/27th 2008: DfT discussions with Network Rail regarding timetable development (the DfT stated that the ORR's timetable specification should assume a complete rewrite of the ECML timetable, which should not be constrained by the exact timings of then current services).
– July 16th 2008: the ORR issues a response to stakeholders (all passenger and freight train operators, all holders of and/or potential applicants for track access rights or track access options, Network Rail, Transport Scotland, the DfT, South Yorkshire, West Yorkshire and Tyne & Wear PTEs, Passenger Focus, London TravelWatch).
– In the meantime, Network Rail continues to work on the various bids for the December 2008 timetable in the usual way through the timetabling development process.
– August 8th 2008: further update to stakeholders from the ORR.
– August 19th 2008: the ORR commissions Network Rail to produce a capacity and performance report by September 19th 2008, analysing the various applications against the available capacity, and against the background of the anticipated CP4 settlement and the established freight and ECML RUSs.
– August 8th–December 23rd 2008: further stakeholder responses.
– September 26th 2008: Network Rail publishes ECML Capacity Assessment Report.
– October 6th 2008: DfT response to Network Rail report.
– September 23rd–October 14th 2008: stakeholder responses to Network Rail report (16 stakeholders).
– December 18th 2008: Network Rail issues second Capacity Assessment Report
– December 23rd 2008–January 26th 2009: stakeholder responses to second Network Rail report.
– January 29th 2009: the ORR issues proposed decisions on key aspects of applications for track access rights.
– February 2nd–25th 2009: stakeholder responses to ORR proposed decision(s).
– February 27th 2009: the ORR publishes final decision on applications for track access rights for passenger services.
– July 3rd 2009: Network Rail develops standard pattern timetable options.
– August 10th 2009: Workshop 1, covering London to Doncaster; Workshop 2, covering Doncaster to the Border; Workshop 3, covering Scotland.
– October 16th 2009: Network Rail's SLC2 timetable.
– October 21st–November 5th 2009: stakeholder responses to SLC2.
– December 18th 2009: Network Rail SLC2 timetable issued.
– February 11th 2010: ORR decision letter.

It is noteworthy that the ORR has been reviewing its approach to capacity allocation in recent months.[113] This review has covered 'moderation of competition' rules that were brought in at the time of privatisation to protect franchised operators from competition from open access operators, and which have been progressively removed; trading off different types of services (eg, passenger versus freight, operational versus commercial); and trade-offs between performance and capacity.

An alternative to current approaches would involve auctioning capacity on the network, as seen in other sectors such as broadcasting and gas.

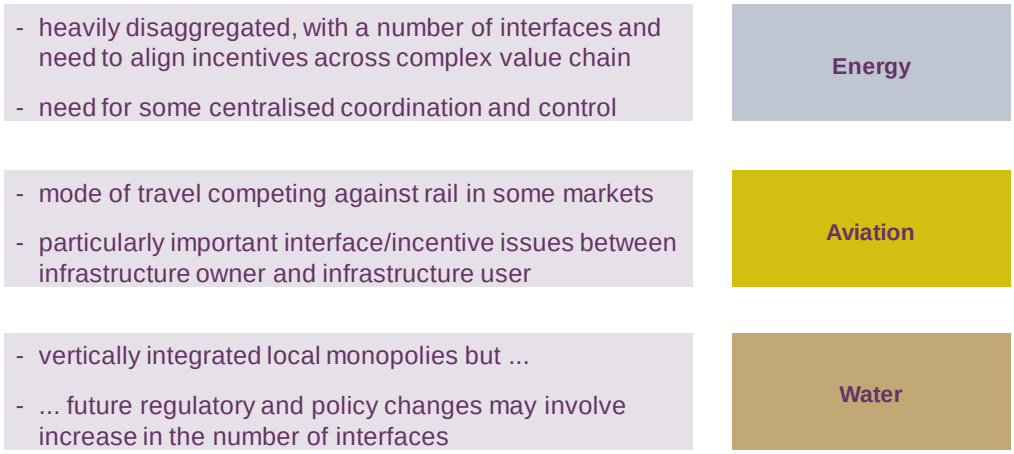
[113] Current progress with the Review of Access Policy may be viewed at: http://www.rail-reg.gov.uk/server/show/category.2254

A3 Lessons from other sectors

This section looks at experience in other sectors to identify potential lessons for the review of interfaces, incentives and structures in rail. In line with the agreed scope of the project, experience in three GB sectors has been examined: energy, aviation and water.

Each of these sectors offers a history of different models of separation and integration, ranging from a highly disaggregated structure in energy to a vertically integrated industry in water that is, nevertheless, the subject of active debate that is likely to lead to an increase in the number of interfaces, as shown in Figure A3.1.

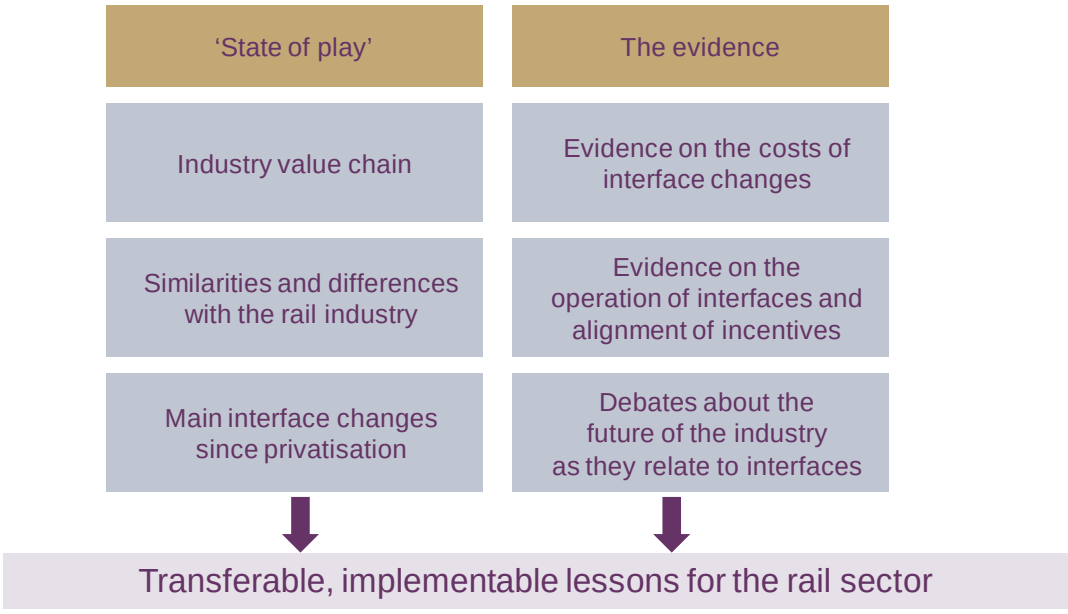
Figure A3.1 Sectors considered



Source: Oxera/Arup.

The approach to analysing these other sectors is summarised in Figure A3.2.

Figure A3.2 Approach to analysing other sectors



Source: Oxera/Arup.

The following describes these issues for each sector, before considering lessons for the rail sector.

A3.1 Energy

The energy sector is one of the most disaggregated of all GB infrastructure networks. Not only have regulators enforced the same basic split between monopoly and competitive businesses as in other sectors, but they have also drawn further distinctions and boundaries *within* these broad categories—for example, between system operations and asset ownership in network businesses, or between generation and supply in energy delivery. Regulators and industry participants have had to design sophisticated mechanisms for dealing with coordination issues and ensuring that the sector as a whole meets the needs of energy consumers and the policy goals of government.

The current debates in relation to the structure of the industry are also informative.

– On the one hand, Ofgem is contemplating mechanisms for strengthening the contestability in CAPEX delivery for regulated networks. These mechanisms might create a new, complex interface between the network operators, the asset owner and the third parties responsible for delivering the additional pieces of infrastructure.

– On the other hand, there is a growing recognition that the current configuration of the industry is already creating coordination issues, and that the innovation needed to create 'smart grids' and promote sustainability throughout the sector might potentially require a degree of reintegration in some areas. For example, as an increasing share of generation capacity is connected at the distribution level (eg, onshore wind), distribution network operators must play a greater role in system balancing; moreover, their interface with the transmission network has become so complex as to prompt a debate about further integration between distribution and transmission.

The GB energy network sector therefore presents a case study of particular relevance to the issues being discussed in this study.

A3.1.1 The value chain in electricity and gas networks

The GB gas and electricity value chains operate at a relatively high level of disaggregation.

Electricity

Here, the value chain is based around four key activities that are required to provide electricity to end-consumers.

– **Generation**—the production of electricity through generating assets such as nuclear power stations, gas- and coal-fired power stations, and wind turbines.

– **Transmission**—the long-distance transportation of high-voltage electricity produced by generating assets via the transmission network.

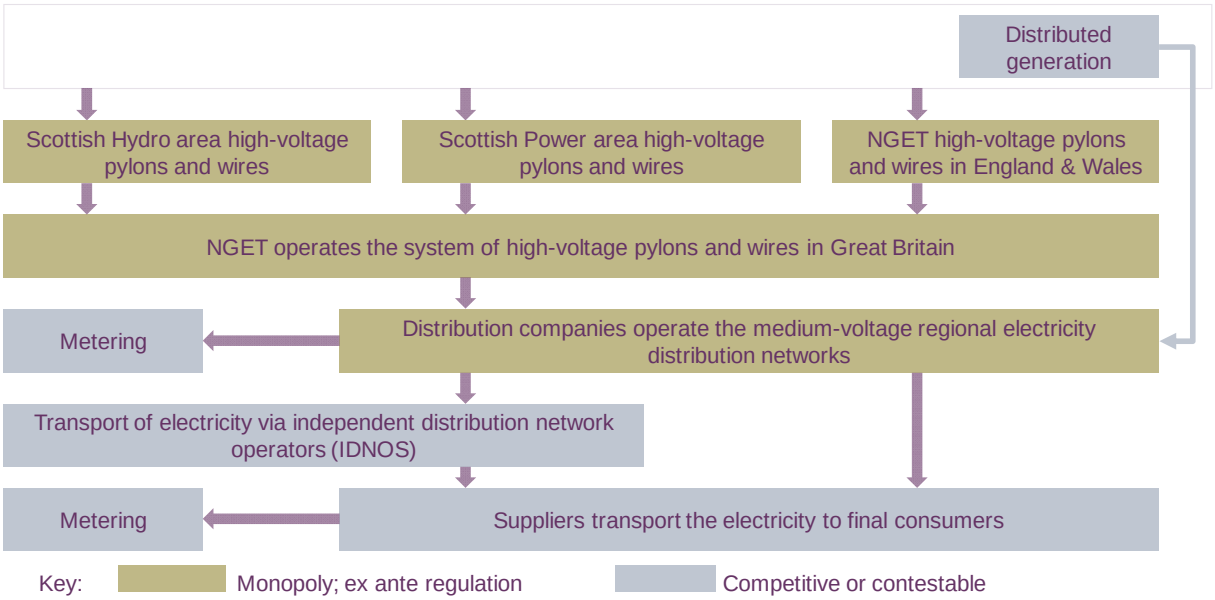
– **Distribution**—the transportation of low-voltage electricity (which has been transformed from high-voltage), typically from the transmission network to end-consumers via the distribution network. Distribution networks generally cover a localised area.

– **Supply**—the provision of electricity to consumers. This includes obtaining the electricity to sell on, and the billing and marketing of electricity (metering).

It is also possible to identify a separate fifth part of the chain—system management—which is concerned with ensuring that electricity supply balances demand in the system as a whole.

Figure A3.3 illustrates this value chain, and the main interfaces in the GB electricity sector, although it is not exhaustive with respect to the full set of interfaces.

Figure A3.3 Value chain and key interfaces in the GB electricity sector



Source: Based on Ofgem (2009), 'History of Energy Network Regulation', February, Figure 2.2, p. 15.

The costs passed through this value chain are a mix of monopoly regulated costs (largely costs associated with the distribution and transmission networks) and costs determined in competitive markets (such as in the upstream generation market and downstream retail markets). For companies operating in competitive parts of the value chain, normal commercial pressures exist to minimise costs. For network companies, incentives to minimise costs emerge from the RAB/WACC system of regulation and approach to price capping that is overseen by Ofgem. The role of the system operator in coordinating activities and cost minimisation across the industry as a whole is discussed in detail below.

Gas

The physical gas chain is characterised by successive vertical stages of production. These stages have differing economic characteristics driven (primarily) by their cost structures and their potential for contestability.

- **Exploration, production and importation**—this activity first consists of prospecting oil and gas fields. Once found, the oil and gas are extracted and transported to the receiving terminal (located at the beachhead, where extraction occurs offshore) through pipelines, at which point the gas enters the onshore system; alternatively, it may be imported from other jurisdictions.

- **Trading**—trading involves the exchange of large volumes of gas, and is conducted by two basic types of market participant:

  - suppliers buy and sell gas in the wholesale market in order to fulfil the supply requirements of their end-consumers;

  - traders are agents that do not necessarily serve end-consumers, but rather buy and sell gas in the wholesale market in order to exploit opportunities for arbitrage arising from price differentials between different markets.

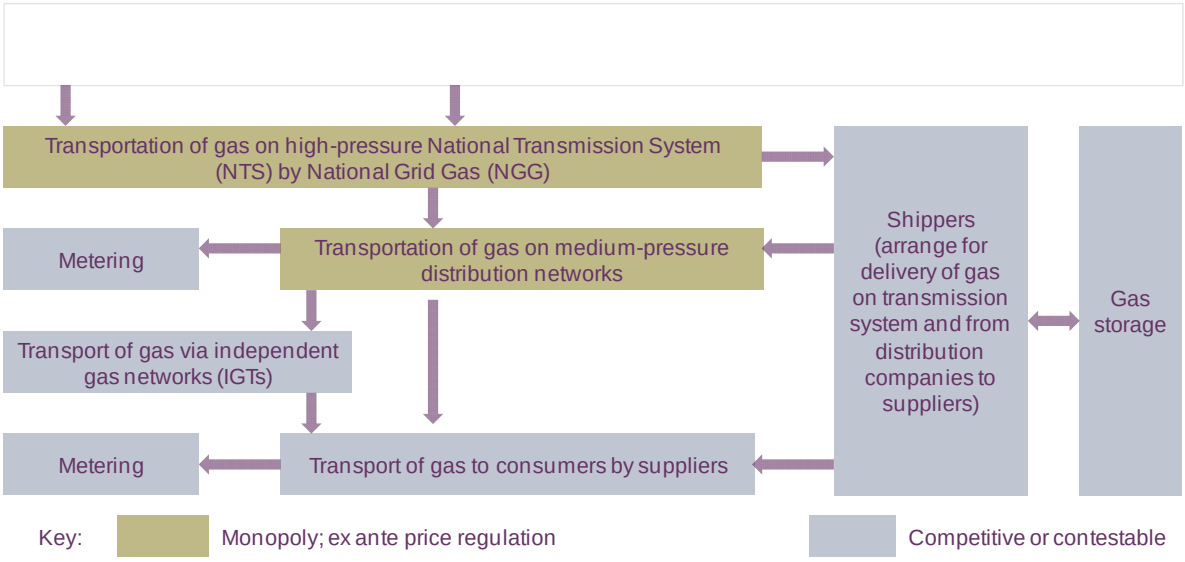
- **Transmission (transportation) and distribution**—these activities consist mainly of moving gas from point A to point B, through pipelines. The difference between gas transmission and distribution is that, in the former, gas is usually transported over longer distances and at high pressure, while in the latter it is distributed at low pressure on a regional or local basis.

– **Storage**—gas suppliers use storage capacity to meet variable patterns of gas demand and thus to keep the system in balance.

– **Retail supply**—suppliers (sometimes called 'shippers' or 'traders') purchase gas upstream and sell it to the end-consumer. This part of the value chain includes activities such as metering and customer service.

A schematic representation of the gas value chain is provided in Figure A3.4.

Figure A3.4 Value chain and key interfaces in the GB gas sector



Source: Based on Ofgem (2009), 'History of Energy Network Regulation', February, Figure 2.1, p. 11.

A3.1.2 Post-privatisation landscape and subsequent changes

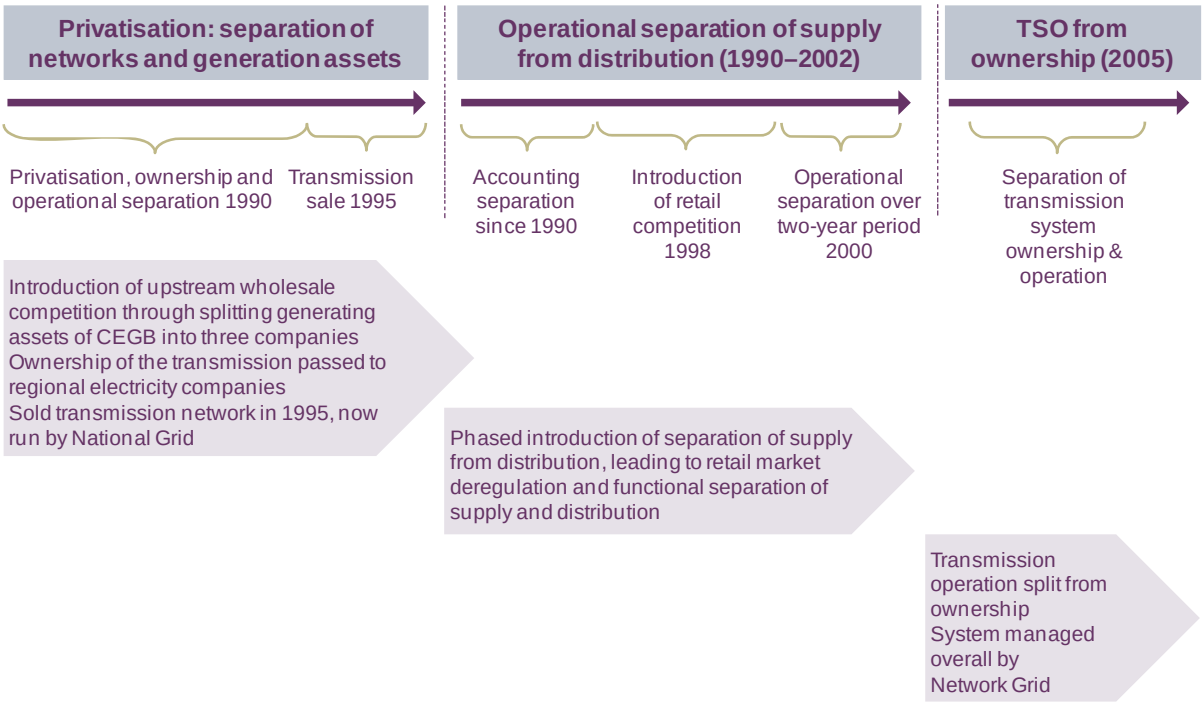
Electricity

Before privatisation, the Central Electricity Generating Board, a state-owned monopoly, operated transmission network and generating assets in England Wales. It supplied 12 state-owned regional electricity companies, which acted as monopoly regional distribution networks and monopoly regional electricity suppliers (RECs).

In Scotland there were two fully vertically integrated electricity providers. As the Scottish transmission system (and, post-privatisation, wholesale market transactions) were operated on a separate basis from England and Wales until 2005, the process of privatisation was slightly different.

The privatisation, restructuring and deregulation of the British electricity industry took place in stages. The main forms of vertical separation implemented since this process began are summarised in Figure A3.5, and discussed below in more detail.

Figure A3.5 Timeline of separation in GB electricity industry



Source: Oxera.

The high-level developments with respect to separation and interfaces were therefore as follows.

– **Operational and ownership separation of the transmission network from generation assets**. This occurred with the privatisation of the industry in 1990 and the introduction of upstream wholesale competition when the generating assets of the CEGB were divided between three companies: PowerGen, National Power, and Nuclear Electric. Ownership of the transmission network was then passed to the RECs, which sold it in 1995. The transmission network is now run by National Grid.

– **Operational separation of electricity supply from distribution**. Following privatisation in 1990, the RECs became public electricity suppliers (PESs), with a licence covering both operation of the distribution network and supply (accounting separation was enforced and these elements were separately regulated). Following the introduction of retail competition into the downstream domestic retail electricity market in 1998, operational separation measures were instituted over a two-year period from 2000.

– **Transmission system operation from ownership**. This occurred when the Scottish electricity market was integrated with that of England and Wales in 2005. Although the Scottish incumbents maintain ownership of the transmission network, the overall system is managed by National Grid.

Regulatory interventions have reduced the vertical integration of the industry, but mergers and acquisitions in the sector have increased vertical integration between electricity suppliers and generators, with many companies now owning both generation and electricity supply companies. There are currently six large electricity suppliers in the UK, all of which own generating capacity.[114]

[114] Ofgem (2008), 'Energy Supply Probe: Initial Findings Report', p. 28, para 2.40.

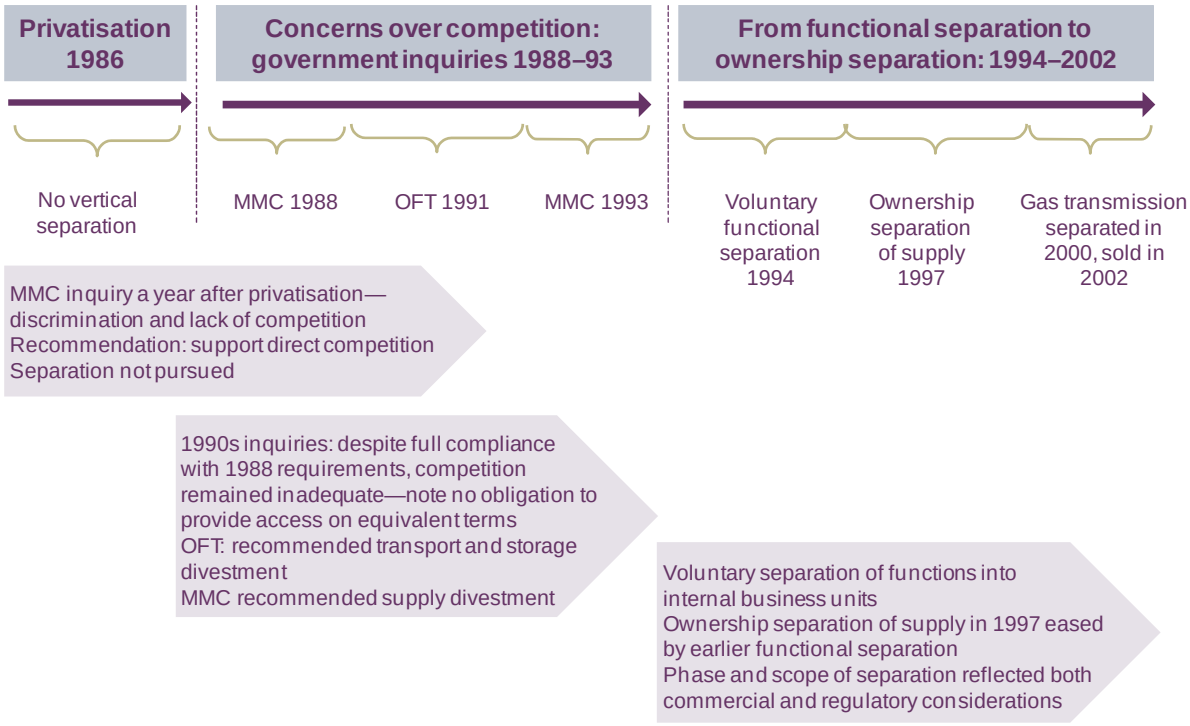
Gas

The Gas Act 1986 established the structure and regulation of the post-privatisation gas industry and paved the way for the privatisation of British Gas.[115] The company was privatised as a vertically integrated entity: there was no vertical separation of transportation and storage from supply, nor was there horizontal separation into separate regional companies.

Its management headquarters was responsible for all centralised activities (gas exploration, purchasing and research & development) and management across 12 regions, each of which was responsible for all customer-related activity (supply, retail sales and service).

As with electricity, the privatisation, restructuring and deregulation of the industry took place in a number of stages, summarised in the timeline given in Figure A3.6 below, which focuses on vertical separation (and not subsequent horizontal separations).

Figure A3.6 Timeline of gas separation



Source: Oxera.

The high-level developments with respect to separation and interfaces were as follows.

– **Internal separation of transportation and storage from supply**. This occurred under the impetus of regulatory concern over competition in the non-tariff market, which led to internal separation under the Gas Act 1995.

– **Formal separation of British Gas**. British Gas demerged its supply business from Transco, and BG Storage was established as a separate stand-alone business. In 2002, National Grid Group merged with the owner of Transco to create a single gas and electricity transmission company.

– **Horizonal separation and sale of regional distribution networks**. National Grid Gas sold four of the gas distribution networks to new owners in 2005, and retained the remaining four networks.

[115] Armstrong, M., Cowan, S. and Vickers, J. (1994), *Regulatory Reform: Economic Analysis and British Experience*, Massachusetts Institute of Technology, p. 254.

A3.1.3 Similarities with and differences to the GB rail value chain and industry structure

The following considers some of the similarities and differences in industry structure between rail and electricity and gas. These issues are reflected in the discussion of transferable lessons that forms part of the conclusions to this review of other sectors (see section A3.4).

Electricity

- **Homogeneous product**. Consumers tend to be indifferent between who physically provides electricity, although there are differences in quality of service that may make the choice of supplier significant.[116]
    - Rail users require journeys differentiated on a number of dimensions, including origin/destination, route/number of changes, time of travel, quality of service, etc.
- **Continuous market clearing**. Electricity is difficult to store, and as demand is continuous and the extent to which the electricity produced by generating assets can be controlled in the short term varies (eg, generation by wind turbines is uncontrollable and that of nuclear power stations cannot be easily altered), the market must be continuously managed by a systems operator to ensure that supply is available to meet all demand on a real-time basis.
    - Rail demand and supply adjust over years, rather than continuously. However, issues of network planning over the long term are also relevant in the energy sector, as set out in section A3.1.5 below.
- **Price volatility**. Both electricity supply and demand may be subject to significant changes which, combined with the need for continuous market clearing, can result in significant price volatility in the wholesale market, which may in turn affect the retail market. Price volatility can arise on both the demand side (eg, seasonal changes in demand) and on the supply side (eg, changes in the international price of inputs used to generate electricity).
    - Rail prices are subject to less volatility, although this is an administrative choice, which does not necessarily reflect volatility in the use of capacity throughout the day or year.
- **Environmental issues**. The generation of electricity raises significant environmental concerns.
    - Environmental issues are relevant to the rail industry, especially in evaluating the costs associated with competing modes of infrastructure, but in general rail is a low-carbon form of transportation.

There are also some high-level similarities between the industries.

- **System-wide coordination and real-time demand management**. There is a need for day-to-day and minute-by-minute balancing on the system by a responsible centralised authority.
- **Long-term capital investments**. Capital investments in generation and network assets are typically undertaken over the long term; many generation and network assets in the GB energy sector have been in operation for a number of decades, as have certain assets in GB rail. There is an ongoing need to maintain, enhance and renew these assets.

[116] It is possible that increased environmental concern may make consumers more sensitive to the means by which their electricity is generated in future.

Gas

– Like electricity, **Gas is a relatively homogeneous product**.

– **Gas prices are determined in internationally traded markets** and are linked to those of other commodities, such as oil.

There are also some high-level similarities between the industries.

– **System-wide coordination**. As in electricity, although to a lesser extent, there are some similarities between the need to coordinate the use of the gas network by different parties, which must be managed by a centralised authority.

– **Nature of the physical network** The physical infrastructure of the gas network is less complex than other networks, such as electricity grids and telecoms networks. This relative simplicity is a characteristic shared with the physical assets making up the railway network, although this is not to say that the management of the network is not, in itself, complex.

A3.1.4 Evidence on the cost of changes

This section reviews the available evidence on the costs of restructuring in the GB electricity and gas industries as they relate to interfaces. Studies on the costs associated with separation, and the introduction of interfaces, can best be understood by comparing them to the desired policy objectives, and the motivation for breaking up integrated monopolies, in the face of the relevant counterfactual. In both cases, the drive to separation, and the concomitant increase in interfaces, was founded on a belief that this would lead to greater competition.

Identifying and quantifying the impact of the movement away from the relevant counterfactual is a complicated exercise, given the staged and progressive liberalisation programme that featured in both industries. However, some attempts have been made to quantify the effects of change. For example, in its 2001 assessment, the National Audit Office (NAO) concluded that, due to greater competition in electricity, there were annual savings of £143m.[117] These arose from consumers switching to cheaper suppliers and being able to switch to cheaper payment mechanisms, although there has been some criticism of the methodology used to generate these estimates.

Having estimated the costs of putting in place arrangements for competition (rather than separation and the creation of interfaces per se), Ofgem allowed as a pass-through item to consumers in the price controls for distribution networks, and for incumbent electricity suppliers' tariffs in the transition to competition. Ofgem allowed companies to recover some £850m from customers over the seven years 1998/99 to 2004/05—an average of £121m each year—to meet the extra costs that the companies incurred. This was equivalent to around £4, or 2%, on the average annual electricity bill.[118] However, most of this cost related to the establishment of the arrangements for competition rather than separation per se. To put this in context, Ofgem allowed each PES £1m per year on the distribution business's ongoing operating cost from 1998/99 to 2004/05, and £0.2m per year on the supply business, as a result of the costs of separation.[119]

Newbery and Pollitt (1997)[120] examined whether the restructuring of the GB electricity sector was worthwhile by undertaking a social CBA, which involved constructing various counterfactuals around how the industry might have developed if a vertically integrated

[117] National Audit Office (2001), 'Ofgem, Giving Domestic Customers a Choice of Electricity Supplier'.
[118] Ibid.
[119] Ofgem (1999), 'Reviews of Public Electricity Suppliers 1998–2000: Distribution Price Control Review—Final Proposals', p. 21 and 'Supply Price Control Review: Final proposals', p. 30.
[120] Newbery, D. and Pollitt, M (1997), 'The restructuring and privatisation of the CEGB – Was it worth it?', *Journal of Industrial Economics*, **45**, pp. 269–303.

industry structure had been retained. The authors estimate the effect of the restructuring as a permanent reduction in costs of 5% a year. This consists of benefits arising from improved productivity and efficiency by generators, switching from nuclear power, and efficiency gains, while the main costs included premature investment and the costs of the restructuring itself.

In gas, the *direct* costs of the functional separation of transportation and storage had been estimated by British Gas during the 1993 MMC inquiry at £50m per annum, or £330m at discounted present cost over ten years.[121] This was the least costly of the separation measures considered at the time. However, the restructuring ultimately undertaken in 1994 created six separate business units, of which one was the transportation and storage business. On that basis, the actual direct costs of the functional separation are likely to have been higher.

Table A3.1 summarises the direct cost estimates produced by British Gas during the course of the MMC inquiry.

Table A3.1 Direct costs of vertical separation measures considered

	Annual cost (£m)	Discounted cost over ten years (£m)
Structural separation		
Divestment of trading business	130	870
Divestment of trading unit into several regional companies	n/a	2,300
Functional separation		
Business unit dedicated to transportation and storage	50	330
Hybrid arrangements		
Joint ownership of transmission network by integrated regional distribution and trading companies	n/a	2,900

Source: MMC (1993), 'Volume 1 of reports under the Fair Trading Act 1973 on the supply within Great Britain of gas through pipes to tariff and non-tariff customers, and the supply within Great Britain of the conveyance or storage of gas by public gas suppliers', August.

The subsequent separation of the gas distribution networks has been noted above. Ofgem's 2004 impact assessment examined the consequences for price regulation of having more comparators, and found that estimated benefits increased from £145m for one comparator to £355m for four comparators.[122]

The NAO's subsequent review of Ofgem's analysis, in 2006, found that the regulator's estimate of the potential benefit of £325m (95% of which was due to the existence of more comparators) was somewhat conservative, but also that this was a negligible benefit compared with the average end-consumer bill.[123] The creation of associated interfaces was associated with an estimated annual cost of £7m, and a one-off cost of £25m. It is notable that a mechanism was explicitly created to minimise the overall number of interfaces confronting shippers.

A subsequent report for Ofgem identified scope for efficiency gains among gas DNs of 1.8–3.7% a year, although this does not account for a counterfactual of continued ownership.[124]

[121] MMC (1993), 'Volume 1 of reports under the Fair Trading Act 1973 on the supply within Great Britain of gas through pipes to tariff and non-tariff customers, and the supply within Great Britain of the conveyance or storage of gas by public gas suppliers', August, p. 42.
[122] Ofgem (2004), 'National Grid Transco – Potential sale of gas distribution network businesses Final Impact Assessment', November.
[123] National Audit Office (2006), 'Sale of gas networks by National Grid', February.
[124] Europe Economics (2007), Top Down Benchmarking of UK Gas Distribution Network Operators', A European Economics report to Ofgem, April 10th.

A3.1.5 Evidence on the operation of interfaces, alignment of incentives, and future debates about interfaces

This section looks at what the most problematic interfaces are in the electricity and gas industries between network entities and the rest of the industry, and summarises current debates about how problems might be resolved, or at least mitigated.

After setting the scene, the questions addressed in the following sections are as follows.

– How is investment coordinated between network companies and the other entities in the industry?

– How is long-term network planning encouraged? This issue is being covered in more detail in the contemporaneous study by SDG/CEPA for the VfM Review team.

– What is the link between long-term planning and day-to-day delivery?

– How are the networks encouraged to reduce costs, while continuing to deliver the required outputs?

– What is the role of the National Electricity Transmission System Operator (NETSO), and how is the NETSO incentivised to manage the electricity system efficiently?

Background

The two principal current drivers of investment in the electricity industry are to:

– ensure security of supply against the background of substantial closure of generating plant in the medium term—partly owing to the ageing of plant, including nuclear plant, and owing to legislation on emissions from generating plant;

– achieve substantial decarbonisation of the sector, particularly by building and connecting a substantial amount of renewable generating capacity (mainly onshore and offshore wind generation) and by at least replacing the existing nuclear power stations.

This background will require substantial investment in networks in order to transport the electricity from new generating plants, in particular, because:

– most of the new renewable generating capacity will be located far away from where the demand is, and from the existing network infrastructure. (Much of the onshore wind generation is, and will be, in Scotland, and there are already significant constraints on transmitting electricity from Scotland to England);

– new nuclear plants, although likely to be built on the sites of existing nuclear plant, will have significantly more capacity per power station than the existing plant, requiring a significant uprating of the existing power lines.

Although, in the longer term, much of the new network investment may be on (low-voltage) distribution networks (the creation of ‘smart’ distribution grids), the bulk of the short- to medium-term requirement will be on the GB electricity transmission network—to transport bulk power from remote locations to where the demand is. For this reason, the focus in the following sections is primarily on transmission.

The network companies involved in electricity are:

– the three transmission networks operated by National Grid, Scottish Power, and Scottish & Southern Energy in their role as transmission operators (TOs);

– the 14 licensed DNOs;

– the NETSO, which is National Grid (hereafter referred to the ‘SO role’).

How is investment coordinated between transmission network operators (TOs) and the other entities in the industry?

Before privatisation in 1990, the CEGB operated both generation and transmission in England and Wales (and the same combined functions were carried out by the two state-owned Scottish electricity companies in Scotland). When considering new generation—either to meet the demand for extra generating capacity in aggregate, or to substitute cheaper new generation for more expensive existing generation—the CEGB’s investment appraisal internalised the combined costs of generation and transmission, and, in principle (and subject to non-financial objectives to which the industry was exposed), preferred schemes with the lowest combined generation and transmission costs.

In the current disaggregated industry, a number of mechanisms exist for achieving the same objective.

– National Grid (in its SO role) sets a structure of charges for use of the transmission system (transmission network use of system, or TNUoS, charges), which has a significant locational element. This confronts putative new generating plant with an approximation of the (long-run) incremental costs of reinforcing the transmission system to accommodate the new generation, and encourages the generator to internalise the combined marginal costs of generation and transmission in making its investment. Ofgem has oversight of the structure of TNUoS charges and National Grid operates under an overarching obligation (both statutory and in its licence) to develop an efficient transmission system. Until now, the requirement for ‘efficient’ transmission network charges has broadly been interpreted as requiring some sort of cost-reflectiveness, albeit various parties (especially the Scottish National Party and generators in Scotland) have lobbied vigorously for postalised prices. Recently, these parties have tended to use the government’s objectives for decarbonisation of the electricity sector as support for their position, given that much of the likely renewable generation in GB tends to be located some distance from where the demand is and consequently has high associated transmission costs.

– Would-be generators are required to make a firm financial commitment to the SO in order to obtain a ‘connection agreement’ with the SO, a necessary pre-condition for connection to the transmission system. Although the form of this financial commitment has evolved over the last few years, the intention underlying this commitment is to discourage transmission investment that will subsequently not be needed, because the generation project in question has not been completed. However, once the generator is connected to the transmission system, it has to give only very short notice that it is going to disconnect—with the potential for this to trigger unnecessary transmission investment if new generators are looking to connect in the relevant zone. There are current plans to increase this notice period but, in effect, only from a few days to a year or so.

– At their periodic price reviews, the TOs submit plans for enhancement of the GB transmission network which reflect financial commitments that the generators have already made. However, these plans may also include forecasts of generation that is likely to be seeking connection to the network at some point but has not yet reached the stage of financial commitment. This will tend to be more of an issue with generating plant that has relatively short construction times—for example, wind farms and gas-fired plant, rather than, for example, nuclear plant.

An underlying problem in coordinating network investment with investment in generation is that there has been a tension between the approach to the issue favoured by Ofgem in the past and the government’s desire to hit targets for decarbonisation.

– Ofgem’s philosophy has been (and, arguably, still is) to favour an approach in which networks are, to a large extent, planned by network *users*, rather than by the network

companies themselves. Thus, Ofgem has opted for an approach in which major network enhancements go ahead only on the back of firm financial commitments from network users to use these enhancements. This approach reached its apotheosis in the current mechanism for driving investment to expand 'entry' capacity in the gas transmission system (ie, capacity to inject gas into the network at a particular location). Under this approach, biannual auctions are held for long-term rights to inject gas into the transmission system at designated locations. If the rights sold exceed the available capacity at a given location then, in the normal course of events, National Grid will build more capacity and is financially incentivised on both the cost and the timing of the new capacity. Furthermore, in the normal course of events, National Grid will not build new entry capacity in the gas transmission system without users having bought the capacity (and for an appropriate duration) in the auctions. Ofgem's approach is designed to deliver a 'market' solution to network expansion and to lessen the risk of new network assets becoming prematurely stranded.

– On the other hand, one of the problems with Ofgem's approach—as exemplified by the above mechanism for gas entry rights—is that it may delay the building of new network capacity at a time when the government is increasingly committed to volume/environmental targets, rather than to facilitating market outcomes. The Ofgem preferred approach tends to cause delay because:

  – the putative network user will not normally be prepared to make a firm commitment to pay for future access rights until it has secured all its consents and financing;

  – if National Grid starts securing consents and undertaking other pre-construction work only once it has received a firm commitment from users, processes which could have run in parallel become sequential;

  – planning consents have traditionally taken much longer to secure in electricity than in gas (the main reason why the last major completed enhancement of the electricity transmission system—the 'North Yorkshire line'—took around ten years to deliver).

Put another way, one of the key tensions here is between Ofgem's fear that National Grid (and other network companies) will tend to overbuild new capacity in order to grow their regulatory asset values (RAVs)—reflecting Ofgem's dominant preoccupation with the energy *prices* paid by consumers—and DECC's desire to hit *volume* targets for renewable energy.

Faced with this tension between its own desire to have the energy market drive network enhancement and government pressure to expand the electricity transmission network's capacity to handle new sources of renewable generation as soon as possible,[125] Ofgem has looked to introduce 'enhanced incentives', under which National Grid would (as DECC desires) aim to anticipate what new transmission network capacity will be required, but would also be financially exposed to the utilisation of that capacity—thereby putting financial pressure on National Grid not to overbuild new capacity. However, at this stage, Ofgem has made little progress with such new incentives; rather, it has agreed to enough extra revenue for electricity transmission companies to undertake pre-construction work on particular projects, in order to keep open the option of relatively speedy delivery in the event that particular generation projects do eventually make firm commitments to pay for new network capacity.[126]

[125] The current guidance to the Gas and Electricity Markets Authority exhorts the Authority (ie, Ofgem) to achieve 'an early start by network companies in identifying and planning necessary works, in dialogue with developers, to ensure that those plans are better placed in relation to new generation, including renewables, nuclear and other low carbon developments. The Government expects this to mean that more preparatory work will need to take place before firm commitments are given by generators'. DECC (2010), 'Social and Environmental Guidance to the Gas and Electricity Markets Authority', January 18th, para 11.
[126] Ofgem (2010), 'Transmission Access Review – Enhanced transmission incentives: Final Proposals, January 19th.

How is long-term network planning encouraged?

As noted above, Ofgem's core approach to encouraging long-term network planning by the transmission companies has been to:

– require financial commitment from network users, particularly generators, for network enhancements;
– confront those generators with broadly cost-reflective locational charges in order to encourage the generators themselves to minimise the combined costs of generation and transmission.

Beyond this, in the core approach, the role of National Grid has been to provide information about current and future utilisation of the transmission grid to help inform generators' investment decisions. The main channel for disseminating this information has been the annual 'Seven Year Statement' (with a similar 'Ten Year Statement' for gas transmission), although, in keeping with National Grid's role as market facilitator—rather than as 'planner' of the electricity system—the assumptions in these statements about future network use reflect information provided to National Grid by users, rather than National Grid's own judgements. Given that many potential generation projects will not, in the event, go ahead, the Seven Year Statement is not necessarily a realistic guide to future generating capacity in the transmission system.

However, as noted in the previous section, this core approach does not sit comfortably with the government's security of energy supply objectives and, even more, with its environmental/climate charge targets for the energy sector. Thus, Ofgem has been reappraising, primarily as part of its 'RPI – X@20' review, how network regulation should work and how it should encourage networks to plan for the longer term.

In the light of RPI – X@20, Ofgem may modify how it encourages longer-term thinking by energy networks. (Ofgem typically steers clear of the notion of 'planning' of energy networks.) The ideas floated include:

– longer-term business plans at the periodic reviews;
– longer-term price controls. (Ofgem's current preference is for eight-year control periods, with a 'mini review' midway, after four years);
– providing an 'innovation stimulus' fund—building on the Low Carbon Networks Fund which was an outcome of the most recent electricity Distribution Price Control Review ('DPCR5')—in which parties (not just network companies) bid for a fixed pot of money for projects that are difficult to incentivise networks to undertake, partly because of the current length of the price control periods, and partly because the benefits may accrue to a number of parties other than the networks themselves;
– more focus at price reviews on the 'outputs' that networks are expected to deliver, including outputs relating to the underlying health of existing network assets.

What is the link between long-term planning and day-to-day delivery?

As with other UK privatisations, at least the earlier ones, the original gas and electricity regulatory regimes were fairly non-specific about what the privatised companies were expected to deliver. All networks were required to develop and operate 'efficient' networks, but there was little specification of what efficiency might entail beyond the various 'Overall' and 'Guaranteed' Efficiency Standards that applied to the distribution businesses in particular.

Since privatisation—especially since the late 1990s—the incentives on network companies, and in particular distribution companies (not least because of their direct interaction with end-consumers), have been both broadened and sharpened in order to deliver a range of outcomes. The 'Incentives and Obligations' paper, published as part of the final proposals for

DPCR5, gives a flavour of these incentives.[127] Thus, even a glance at the Contents page reveals:

- a Distributed Generation Incentive, to encourage DNOs to anticipate the requirements of generators that will want to connect to their networks;
- a Losses Incentive, to reward the reduction of electricity losses on distribution networks;
- penalties for poor performance in providing new connections to customers;
- proposals for an incentive based on a 'broad measure of customer satisfaction';
- an incentive in relation to telephone response (due to be replaced eventually by a customer satisfaction incentive);
- the Interruptions Incentive Scheme, largely based on the frequency and duration of customer interruptions;
- an incentive based on service provided to 'worst-served customers' (ie, those who typically experience a relatively large number of interruptions);
- the Information Quality Incentive, sometimes known as 'menu regulation', which is designed to encourage companies to submit accurate forecasts of their expenditure requirements;
- new 'network output measures' (elaborated on below);
- an Innovation Funding Incentive, the model for the Innovation Stimulus referred to in the previous section.

One particular feature of the above is worth noting (especially in relation to the attempt to link day-to-day performance with network investment and longer-term planning): Ofgem has often noted the improved 'performance' of energy networks since privatisation. In relation to electricity networks, this has usually referred to improved network reliability performance—ie, fewer and/or shorter customer interruptions.[128] However, in the more recent past, Ofgem has shown concern about whether this improved customer performance has co-existed with a decline in the underlying health of the networks—in a similar way to Railtrack's relatively good punctuality performance in the period running up to the partial shutdown of the rail network in the wake of the Potters Bar and Hatfield incidents. This has led Ofgem to focus increasingly on the delivery of outputs that are more closely related to the underlying health of network assets—asset failure rates and other broader assessments of asset health. This work is still at a relatively early stage, as is the way in which new measures may feed into regulated revenue (both are covered in Chapter 19 of the Incentives and Obligations paper, referred to above). Broadly, the intention is that any company that has failed to deliver required outputs would suffer a financial penalty that would be imposed at the subsequent price review.

How are the networks encouraged to reduce costs, while continuing to deliver the required outputs?

The answer to this question follows on from the previous section and can be illustrated by the mixture of cost-reduction incentives and output-delivery incentives put in place for DNOs by DPCR5.

As far as cost-reduction incentives per se are concerned, the DPCR5 outcome has refined these incentives in one main way—specifically, there has been an attempt to equalise the incentives between OPEX and CAPEX. This has been achieved by allocating *network* spend (as opposed to *'business support'* costs which continue to be treated as OPEX) between CAPEX and OPEX (for price control purposes) according to a formula, rather than allowing the company flexibility as to whether it counts its spend as OPEX or CAPEX. However, it needs to be recognised that, as long as Ofgem continues to *benchmark* OPEX and CAPEX separately (as it did in DPCR5), the incentives remain for companies to lean towards CAPEX solutions, for example.

[127] Ofgem (2009), 'Electricity Distribution Price Control Review Final Proposals – Incentives and Obligations', December 7th.
[128] On gas networks, unplanned customer interruptions are relatively rare (and typically far more serious in terms of what is entailed for restoration of supply).

The main way in which Ofgem has tried to balance incentives between cost reduction and output delivery is as implied at the end of the previous section—ie, it has maintained the threat that failure to delivery on outputs will result in financial penalisation at the next price review. However, companies might still be seen as having stronger financial incentives to cut costs than to deliver outputs, in that:

– the benefits of cost reduction are built into the price control mechanics, whereas the penalties for failure to deliver some outputs (especially those relating to asset health) are not;

– the new outputs regime is still very much 'work in progress', with significant scope for subjectivity and self-assessment by companies in judging whether the required outputs have been delivered.

What is the role of the SO and how is it incentivised to manage the electricity system efficiently?

The role of the SO has expanded over time, not least in response to changes in the industry structure. At the core of the role is management of the electricity system in real time—to match generation with demand in real time, to minimise the cost of transmission constraints in causing generating plant operation to depart from the wishes of generators, and to minimise the other costs of day-to-day system operation.[129] However, the role of the SO has also changed significantly in response to:

– the integration of the Scotland and England/Wales electricity markets with the introduction in 2005 of BETTA (the British electricity trading and transmission arrangements);
– the ongoing creation of a commercial and regulatory framework for offshore transmission.

BETTA has meant that National Grid is now both the SO for the whole GB electricity market (whereas, previously, Scottish Power and Scottish & Southern Energy were responsible for system operations in their respective areas of Scotland), and the commercial face of transmission for all GB users of the GB transmission system. For example, National Grid, in its SO role, is responsible for proposing how transmission is charged for—within the constraints set by transmission price controls—and for making connection offers to generation developers wanting to connect to the GB transmission system.

Offshore transmission has posed new issues. National Grid is the SO but will also be competing with other potential offshore transmission operators (OFTOs) to provide links to offshore wind farms. As a result of the potential conflicts of interest arising from this—and under current plans—Ofgem will run the competitive tenders which will lead to the appointment of OFTOs. However, while in opposition, the Conservative Party implied that it had a rather different vision of how offshore transmission should be developed, and it remains to be seen whether the offshore regime will evolve as Ofgem and DECC were previously planning.

In general, incentivisation of the SO will depend on how 'deep' the role of the entity is: a 'shallow' SO would be simply about day-to-day balancing of electricity supply and demand; a 'deep' SO would be about the planning and development of the transmission system, and could have wider responsibilities.

[129] Unlike with most electricity systems round the world, the role of the NETSO is not to 'despatch' generating plant according to its view of what are the cheapest (ie, lowest short-run marginal cost) plant in the system. The GB wholesale electricity market works, in the first instance, on the basis of self-despatch by generators (ie, generators telling the NETSO how much they want to generate in each half-hour period)—it is only close to real time that the NETSO, in effect, takes over and buys 'increments' and 'decrements' of generation from generators (relative to the generators' own submitted plans) to balance the system.

In effect, the investment/deeper aspects of transmission are handled by the transmission network operators (TOs)—ie, National Grid (in its TO role) and the two Scottish transmission businesses—and incentivisation of this role is broadly as outlined above.

Thus, National Grid, in its SO role, has two main areas of activity:

– day-to-day system operation;
– the commercial relationship between transmission and network users, including the charging for, and rationing of, existing network capacity.

The second of these two activities has been the subject of a long-running government/Ofgem/industry process called the 'Transmission Access Review'. This has been broadly about the terms on which new generation should connect to the transmission system, and, in particular, who should pay for the cost of any transmission constraints arising from new generation connecting to particular parts of the system. Broadly, those who emphasise economic 'efficiency' have preferred the idea of focusing the incremental transmission costs on those who cause them (the new generators in the relevant locations), while those who emphasise the need for a speedy build-up of renewable generation (much of which is, as already noted, located far from demand and where the existing transmission system is either weak or non-existent) advocate socialisation of the incremental costs.

At least at a high level, this debate would look to have been settled with DECC's publication of its latest consultation on transmission access,[130] albeit the new government may take a different view. In effect, the DECC consultation proposed that:

– the SO should offer to connect any new generation that is prepared to make a certain level of firm financial commitment and, subject to the local connection existing, on the generator's desired timescale;

– any extra transmission constraints caused by the new connection—as a result of lack of capacity in the 'deeper' transmission system—should be socialised across all transmission users (both generators and suppliers).

If these proposals are implemented, the main respect in which the SO will need to be incentivised is to minimise resulting transmission constraints for any given level of transmission capacity, in light of the fact that the TOs will be responsible for building the new transmission capacity. In other words, this reinforces the need to incentivise the SO to perform efficiently its task of managing the system efficiently day to day. (This is not to say that there is not an alternative model here—where the SO would be responsible for determining the optimal pattern of investment in new transmission capacity.)

As far as incentivisation of the SO's day-to-day system management is concerned, this is mainly handled through an SO incentives mechanism.[131] This mechanism is set annually, although Ofgem is looking to move to multi-year arrangements, as this might encourage the SO to exert more pressure on the TOs to invest optimally in the transmission system in order to reduce transmission constraints. The main argument for a one-year mechanism has been the difficulty in forecasting many of the costs facing the SO, and the extent to which these are outside the SO's control.

Under the one-year SO incentive mechanism:

– a target level of external SO costs is set;

[130] DECC (2010), 'Improving Grid Access – Technical consultation on the model for improving grid access', Consultation Document, March 3rd.
[131] This SO incentives mechanism covers purely the 'external' costs of system operations—for example, the cost of transmission constraints and of paying reserve generating plant. Revenue to cover the SO's 'internal' operating costs is set at the normal periodic review.

– to the extent that costs turn out lower or higher, the SO is rewarded or penalised—with both the upside and downside limited by a cap and collar respectively.

A3.2 Aviation

This section evaluates interface issues in the GB aviation sector. The boundary between the main players in the market—airports, airlines, and air traffic control—has been relatively stable since the privatisation of BAA in the mid-1980s. Nevertheless, the management of the interface between these three players has undergone some change over the past decade. For example, the sector has pioneered the development of the (not wholly successful) regulatory mechanism of 'constructive engagement', whereby elements of capital projects and capacity development are negotiated between the regulated airports and the carriers (and, more recently, between the air traffic control provider, NATS, and the carriers).

The experience of the sector is also of interest given that it competes with rail services in certain markets. It is an industry with high public visibility, in which safety and security concerns are paramount, and which has witnessed an increasing degree of conflict between third-party users of infrastructure (airlines) and the owners and operators of that infrastructure (airports), including with respect to operational performance, and the funding and specification of capital investment.

A3.2.1 The value chain in the aviation sector

As noted, three key players are involved in conveying passengers from one location to another: airports, airlines and air traffic control.

Airports constitute a number of assets used to facilitate travel by passengers and freight users, the most important being runways and associated facilities used by aircraft (such as aprons and taxiways), and terminals for handling the processing of passengers and freight. Airports will also typically use other complementary infrastructure, such as maintenance hangars, user car parks, and surface-access connection points. A significant proportion of the floor space of passenger terminals is usually let on a concessionary basis to commercial third-party providers to provide non-aeronautical services to users.

The OFT's reference to the CC in relation to the supply of airport services by BAA in the UK provides a useful summary of the scope of airport services:

> 'airport services' means all airport services whether they are supplied to airlines, ground handlers, passengers or any other person and includes all or any of the following:
>
> *(a)* the provision of airport infrastructure services (including the development, maintenance, use and provision of access roads, runway facilities, fuelling facilities, taxiways, aprons, stands, loading bridges, piers and gaterooms, check-in and arrivals concourses, check-in desks, customs and immigration halls, baggage facilities, passenger care facilities, lifts, escalators, passenger conveyors, terminal offices, ramps, and other airport structures);
>
> *(b)* the co-ordination and control of the activities performed on or in airport infrastructure and the provision of associated services including security services; and
>
> *(c)* the provision (including the provision by persons other than BAA under arrangements made with BAA) of associated commercial services (including catering services, retail services, car rental services, the sale of advertising space, the provision of car parking, and activities relating to commercial property).[132]

Airports fund themselves through a mixture of aeronautical revenue (derived from charges for use of infrastructure such as landing charges) and from non-aeronautical revenue, which encompasses the commercial activities described above. The aeronautical charges of three

[132] OFT (2007), Terms of reference for investigation into supply of airport services, March

GB airports—Heathrow, Gatwick and Stansted—are subject to ex ante price regulation by the UK Civil Aviation Authority (CAA). The overall yield paid by users—including at the three regulated airports—is calculated on a 'single till' basis (similar to that for Network Rail)—where commercial income is netted off the allowed revenue of the airports that is subsequently recovered from users.

Although there are local variations, airports are responsible for maintaining this infrastructure, and for providing related services, such as perimeter security, fire-fighting facilities. They also have responsibilities in relation to the allocation of terminal resources between airlines and between concessionaires of terminal space. In contrast to terminal capacity, 'congested' airports do not have discretion as to how runway capacity is allocated to different airline users. Under Council Regulation EEC 95/93, capacity at congested airports is allocated according to incumbency ('grandfathering'), so that existing users retain slots (the right to land aircraft at a particular time, at a particular airport), subject to rules governing the frequency with which these slots are used. Some slot switching and trading of slots does occur; future debates in relation to slots are considered below.

A number of other airport services are not provided by the airport operator, such as check-in processing, baggage-handling, and the maintenance of aircraft, which may be performed by different commercial parties in each case. Government agencies are responsible for operations such as policing, immigration and customs control. It is notable that the performance of all of these parties, over whom the airport often has limited or no control, affects user perception of the quality of the airport.

One important area of interfaces for the industry as a whole is security. In some cases, airports themselves undertake passenger search, whereas in others these services are contracted out. In both cases, the government will undertake close oversight of standards. There are also some international examples of security being provided by the government (eg, in the USA). Despite the different models of security provision, there are no obvious interface problems. This is due in large part to the fact that security is, for the most part, a well-defined and 'stand-alone' activity and there is perhaps also a recognition and acceptance of its importance by all stakeholders.

Airports are owned by a variety of public and private sector operators. Some airports are wholly in the private sector (eg, Gatwick, London City, Exeter, Aberdeen), some wholly in the public sector (eg, Manchester Airport, which is owned by the Council of the City of Manchester and nine borough councils), and some are part-public, part-private (eg, Birmingham Airport, which is 49% owned by seven West Midlands District Councils, 2.75% by an employee share trust, and the remainder by private investors).

Airports also interface with surface-access providers, such as buses, taxis and train companies.

Airlines, which convey passengers and freight from one destination to another, are subject to a wide variety of regulations, but not ex ante economic regulation per se. This is a consequence of the progressive liberalisation of the USA and European markets (largely completed within Europe in the early 1990s), until which point they were regulated ex ante by national governments with respect to prices and frequencies.

Airlines differentiate themselves in several ways, including by type of route offered (long or short haul) and by business model, with three main models being commonly referred to:

– full-service airlines offer a reasonably extensive range of routes and services, which may include both long- and short-haul services;

– low-cost carriers (LCCs) operate a business model based on rapid turnaround of aircraft fleet (which, in order to simplify maintenance, is usually of a single aircraft type), and unbundling of many of the services offered by full-service airlines;

  - Except for core in-flight activities, LCCs tend to contract out most activities and therefore have interfaces with aircraft maintenance and baggage handlers;

  - LCCs also differ from full-service airlines in how they interface with passengers, and generally focus on telephone or online contact, and do not operate through travel agents;

- charter airlines sell blocks of seats to package holiday operators or to other groups for infrequent or 'one-off' journeys.

There are also non-passenger carriers serving freight markets.

Air traffic control involves two types of activity:

- at the airport itself, it covers the take-off, landing, and surface movement of aircraft—this is analogous to platform management at rail stations;

- en route, it involves regulating traffic over designated airspace—ie, a much wider area than airport air traffic control. This can be broken down further into aircraft management in cruise, and approach services, which are only material in the congested London area.

These two types of activity are carried out at two separate locations: from a control tower at the airport in the case of the former, and from an operations centre in the case of the latter. There are two control centres covering the UK, one in Hampshire (responsible for the London area among others) and one in Prestwick (covering Scotland and the North of England among other areas).

The two activities must be closely coordinated. There are therefore repeated interfaces in the sky between air traffic control and the aircraft; an aircraft on a European flight can be passed between many different controllers as it crosses European airspace. No significant problems are apparent with the operation of this regime, despite the number of possible interfaces and their safety-critical nature, although there are movements to streamline the system.[133]

NATS is the main UK air navigation service provider, providing en-route services and services to 16 UK airports, for which it competes with other providers. Some airports perform their own airport air traffic control, while others contract to NATS or their competitors. The charges associated with NATS' en-route services are regulated by the CAA, which also oversees the licence imposed on NATS, which includes responsibilities in relation to providing access to airspace for all users, and making the most efficient overall use of that airspace.

The main control on the level of costs in the industry is competition. Only three of the UK's 24 airports with more than 500,000 passengers per annum are subject to ex ante price regulation, and the number of price regulated airports could fall further. As such, competition between airports and between airlines, is an important determinant of the level of cost in the industry, and therefore on how costs fall on different parties in the industry value chain.

Depending on the type of journey undertaken (eg, distance, full-service carrier versus low-cost carrier), a significant proportion of the costs of an airline ticket can be made up of taxes and other government duties, and the margin charged by airlines. The proportion of costs associated with regulated airport charges can therefore be relatively small, although this proportion is higher in the case of cheaper tickets, such as those typically sold by LCCs on short-haul, point-to-point journeys.

[133] These include the Single European Sky ATM Research (SESAR) programme, and the European Commission's 'Single European Sky initiative'.

For example, the 2012/13 regulated yield set for Gatwick Airport at the most recent price control review was £7.34.[134] For some tickets from that airport (eg, transatlantic flights), it is apparent that the proportion of the ticket borne by the user will be very small, whereas for a short-haul LCC flight the proportion will be higher.

A3.2.2 Post-privatisation landscape[135] and subsequent changes

The privatisation of BAA in 1987 did not have a large impact on the number of interfaces operating between different parties in the industry. The government chose not to divide BAA horizontally into separate airports, and did not promulgate any form of vertical separation, since such separation was, for the most part, already present in the industry at the key junctures of the value chain.

It is also notable that a significant proportion of employees at the BAA airports at the time of privatisation were not directly employed by the operator. This indicates the extent to which interfaces between BAA and other parties were already established and understood when the assets were sold to the private sector. However, there were some initial tensions between BAA and other parties after privatisation in the face of a more aggressive commercial stance (eg, the introduction of charges for taxis to use Heathrow's taxi compound—see Doganis 1992[136]).

The privatisation had been preceded by an MMC inquiry in 1985 in relation to BAA's policies towards commercial activities. Despite concerns about abuse of a monopoly position, the MMC concluded that BAA was not pursuing a course of conduct contrary to the public interest in this area; however, the MCC did make certain recommendations concerning, for example, the promotion of competition in tendering and the presentation of accounting information showing the separate contributions of aeronautical and non-aeronautical activities.

Since privatisation, there have been only a limited number of developments with implications for interfaces and associated incentives, albeit there have been other industry developments of wider and more general interest. BAA sold Gatwick Airport in 2009, reducing from three to two the number of large airports it controls in the south east of England.

One of the more significant developments for interfaces—EC Regulation 95/93—has been mentioned in the previous sub-section. This ultimately gave effect to an interface between airports, airlines and Airport Coordination Limited (ACL), the body responsible for slot coordination at all the major UK airports. In addition to slot allocation, ACL provides schedule facilitation and schedule data collection services. It is an independent body owned by nine of the largest UK airlines; any airline is eligible to join and member airlines are not given preferential treatment in allocation and other activities. Most of its funding comes from the airports to which it provides services, with the remainder coming from airlines and commercial activities such as training and consultancy.

Changes in ground-handling have affected interfaces in the industry. Ground-handling refers to a variety of services, including baggage-handling, passenger check-in, aircraft cleaning and aircraft refuelling. Directive 96/67/EC opened up the market for ground-handling services at airports, although it allowed Member States to impose certain limitations on the number of suppliers in four service categories, subject to compliance with related obligations, including the transparency of the tendering process. Airports retain responsibility for appointing handlers, subject to the provisions of the Directive.

Air traffic control has also changed in the period since BAA's privatisation. The Transport Act 2000 mandated a form of public–private partnership for NATS, which was subsequently partially privatised in July 2001, with a consortium of seven airlines taking a 46% stake in the

[134] CAA, (2008), 'Economic Regulation of Heathrow and Gatwick Airports 2008–2013', March.
[135] In this section, 'privatisation' refers to the privatisation of BAA.
[136] Doganis, R (1992), *The Airport Business*, London: Routledge, p. 32.

business, which involved debt of £748m and equity of £50m. NATS was exposed to revenue risk under the system of charges regulation overseen by the CAA, and the highly geared structure collapsed after the decline in air traffic following 9/11. Ultimately, new equity was introduced by the government, and by BAA as a new equity partner.

There have also been significant changes in the airline industry, particularly in relation to the emergence of LCCs, which have supported traffic growth and increased the pressures on capacity at many airports. The growth of these carriers was, in turn, supported by efforts from the 1980s onwards to liberalise European airspace. From an interface perspective, these types of carrier have different requirements to more traditional, full-service carriers, and this has led to tensions and new forms of regulatory engagement, explored in more detail below. More generally, there has continued to be a degree of consolidation in the industry.

A3.2.3 Similarities with and differences to GB rail value chain and industry structure

The most immediate similarity between the rail and aviation value chains is that they are both modes of transport, and indeed competing modes on some domestic and international routes. The following describes some differences.

- **No physical network** linking airports to one another. Unlike rail, there is no physical infrastructure that must be maintained and renewed in order for journeys to be made. Therefore problems in relation to possessions do not arise to the same extent, albeit congestion and delays are nevertheless features of airspace management.

- **Significant international dimension**. The international aspect of air travel is significantly more important than in rail, which has interface implications with respect to the need to coordinate airspace movements with a number of sovereign countries.

- **Economies of scale**. The phenomenon of hubbing is particularly related to the international dimension of aviation, and exists to a greater extent than in rail. For example, the fact that Heathrow Airport is the UK's only international hub has interface implications with respect to the requirement to fulfil carrier needs in relation to transferring passengers and baggage between gates, terminal.

- **Compensation regime**. Both industries oversee a compensation regime, although it is sufficiently different in substance to qualify as a difference rather than a similarity.

Some of the main similarities between the sectors include the following.

- **Mix of public and private interfaces**. As in rail, there is repeated and frequent interaction between the public and private sector, both within organisations (such as NATS), between privately owned airports and publicly owned carriers, and between publicly owned airports and private carriers. Some of the core infrastructure remains in public ownership, although its size and value is less than that owned by Network Rail.

- **Safety issues and public visibility interfaces**. As both industries have extensive dealings with the general public, there is a variety of interfaces connected with safety in both rail and aviation, as well as in the context of more general consumer protection issues.

- **Timetabling and central coordination**. Airspace and airports have limited capacity, and there is a need for a degree of central coordination and of timetable management between airports.

- **Infrastructure owners' control of infrastructure**. In both industries, the infrastructure owner cannot fully control the identity of the intermediate infrastructure user. In the case of rail, passenger franchises are awarded by the DfT; at congested airports, slots are allocated according to the grandfathering principle and their use is governed by ACL.

– **Single till**. Both industries operate under a single till, although its significance is much greater in aviation. Nevertheless, regulatory tills can have important implications for incentives, and are therefore relevant to this project.

A3.2.4 Evidence on the cost of changes

There is little evidence on the changes in interfaces and vertical structure in the industry that is of direct relevance to this study. This is because many of the interfaces were already present in the industry when the BAA airports were privatised (which is used as a focal point in this case study) and the broad scope of interfaces between airports, airlines and air traffic control is established. This is not to say that the operations and the alignment of incentives in these industries is without interest for this study, since there have been tensions—constructive and otherwise—over this period, but these have been less numerous and less wide-reaching than in rail and in the energy networks. The evidence on the operation of these interfaces, and the incentives on the different parties involved, is the subject of the next section.

Evidence on the operation of interfaces and alignment of incentives

One of the most contentious interfaces in aviation, particularly at BAA's regulated airports, has been in relation to **CAPEX planning**. The CAA introduced a process of constructive engagement in essence to address the problem of how to decide on the appropriate amount of CAPEX to allow when making regulatory determinations and determining an overall revenue requirement for the regulated businesses. In one sense, it could be seen as an attempt to improve the operation of the interface between infrastructure owner and operator (airport) and intermediaries providing services to end-users (airline carriers). It also reflected, to some extent, a belief that direct negotiation between providers and users, with limited regulatory intervention, could result in regulation that was more simple and transparent than alternative means of determining the appropriate quantum of CAPEX.

However, the process encountered serious problems. It was not suited to resolving rational differences between airports and their users on investment priorities. In particular, the model arguably did not place sufficient emphasis on the needs and opinions of important stakeholders such as government and passengers. In effect, the interface failed because it did not properly account for the inherent misalignment in the incentives of the two main parties.

The CAA's intention was that the 'Q5' review (covering the period of charges at Heathrow and Gatwick from 2008 to 2013, and from 2009 to 2014 at Stansted) would be based 'to the greatest extent possible'[137] on direct negotiation between airlines and airports, so that any agreements made in this negotiation process would be adopted by the CAA in setting the price control.

The initial intention was that some key elements of the control would be resolved by negotiation. These elements included future volume and capacity requirements, the capital investment programme, the level of service quality, opportunities for OPEX efficiencies and commercial revenues.

A Beesley Lecture in 2008 analysed the outcome of constructive engagement at BAA's three designated airports:[138]

– traffic forecasts were not agreed, and the position of both sides changed over the course of the review process;
– while some service quality parameters were agreed, significant differences remained in key areas;
– nothing significant was agreed on operating costs;
– nothing significant was agreed on commercial revenues;

[137] CAA (2005), 'Airport Regulation: The process for constructive engagement', Industry Seminar, June.
[138] Toms, M. (2008), 'Airport regulation: a case of destructive engagement', Beesley Lecture, October 9th.

– the parties remained far apart on incentive mechanisms.

Ultimately, the regulator could not test whether the outcomes of the negotiations, such as they were, in any sense represented an 'optimal' solution. An important reason for this failure was that the incentives of each party were fundamentally misaligned in many respects. Airlines were being asked to pay for CAPEX that could benefit rivals (since capacity expansions would support entry). Airports were in a regulatory negotiation that affected about 50% of their revenue, while the airlines were playing for a much small proportion of their overall cost base.

The design of the negotiation, and this misalignment of incentives, is reflected in the problems that emerged from the process:

– the process required a cooperative attitude among all the parties in the negotiation;
– there were important disparities in the bargaining resources and time horizons of airlines and airports;
– the process gave precedence to the view of existing airlines, with no account taken of future users;
– there was no role for passengers;
– airlines were being asked to commit to CAPEX without knowing the terms of the regulatory settlement that would remunerate that CAPEX.

What lessons for rail can be drawn from this experience? A basic message is that, even where users of infrastructure have a financial stake in the outcome of a regulatory settlement, negotiation alone cannot overcome misaligned incentives that otherwise exist. In turn, recognising these incentives should encourage regulators to consider alternatives to negotiation. This is not to say that the parties should not be encouraged to share information and to consult each other, but to accept that a regulator will ultimately need a means of satisfying itself that outcomes are acceptable, consistent with its statutory duties, and meet other relevant criteria.

This alignment of incentives also manifests itself in debates over the type of **facility** and **service standard** expected at airports. In particular, LCCs tend to favour basic, low-cost terminal facilities, whereas full-service carriers tend to demand higher standards, and may require different types of facility (eg, to process transfer passengers).

In both cases, airlines are concerned about the long-term quality of service and appearance of the airport, even where there are differences. However, airlines tend to have a shorter planning horizon than airports, which makes long-term contracting difficult. In addition, competition between airports may mean that the mobile capital embodied in airlines can transfer to an alternative facility.

There is therefore a tension and a trade-off between:

– an airport's willingness to commit to providing services heavily tailored to meet the needs of one user, or one type of user, in a world of imperfect contracting; and
– a desire to fulfil the requirements of these users at the risk of making the infrastructure less attractive to other types of user. Regulators also confront this problem when determining the appropriate scale and specification of CAPEX at regulated airports.

Comparisons with the situation in rail are helpful. For example, a long-distance TOC using routes going west from Paddington cannot avoid using Reading Station, but may not be unduly concerned about its appearance after the expiry of a franchise, or in the latter stages of a franchise. An airline can avoid a particular airport if a competing facility allowing the same markets to be served is available, but this will not be possible in all cases. Airlines will put pressure on the airport operator to ensure that suitable facilities are available at an appropriate price—either basic facilities in the case of LCCs, or more elaborate facilities in the case of full-service carriers. However, a degree of friction remains as to how the service

standards are monitored, overseen and defined, as part of either bilateral or regulatory settlements. Regulatory intervention in the future might be in the form of more extensive use of performance metrics.

One problematic set of interfaces has been that between the **regulator, airport** and **government**. The CAA's statutory duties, as *currently* constituted,[139] do not oblige the regulator to implement government aviation policy, or to have regard to any such policy. Consequently, pending future changes, there is no interface that is comparable to the HLOS process in rail, whereby the government sets out the outputs that it requires the sector to produce. The complications at this interface were demonstrated by the debate around the 2003Aviation White Paper.

This White Paper was the result of a five-year process involving a full environmental and economic analysis of the case for airport expansion at several different sites, and consultation with 360,000 people. The economic analysis found that, of the options considered, the highest benefits (in net-present-value terms) would be secured by developing runways at both Heathrow and Stansted. The White Paper therefore proposed that a new runway be built at Stansted by 2012, and a new short runway at Heathrow by 2015–20, but only if strict environmental tests could be met.

BAA began developing a runway at Stansted. However, the expansion was opposed by the CAA, as well as the dominant carrier at the airport (Ryanair). A planning application was submitted in 2007, and BAA spent over £100m in related, initial work on the project. Ultimately, about half of this spend was disallowed in the 2008 periodic review of Stansted charges, despite BAA's pursuit of the government's White Paper policy intentions. In this case, there was a clear disconnect between the government's policy and how the CAA treated the airport operator's expenditure to support that policy.

The situation at Heathrow was somewhat more complicated by environmental and social issues, including with respect to air noise, air quality, surface access, and the loss of homes on sites near the airport. In January 2009, the government confirmed support for the new runway, but this policy was frustrated by a court ruling in March 2010, which found that the government had failed to take account of climate change law and policy and to discharge its own conditions requiring good public transport, and by the policy position of the new coalition government.

Whatever view is taken about the economic and social desirability of new runways in the south east of England, there are problems associated with the regulatory lacuna described above in relation to regulatory risk (for the operator) and social welfare (for society as a whole). The Labour government initiated a review of the economic regulation of airports in 2008, and its proposals included a new set of statutory duties for the CAA, one of which was 'to have regard' to government policy. This is discussed in more detail below in relation to future debates in the sector concerning interfaces.

Investment coordination at the three largest London airports was also affected by the decision in 2003 to regulate Heathrow, Gatwick and Stansted on a stand-alone basis. Previously, all three airports had *not* been regulated on an individual basis, which allowed for transfers from the two larger airports to Stansted in order to support the development of capacity. In turn, the availability of this capacity supported the emergence of low-cost carriers such as Ryanair and easyJet, which helped to drive average air fares lower.

[139] The CAA's existing statutory duties are defined in the Airports Act 1986. The CAA must perform its functions with respect to economic regulation to: 'i. further the reasonable interest of users of airports within the UK; ii. promote the efficient, economic and profitable operation of such airports; iii. encourage investment in new facilities at airports in time to satisfy anticipated demands by users of such airports; and iv. impose the minimum restrictions that are consistent with performance by the CAA of its functions.'

It is a matter for debate as to whether this competition would have emerged without the integrated approach taken to investment planning facilitated by the 'systems' approach to regulating the three airports, and whether the 'system'-based approach to the regulation of BAA's three airports in the south east delivered the right outcomes for users.

In 2002 the European Commission published a study[140] examining the impact of the **ground-handling** Directive (96/97/EC) on the quality and efficiency of these services in EU airports. The report found that the Directive had had a limited impact in the UK, since the market already had several active service providers; for instance, Heathrow had the same number of self-handlers and third-party handlers before and after the Directive.[141]

The report did note that the Directive appeared to have been successful in terms of increasing the number of service providers operating at a number of European airports, although this effect could be attributed to general market developments towards greater competition, rather than the Directive per se. Overall, the conclusion in relation to interfaces appears to be that greater competition has not materially impaired airports' ability to provide services to the standards demanded by users, and that, in some cases, airports have volunteered to increase the number of handlers (ie, interfaces) at their airport. For example, in 1998 Gatwick Airport applied to the CAA to limit to four the number of airside handlers operating at Gatwick, based on concerns regarding space and capacity. Despite the objection of two handling companies, the CAA approved this application, taking into account the relevant provisions of the Directive. Gatwick subsequently applied to have some of these restrictions lifted.[142]

One could interpret this type of application, and the movement towards more contestability in this area, as reflecting a positive view of the benefits of greater competition in handling services, and a confidence in airports' abilities to manage a greater number of interfaces with more suppliers. However, some problems were encountered in relation to apron vehicle congestion caused by the proliferation of ground handlers. The experience also indicates the importance of tailoring regulatory interventions to the market concerned; in the case of ground-handling, there was little reason to assume that the services provided could, in all cases, be supplied by only one organisation. Additionally, entry could be facilitated by the relatively straightforward and labour-intensive nature of the industry.

There has also been friction at the **performance and compensation** interface. Apart from compensation that may be due under airport-specific performance measures of the type discussed above (eg, in relation to cleanliness, availability of departure lounge seating, etc), there is also a performance regime mediated by airlines. EC Regulation 261/2004 established rules concerning compensation available to passengers in the event of denied boarding, cancellation and long delay of flights. The regulation defines three levels of compensation, whose cash amount depends on the distance of the flight. European Court Judgments in November 2009 also entitle passengers whose flights have been delayed to the same rights of compensation.[143] There are also rights in respect of compensation for re-routing and refreshments, communications and accommodation. The regulation is contentious because it entitles airline passengers to compensation for delays or cancellation that may not be under the control of the airline concerned. For example, there is no formal provision for compensation between jurisdictions for delays caused by problems in national air traffic control. Another important cause of delay is adverse weather conditions, an issue brought into sharp focus by the disruption caused by the Icelandic volcano eruptions of April and May 2010.

[140] SH+E (2002), 'Study on the quality and efficiency of ground handling services at EU airports as a result of the implementation of Council Directive 96/67/EC'.
[141] CAA (2005), 'European Council's Ground Handling Directive: Evidence Of Enhanced Competition', Supplementary Memorandum by the CAA (EU 15A), Select Committee on Transport, April 1st.
[142] CAA (2006), 'Notice published in accordance with paragraph 8 of schedule 1 To the airports (groundhandling) regulations 1997. Ground handling at Gatwick Airport', May. http://www.caa.co.uk/docs/5/ergdocs/gatwickmay06.pdf
[143] Joined cases C‑402/07 and C‑432/07.

The regulation exposes airlines to liabilities for compensation that may be significantly greater than the value of the tickets purchased. This raises questions about fairness, and about the sustainability of a regime that holds parties responsible for delays and cancellations that cannot be controlled or reasonably mitigated.

It is notable that many countries, including the UK, have opted out of a similar regime in the rail industry until 2024 under the terms of Regulation 1371/2007.[144]

Debates about the future of the industry as they relate to interfaces

The interface involving government policy is likely to be addressed in a fairly direct way by new statutory duties for the CAA. One of the five proposed supplementary duties set out in the December 2009 decision document was:

> to take account of guidance issued by the Secretary of State, and to assist in delivery of airport infrastructure consistent with the National Policy Statement on Airports unless there are compelling reasons not to do so; [145]

Two uncertainties remain in relation to how this interface will be changed by this duty. The first is the form in which the new coalition government will implement, if at all, the policy position taken by the previous administration. The second—and perhaps more fundamental—uncertainty is whether this duty will resolve the problem it is intended to address, namely the regulatory lacuna with respect to the role of government policy. The drafting of the duty in the decision document would still appear to afford the regulator a degree of discretion in how it interprets the duty, and it remains to be seen how this interface might change if this form of duty is introduced.

It also remains to be seen how a new primary duty to 'promote the interests of existing and future consumers of passenger and freight services' will affect interfaces. One consequence may be to increase regulatory scrutiny in relation to performance, leading to greater formalisation of performance regimes.[146]

Another possible future interface change arising from the DfT's review of regulation is the possibility of within-airport terminal competition. The decision document stated that the regulatory framework should not preclude the 'possibility of separating the operation and development of terminals' at airports with substantial market power and where there is the potential for effective terminal competition to develop. This would involve adding another interface to the operation of airports, and its benefit remains to be demonstrated. The CAA has expressed reservations about the possibility of terminal competition, being 'doubtful about the merits of adding another interface, whether it would enable airlines to negotiate better terms, and whether it would in practice increase efficiency'.

A further source of possible change is in relation to the slot regime. Under existing arrangements, the question of 'who owns slots' cannot be easily answered, with possible claims from the state, airports and airlines. A recently announced review by the European Commission will cover this issue. The long-term consequences of the review could have significant implications for the allocation of capacity in the aviation market.

These consequences will depend on which party is deemed to own slots. For example, the slots could be described as the property of the state (as in the USA), and governments may wish to raise revenue by auctioning rights to use these assets. If airlines own the slot, they could conceivably be capitalised and would support their balance sheets. If airports were deemed to own the slots, perhaps on the grounds that they take the risks associated with

[144] Veneables, R. and Healy, C. (2010), 'The case for reforming EC Regulation 261/2004', Bird and Bird. http://www.twobirds.com/English/News/Articles/Pages/EC_Regulation_2612004.Aspx
[145] DfT (2009), 'Consultation on proposals to update the regulatory framework for aviation', December.
[146] The coalition government's 'South East Airport Task Force', announced on June 15th, does not immediately appear to have major consequences for interfaces or industry structure; rather, it will be looking at how passenger experience can be improved, given announcements concerning restrictions on capacity development in the region.

capacity development, then slots could be priced on a congestion basis, and their value would be determined by a market-driven process of capacity allocation. One of the key provisions of the existing slot regime—the 'use it or lose it' rule—was relaxed in summer 2009 in order to allow carriers to adapt to the effects of the economic downturn, and to reflect the increased availability of capacity in the industry.

A3.3 Water

The water industry has remained vertically integrated since privatisation. This section assesses interface issues that have arisen in this context, including developments that have changed, or are likely to change, the role of interfaces in this sector.

– First, although Ofwat has never mandated any type of separation, some of the companies have, on their own accord, outsourced a substantial part of their operations. The best-known example is Welsh Water, but other water companies have followed this example to some degree. The co-existence of 'integrated' and 'outsourced' business models for regulated activities potentially creates a fertile ground for comparison. This configuration has also generated some regulatory issues, and Ofwat has had to specify licence conditions to make sure that outsourcing would not blur the lines of accountability with respect to licence duties (see below).

– Second, the regulator and the government are currently contemplating the separation of the retail function and introducing competition in this part of the value chain. As a consequence, the discussion here is concerned more with future developments than the above energy and aviation case studies, although other relevant issues are also assessed.

A3.3.1 The value chain in the water sector

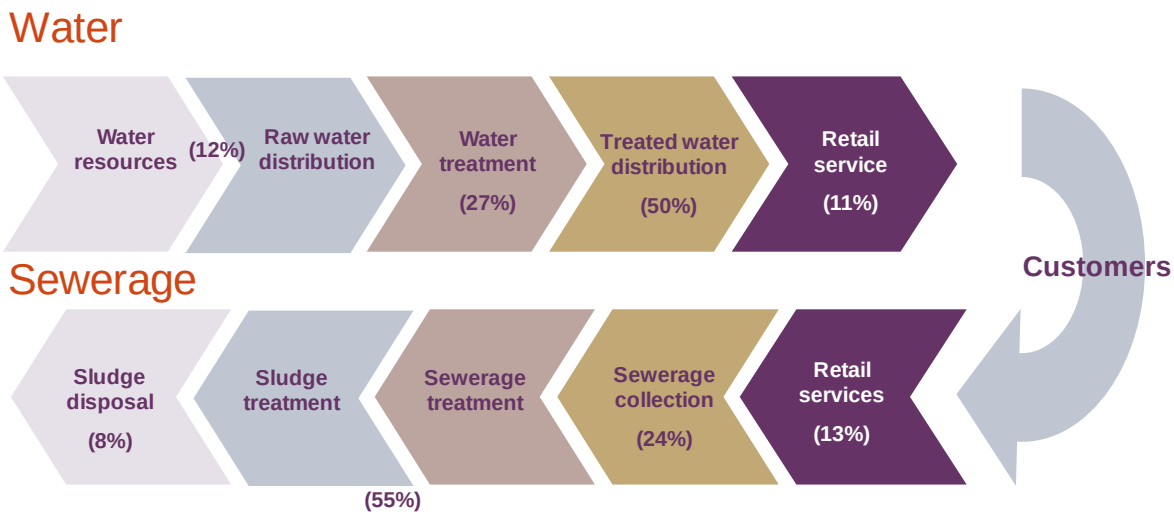
The value chain in the water sector in both England and Wales, and in Scotland, is vertically integrated, but there are differences in the form of integration between both regions with respect to the integration of sewerage operations. In England and Wales, there are 21 water companies[147], comprised of water and sewerage companies (WASCs), and water-only companies (WOCs) to which sewerage services are provided by the former group. These companies were privatised in 1989. In Scotland, Scottish Water is a publicly owned WASC.

At present, water companies in England and Wales, and Scottish Water, are local monopolies, although Scottish Water is now facing competition in the retail market, and there is competition for 'inset appointments' in England and Wales. The future role of competition in the industry, and its implications for investment, are considered below.

The high-level value chain of the industry is summarised in Figure A3.7 below. Within the high-level parts, other activities may also be conducted, such as metering.

[147] Excluding Cholderton, a very small water-only company that supplies 2,000 people; it is not regulated in the same way as the other companies due to its size.

Figure A3.7 Water and sewerage value chain for England and Wales



Note: Percentages correspond to Ofwat's indicative industry cost allocations and sum to 100% in each industry
Source: Based on Ofwat (2008), 'Ofwat's review of competition in the water and sewerage industries: Part II', and Ofwat (2009), 'Accounting separation—consultation on June return reporting requirements 2009-10'.

Briefly, Figure A3.7 describes a value chain in which water is abstracted from underground and surface sources, is treated to meet quality standards, and is distributed and then retailed to consumers. The used water is subsequently collected and treated at sewage treatment works. Sludge is removed, and effluent treated.

Turning to long-term planning and regulatory structures, the Water Act 2003 introduced provisions that require water companies to produce water resource plans forecasting supply and demand over a 25-year horizon and setting out how they propose to meet customers' needs. Although led by the companies and the Environment Agency, input into these plans is received from Ofwat and Defra.

Standards relating to environmental outputs and water quality are determined by the government through Defra upon the advice of the Environment Agency and the Drinking Water Inspectorate. Ofwat, as economic regulator, oversees the delivery of these outputs at the most efficient possible cost. This is sometimes referred to as the 'quadripartite process'.

A3.3.2 Post-privatisation landscape and subsequent changes

Two main changes with implications for interfaces have affected the post-privatisation landscape in the England and Wales water sector. The first is the reduction in the number of companies as a result of mergers, and the regulatory response to these mergers. The second is the emergence of outsourcing arrangements within the overall vertically integrated structure of the industry.

The merger regime for water in England and Wales reflects the fact that most companies are regional monopolies and, accordingly, there is little scope for competition. Instead, Ofwat relies on comparative competition in setting efficiency targets in price controls. A central element of this regime is the availability of sufficient comparators—ie, the existence of enough independent companies whose performance can be compared.

Legislation therefore requires mergers above a threshold of £10m to be automatically referred to the UK Competition Commission (CC), which can be contrasted to the normal test for mergers in most of the rest of the economy, which have to establish whether they may give rise to a 'significant lessening in competition'. Most mergers in the sector have been

prohibited or been subject to significant remedies.[148] Nevertheless, since privatisation, the number of WOCs has fallen from 29 to 13, while the number of WASCs has stayed the same at 10.

The merger regime therefore gives rise to questions of whether preserving the horizontal split and horizontal interfaces of the industry causes, protects or diminishes overall economic welfare. Some of these effects are quantified below.

The second set of changes relates to the way in which different companies have, in effect, chosen to increase the number of interfaces with which they must interact, by entering into different types of outsourcing agreements, while still operating within a vertically integrated structure. The most notable example in the water sector of comprehensive outsourcing is Glas Cymru, which announced in 2001 its intention to outsource the bulk of its day-to-day operations and its capital maintenance programme. In essence, its proposals involved splitting Welsh Water's assets from its operations, with the asset company becoming a 100% debt-financed asset CLG, and operations and maintenance contracts being put out to competitive tender. Other water companies have outsourced other elements of their activities, such as the management of asset delivery, although none have gone as far as Welsh Water. Indeed, Welsh Water has recently elected to take some services back in-house. This development, and the regulatory response to the 2001 change in Welsh Water's structure, is considered below.

A3.3.3 Similarities with and differences to GB rail value chain and industry structure

The following considers some differences between the water and rail sectors.

– **Homogeneous product**. Water is a homogeneous product, and its supply around the network is subject to less customer desire for differentiation than is seen in the rail industry.

– **Private ownership without subsidy**. The physical assets that make up the infrastructure of the water network in England are privately owned (and, in the case of Welsh Water, by a CLG) and do not require subsidy.

Some of the main similarities include the following.

– **Safety**. Water undertakings have tightly controlled responsibilities in relation to the safety of water provided for human consumption.

– **Physical network**. A physical network is used to link water undertakings with their users.

– **Output orientation**. There are high-level similarities between the HLOS process, in which the government specifies outputs that will be produced by the rail sector, and the role played by the quadripartite process in specifying the outputs to be produced by the water industry companies.

– **CLG network provider**: a CLG network exists in the form of Welsh Water.

A3.3.4 Evidence on the operation of interfaces, alignment of incentives, and future debates about interfaces

This section considers some of the evidence on the cost of changes in interfaces in the water industry, and, more significantly, some of the debates surrounding, and analysis that has been conducted in relation to, future increases in the nature and number of interfaces in England and Wales.

[148] With the exception of the 2002 Vivendi case.

Merger regime and value of comparators

Evidence on the costs involved in the merger regime were brought into focus by the merger between two WOCs (Mid Kent Water Limited, MKW, and South East Water Limited, SEW)—which was referred to the CC in 2006. The CC considered whether the merger would prejudice (or be likely to prejudice) Ofwat's ability to make comparisons for performance assessment. Against this were a number of possible benefits to consumers of the merger, including operating cost reductions, CAPEX reductions, and improvements in the management of water resources.

The CC concluded that the merger should be allowed on the grounds that the benefits to customers outweighed any detriment arising from the loss of a comparator, which was found to be of limited impact. The benefits were estimated as being in the region of £4m per year arising from cost reductions and lower financing costs. Water resource benefits were recognised, but not quantified.

The 'Cave' review of competition and innovation in the water sector, discussed in more detail below, recommends that the turnover limit for mergers be increased from £10m to £70m, with mergers above this being subject to OFT stage one merger assessment.

Welsh Water: vertical integration, contracting out and CLG status

Welsh Water's proposals for outsourcing raised concern among regulators, particularly since this had the potential to lead to a weakening of responsibility for quality and service delivery, with negative consequences for users. In response, Ofwat proposed licence conditions to clarity responsibilities that were, in effect, intended to prevent Welsh Water from operating as a procurement agency, but would instead have ultimate responsibility for any failures on the part of the contractors.

Ofwat's intervention also highlighted the critical role of coordination among different interfaces: if the process of regulation specifies objectives, and if contractors deliver the services necessary to meet these objectives, it is still necessary for an entity to be responsible for planning and coordination this delivery.

The main provisions included the following obligations on Welsh Water with respect to coordination:

– it was obliged to submit a detailed procurement plan to Ofwat;

– it had to demonstrate that it had 'sufficient methods of planning and control' to 'discharge its financial and management functions';

– it was prohibited from making any arrangement to transfer or delegate any of its functions, as a statutory undertaker, to a contractor.

To address these concerns, Welsh Water committed to retain 200 key technical staff within the licensed entity to ensure that it had 'sufficient methods of planning and control'. This gives some indication of the transaction costs involved in performing the coordination function.

A significant recent development in this context was the announcement of a corporate restructuring on February 9th 2010 to the effect that the outsourcing contract with United Utilities was not to be renewed, and that all employees were to be transferred back to Glas Cymru by April 2010. This move was driven by the need to meet the PR09 efficiency target set by Ofwat, which required savings of 20% over the period. The restructuring will reduce employment by 300 staff.[149]

[149] Welsh Water (2010), 'Welsh Water to restructure to meet efficiency challenge', press release, February 9th, available at http://www.dwrcymru.com/English/News/displayNews.asp?ID=1615

It remains to be seen whether this restructuring can be effective in the context of the organisation's CLG status. Its efficiency performance has generally been no better than average among the WASCs, a performance that is linked in a 2006 report to its CLG status.[150]

Future changes: increased separation?

There are ongoing debates about the future of competition in the water industry in England and Wales. For example, Ofwat has announced nine work programmes, including a number that relate to initiatives intended to support competition in different parts of the value chain. Separation measures are being introduced in England and Wales to avoid unjustified price and non-price discrimination.[151] The first steps towards accounting separation have already been made, but legal separation is still under discussion and will require legislation. Additional separation measures will depend on the extent to which competition evolves in the market.

In particular, some of these Ofwat workstreams build on the recommendations of the Cave review, [152] while particular forms of separation are already being introduced.

Box A3.1 The 'Cave' review

In early 2008, the Chancellor of the Exchequer, the Secretary of State for Environment, Food and Rural Affairs, and the Welsh Minister for Environment, Sustainability and Housing, commissioned Professor Martin Cave to undertake a review of competition and innovation in the water industry, with the objective of recommending changes to its legislative and regulatory frameworks.

The review outlines a variety of ways in which competition in the England and Wales water sector might be extended in the future, with the use of market-based mechanisms being recommended (as shown in the figure). In particular, Cave has recommended greater upstream competition as a way to stimulate innovation.

Abstraction and discharge	**Current situation**: licences awarded on a first-come, first-served basis; therefore, water resources may not be used efficiently **Cave review**: licences should be fully tradable, subject to modification for direct environmental impacts
Upstream	**Current situation**: provision of water and wastewater treatment dominated by the local vertically integrated incumbent, and there are limited incentives for innovation **Cave review**: upstream competition to encourage innovation, with decision on when and how to extend competition taken by the UK government
Retail	**Current situation**: only customers using above 50 megalitres p.a. can choose their retailer; therefore, the majority of networked customers are supplied by their local incumbent **Cave review**: initial findings, which have been accepted by the UK government, recommended reducing the threshold to five megalitres p.a. Recommendations in the final report were to remove the threshold for non-household customers

Source: Oxera, based on Cave (2009).

The review does not foresee the introduction of widespread competition in the market in water (through the creation of a mandatory pool or bilateral contracts) for some time—at least not for the next 10–15 years. The CBA undertaken in relation to these reforms does not, at present, indicate that

[150] NERA (2006), 'Corporate form, financial guarantees, and efficiency performance: Expectations and evidence', report prepared for the ORR, December 18th.
[151] Ofwat (2008), 'Ofwat's Review of Competition in the Water and Sewerage Industries: Part II', May, p. 42.
[152] Cave, M. (2009), 'Independent Review of Competition and Innovation in Water Markets: Final Report', April.

wider competition in the market would be beneficial to society.

As an initial step, Cave recommended reforming the water supply licensing regime to ensure that it works efficiently and is supplemented with a framework of economic purchasing. As a result of current uncertainties, it was recommended that the decision on when and how to extend competition be taken by the UK government at a later stage.

The CBA of the Cave review work is summarised in Table A3.2.

Table A3.2 Net monetised benefits of water competition options (NPV £ billion)

Recommendation	NPV of review recommendations over 30 years (£m)
Purchasing obligation	1,300
Reform to water supply licensing regime	400
Inset modernisation	300
Retail competition	600

Source: Cave, M. (2009), Table 1A.

In addition to these benefits, the review suggests that there could be significant non-monetised benefits, including improved environmental outcomes and higher service and quality levels. It is notable that the overall benefits from all the models of change and competition proposed is their scale, amounting to over £2.5 billion of NPV benefits. This is relevant in light of the proposals to increase separation and the number of interfaces.

Cave's review identified a number of misalignments in the industry value chain that might be addressed by some of the reforms proposed in the study. These include the following.

– A misalignment of incentives between different elements of the value chain with respect to innovation, leading to misdirected research and development and weak incentives for water companies to innovate.

– A misalignment of incentives such that water companies have a potential bias for solutions that require CAPEX rather than OPEX (in essence, a 'gold-plating' argument).

– The special merger regime, discussed above, is seen as weakening management incentives to secure the best use of scarce water resources.

– In areas where restrictions are placed on the issue of new licences, there are no financial incentives to move the point of abstraction, or to transfer water, from areas where water is abundant to where it is scarce.

– Year-round abstraction licences offer no incentive to abstract water during different times of the year, and therefore do not offer incentives to abstract water at optimal times.

– Abstractors may need to face stronger incentives to reduce their propensity to over consume water.

Another feature of the analysis is that increased competition may lead to a higher cost of capital, which has relevance for the certain changes that might be introduced as a consequence of the VfM Review in rail. The Cave review provides indicative estimates to suggest that the cost of capital might increase by at least 100bp or more, following the introduction of competition.[153] This might be explained by competitive reforms placing

[153] Ibid., p. 65.

elements of the business outside the scope of regulation, thereby leading to an increase in non-diversifiable risks borne by investors in the competitive part.[154]

Analysis for Ofwat by Nourse[155] estimates that uncertainty over the eventual path of the competition reforms could increase the cost of borrowing by as much as 10bp to 30bp.

A3.4 Messages for GB rail

This section identifies some of the key transferable lessons that emerge from a review of these three sectors, and which have been used to inform the assessment of options presented in section 4 of the main body of the report.

- Separation and radical reforms of industry structure, potentially involving significant increases in interfaces, can have positive outcomes for users, but the detail depends upon the circumstances and wider market design considerations. (This is the experience from the electricity and gas industries, and reflects the conclusions of the Cave review in water.)
    - This suggests that, given a workable market structure, incentives matter more than transaction costs when it comes to long-term outcomes.
- A system operator can be established that reflects operational considerations, and allows for some areas with system-wide implications to be retained by networks. (This is the experience in the electricity industry.)
- Vertical integration poses problems for competition. (This is the experience in the electricity, gas, and water industries.)
- Coordination of functions is an important aspect of a utility's operation, and its implications in relation to separation or other forms of structural change requires careful analysis. (This is the experience in the water industry with Welsh Water, and with the design of the overall market structure in electricity and gas.)
- Users having direct financial interests in infrastructure spending by the regulated organisation is not sufficient, in itself, to secure optimal outcomes. (This is the experience in the aviation industry.)
- There is value in comparative competition and therefore potentially net benefits from horizontal separation. (This is the experience in the gas and water industries.) However, the ways in which this information can be used are subject to debate and change. (This is the experience in the water industry.)
- Contestability pressures on companies with market power can reduce cost, but there is a need to ensure that there are no negative consequences for users and for operations of networks more generally. (This is the experience in the electricity and aviation industries.)
- Consistency between government policy and regulatory policy is important for the outcomes for end-users. (This is the experience in the energy and aviation industries.)

A general finding from reviewing these sectoral experiences is that a track record of profitability before privatisation appears to have supported outcomes in what was often a radically different post-privatisation industry landscape.

[154] Nourse, R. (2009), 'Competition proposals and financing issues: a report for Ofwat', January, p. 2.
[155] Ibid., p. 6.

Park Central
40/41 Park End Street
Oxford OX1 1JD
United Kingdom

Tel: +44 (0) 1865 253 000
Fax: +44 (0) 1865 251 172

Stephanie Square Centre
Avenue Louise 65, Box 11
1050 Brussels
Belgium

Tel: +32 (0) 2 535 7878
Fax: +32 (0) 2 535 7770

Thavies Inn House
7th Floor
3/4 Holborn Circus
London EC1N 2HA
United Kingdom

Tel: +44 (0) 20 7822 2650
Fax: +44 (0) 20 7822 2651

www.oxera.com