IMP-RCM Stage 1 Report

RSSB support for the Crossrail / NR Great Western track condition data interchange project

Pete Johnson, Vivacity Rail Consulting Ltd

A Executive Summary

Vivacity Rail Consulting (VRC) has been commissioned by RSSB to consider the use of the toolkit from research project T1010 (Remote condition monitoring Phase 2) to support a project being undertaken by Crossrail and Network Rail Western Route. This will share track condition data gathered by unattended track geometry measurement (UTGM) equipment mounted on some Crossrail Class 345 trains as they run across Western Route infrastructure. The stakeholders have recognised the potential value of the T1010 approach to provide a structure to the project. It can be used as a platform for future expansion to other consumers and providers of data, and to other types of RCM such as overhead line equipment monitoring.

VRC has assessed how the T1010 toolkit can be applied to the project. We have investigated the use of:

- the Business Case Tool (from T1010-04) to clarify where costs and benefits fall
- the process map and template contract (from T1010-02) as a basis for agreement between Crossrail as data suppliers, NR Western Route as data consumers and RSSB as project facilitators
- the data architecture (from T1010-01) as a guide to the technical solution and standards to be adopted.

VRC has put together terms of reference for Stage 2 of this work, which will:

- Assist the stakeholders to use the T1010 approach to scope and deliver the project.
- Assist the stakeholders in building a prototype data broker to implement the T1010 concept of a data bus with systems connecting to it using data adapters.
- Augment and correct the T1010 documentation, based on experience gained in applying it.
- Demonstrate the T1010 approach and the data broker to the rail industry to show the benefit of using it.

An outline programme for the work suggests a duration of 6-7 months for the prototype phase.

Conclusions of the work are:

- The toolkit can be applied almost in its entirety and will provide considerable value to the project, though there are some gaps. Some of these can be addressed by additional documentation to be worked on in Stage 2 of RSSB's programme; others will need amendment to aspects of the toolkit. The main gaps are:
- The business case tool uses an outdated concept of what IT solutions to RCM data sharing tasks would look like.

- The template contract does not consider the role of intermediary data brokers that may be provided by 3rd parties.
- The data architecture has a somewhat outdated view of how best to implement the data bus. The principle has gained currency since the document was written but the techniques available have improved.

- More work is needed to identify the key use cases for the transferred data and thus the most useful focus for the data broker.
- The Stage 2 work can be done in 6-7 months but has dependencies on the completion of testing of the UTGM equipment and its integration with the trains and their IT systems.
- There are some elements of the data broker solution that are best managed outside the immediate project if the long-term benefits of the approach are to be realised.

Recommendations are:

- The data interchange project should set up a formal sponsor and management board to organise and direct the many strands of activity necessary to set it up.
- Work should start very soon to clarify the role of RSSB in the project and the extent to which it will contribute funding for it.
- The scope of the Stage 2 work should include effort to facilitate and speed up identifying the best use cases for the transferred data at NR Western Route.
- Several other data integration initiatives will benefit from knowing about the current work and vice versa. There should be information-sharing and liaison with them.
- RSSB should consider updating the T1010 toolkit to pick up those gaps that cannot be addressed with user documentation.

B Introduction and background

This document is the output of Stage 1 of the T1010 extension project IMP-RCM, prepared by Vivacity Rail Consulting in response to RSSB's request. It comprises:

- A description of the pilot project on which the work is to be based: the project considers the use of information from unattended track geometry measuring equipment fitted to Crossrail trains (Section C).
- Findings of work done so far in Stage 1 which serve to establish the scope and requirements for Stage 2 of this work, including how well the T1010 principles and toolkit fit to the pilot project (Section D).
- Conclusions and Recommendations: suggestions for ways to structure and implement the project and to facilitate cross-industry RCM data sharing, based on the work of Stage 1 (Section E).

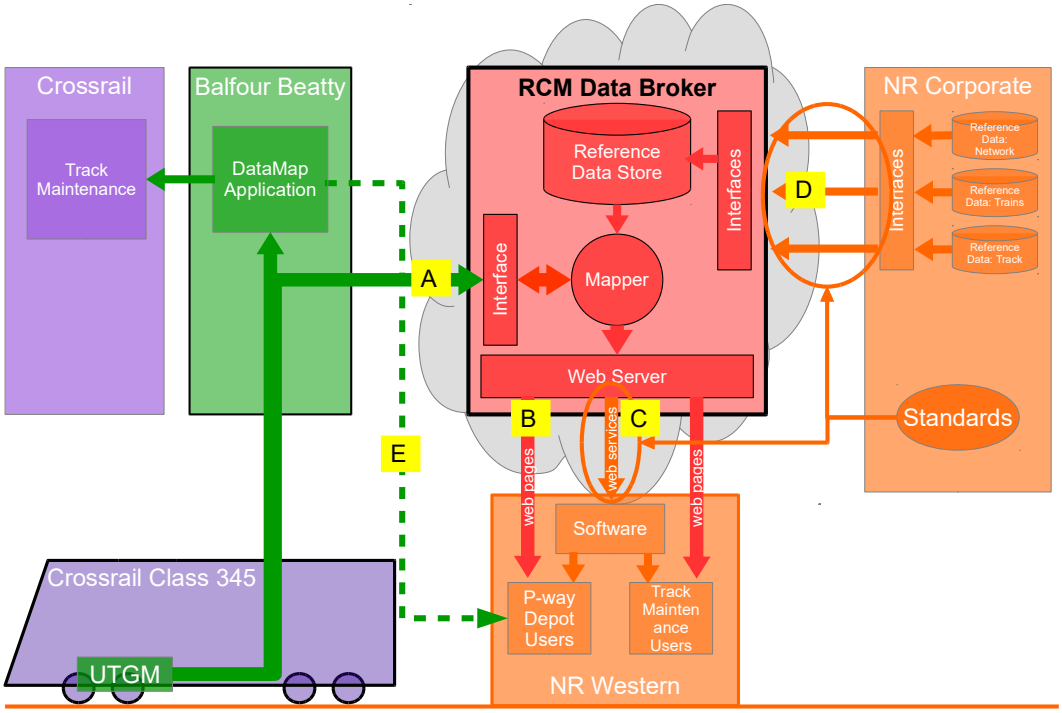
A separate document contains a proposed specification for Stage 2 of this work, with these elements:

- Specification of the main tasks to support the data interchange project
- Specification of the T1010 documentation update to be supplied

C The Data Interchange Project

Figure 1 shows the data interchange being considered as the pilot project, with the proposed role of a Data Broker corresponding to T1010 principles.

Figure 1: **The Data Interchange Project**



The project stakeholders (Network Rail Western Route and Crossrail) could have achieved their goal of sharing the UTGM data from the fitted Class 345 trains simply by allowing NR Western users direct access to Balfour Beatty's DataMap visualisation and analysis application. This is shown as data flow E in Figure 1. While feasible, this approach would not offer the benefits of standardisation, extensibility, and further data sharing that a T1010-based approach would bring. The stakeholders have therefore decided to try to implement data sharing using these principles.

The proposed solution uses a shared data broker to act as an intermediary between the data provider (Crossrail) and the consumer (NR Western). The broker will initially map the data provided by the UTGM equipment into a standard format and make it available to NR in a number of standard ways.

The use of the data broker will generate benefits for both parties:

- For Crossrail, it enables the extension of data sharing arrangements to the other NR route – Anglia – over which it runs trains without having to make any further change to their data-gathering process or analysis software.

- For Network Rail Western, it sets up a standard mechanism that can be extended in future to include data from the new Class 800 Hitachi trains that will also be fitted with unattended track geometry measurement equipment. The same mechanism can also be used to support data on other assets such as OLE.

The approach will also enable other similar initiatives elsewhere to share the same broker infrastructure. Services set up as part of this work (for example, a possible location mapping service) will be able to be re-used, lending consistency of approach and lowering barriers to entry.

The straightforward data flow marked E in Figure 1 will be replaced by a number of data flows, indicated by A-D in the figure. These are:

- A: data from the on-board equipment is picked up by the data broker
- B: the data broker maps and translates the data into a standard format and makes it available to NR Western staff via a simple web interface.
- C: the data broker makes the data available to software used by NR Western staff in a suitable format, accessed by a web service interface.
- D: reference data and lookups are taken from Network Rail systems, as needed, to do the mappings necessary.

Since this is proposed as a pilot or demonstrator system, the data flows in B and C will be the simplest and quickest to deliver, consistent with them providing some value to NR Western. The capabilities of these flows can be enhanced in future work, building on the basis established here.

D Findings of Stage 1 work

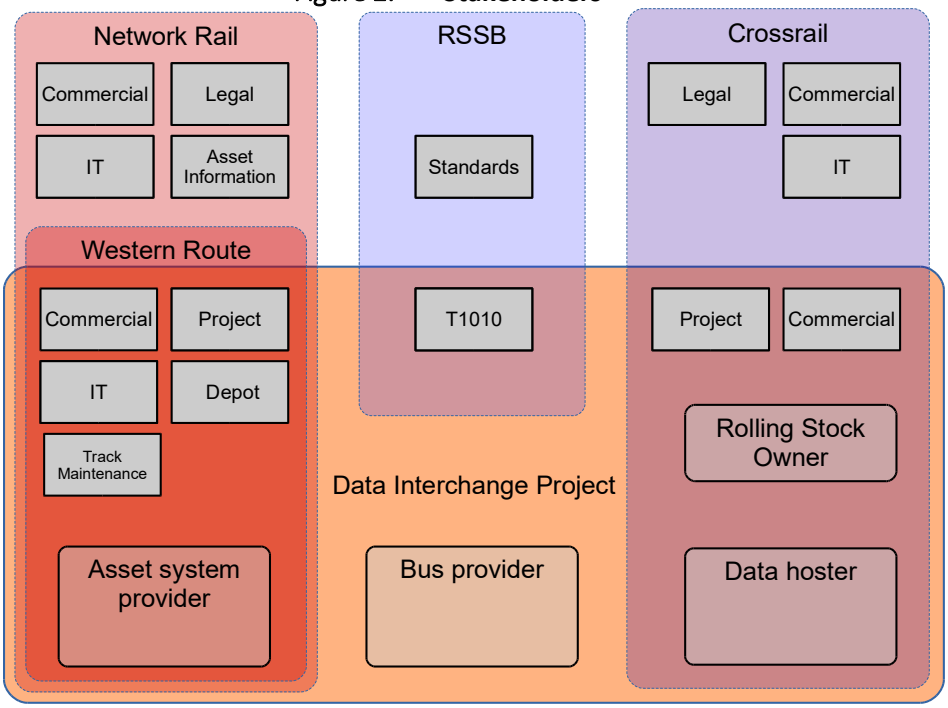
This section describes the work done during Stage 1 and lists the points of note that the work has revealed. The findings are grouped into these categories:

- Context: stakeholders, interested parties and the technical / commercial boundaries across which agreements need to be made (Section D.1)
- Application of the T1010 toolkit: which elements of the toolkit apply and how well do they match the requirement (Section D.2)
- Use cases identified for the track condition data and what this means in terms of the data flows and data broker (Section 0)
- The data that will be included in the data flows (Section D.4)
- The business case from the point of view of the main participants and what this means for the commercial arrangement between them (Section D.5).

D.1 Context and main stakeholders

Figure 2 shows the context of the data interchange project, identifying the stakeholders. The organisations concerned are Network Rail (Western Route as well as central bodies), Crossrail and RSSB. The project will also involve a supplier of the data bus (= data broker).

Figure 2: **Stakeholders**



Within the data-supplying and data-consuming organisations (Crossrail and Network Rail respectively), there are units representing commercial, legal and IT interests as well as project sponsors and the actual users of the data (depot and track maintenance).

Also having an interest are the owner of the rolling stock (Bombardier) and the gatherer of initial data (data hoster) – in this case Balfour Beatty who provide the unattended geometry measurement equipment.

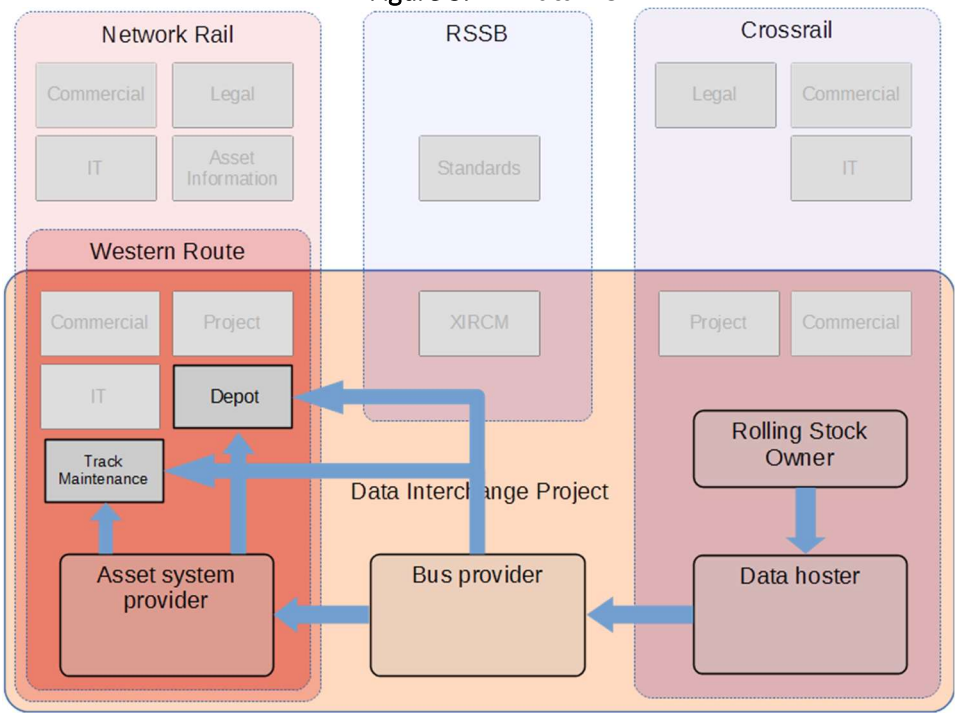
D.1.1 Data flow

Figure 3 shows the basic flow of track condition data and the parties involved in it.

Data flows from the rolling stock, transferred by the owner, via a data repository to the data bus and thence to the consumers.

The consumers of the data are the depot and track maintenance (= strategic) functions at Western Route, who can make use of the data directly from the bus provider (= data broker) or indirectly, via asset data systems that they already use (such as LADS).

Figure 3: **Data Flow**



D.1.2 Commercial concerns

Figure 4 shows the commercial relationships between stakeholders and between the central and project-specific commercial interest groups within stakeholders.

The key relationship is at the project level, between the Crossrail and Network Rail Western Route commercial teams. This is informed by the T1010 template contract and commercial principles.

There is also a commercial relationship with the provider of the data bus (or data broker). The exact form of this relationship is not yet determined, but at least initially it will be between RSSB and the supplier of Stage 2 of this work.

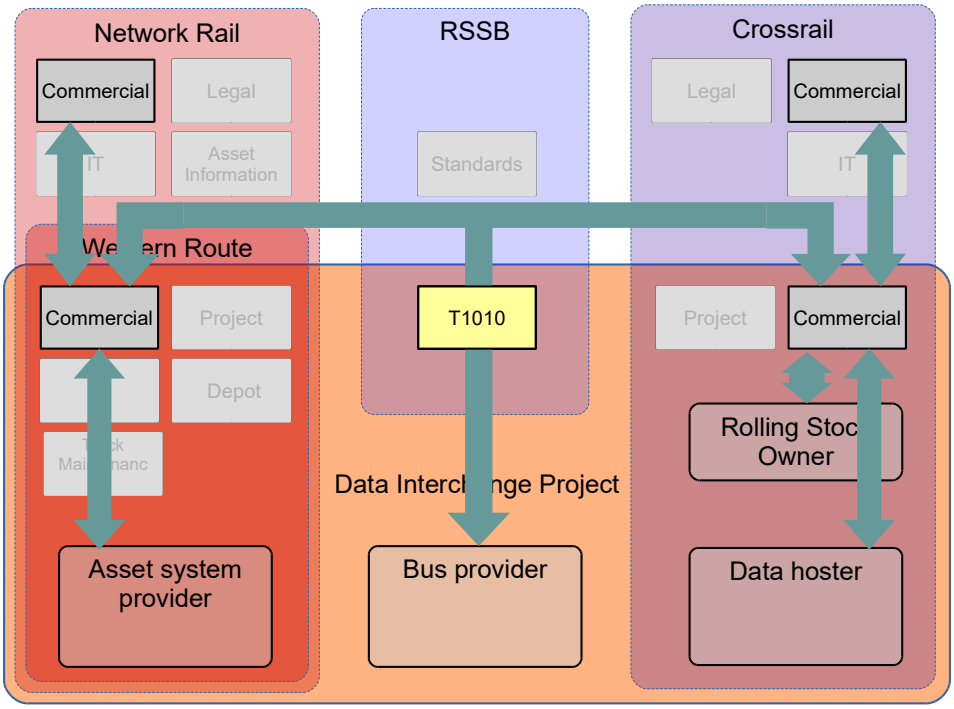


Figure 4: **Commercial concerns**

The project teams will need to respect commercial principles and guidance set out by organisation-wide commercial managers.

D.1.3 Standards and references

Figure 5 shows how reference data and common standards will be applied to the data flow and the systems which generate, mediate, and consume it.

The main source of standards is the T1010 data architecture, which informs the IT organisations of the data provider and consumer as well as the data bus provider.

Also important is the use of reference data and standard techniques defined and managed by Network Rail's asset information function.

The data bus provider will need to conform to IT standards – in such as security, and data formats – established by the IT organisations of Crossrail and Network Rail Western Route.

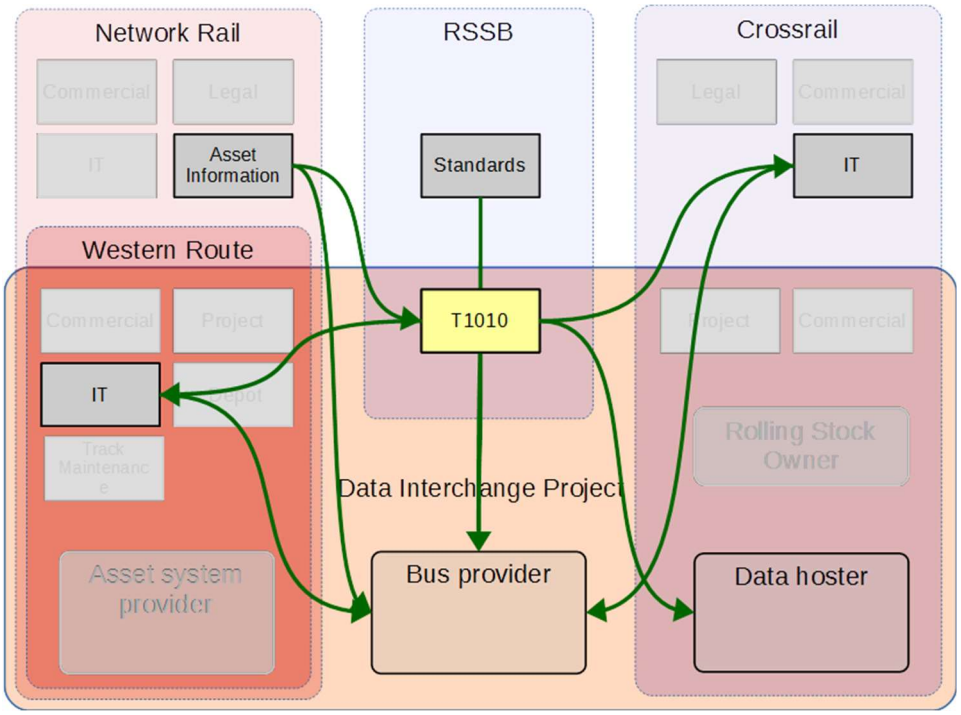


Figure 5: **Standards and References**

D.1.4 Additional stakeholders and interested parties

Beyond the parties directly concerned with the data interchange project, there are others with an interest in it. Among these are:

- Network Rail information architects who are keen to see how the services-based approach proposed here aligns with their own information strategy.
- The Network Rail ORBIS team which is the main supplier of reference data to be used by the mapping process; and which may provide some of the mapping services.
- The team at Network Rail which is working on a tactical data hub as a means of making RCM data available for analysis – starting with OLE data captured by a single fitted train on the London North Western route and passed to a team at Coventry University for storage and analysis.

D.2 Application of the T1010 Toolkit

An initial application of the T1010 toolkit was carried out to help with scoping the Stage 2 work and to get an initial view of the potential value of the toolkit to the project. The three core elements of the toolkit – the business case tool, the template contract, and the data architecture – were reviewed and considered against the requirements and specific circumstances of the data interchange project. This application is described in Sections D.2.1, D.2.2 and D.2.3.

The overall process of applying the toolkit is summarised in Figure 6.

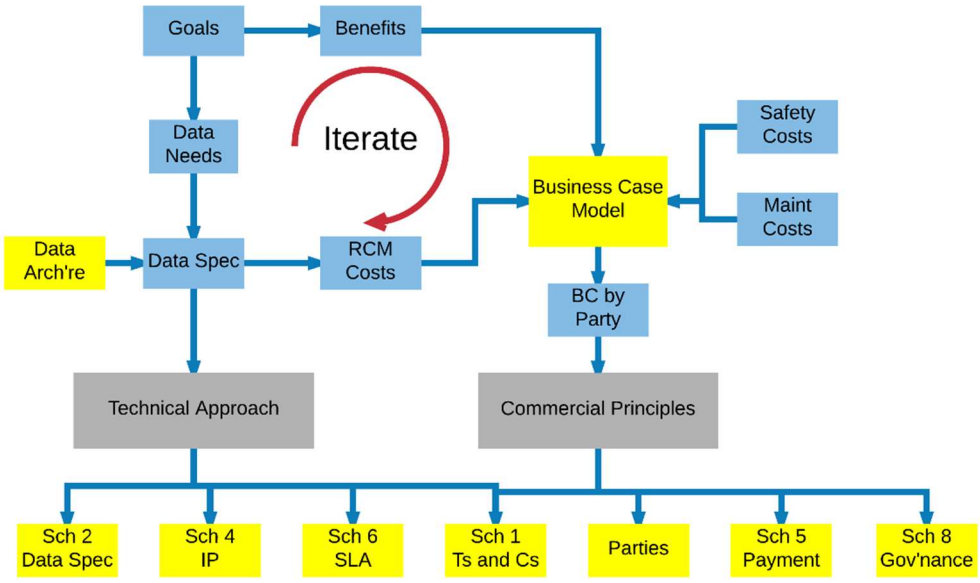


Figure 6: **Overall process map**

There are some areas of misalignment between the toolkit and the circumstances of the current project or others that may grow from it. These have been summarised in a gap analysis in Section D.2.4.

D.2.1 Business Case Tool

This tool, the output of T1010-04, is an Excel spreadsheet with a user interface based on a data entry wizard. Version 3.0, dated 12/08/2014, was used. Details of the proposed data interchange project were entered as estimates; also estimated were the likely costs and benefits of the proposed approach. The various data inputs and the assumptions underpinning them are listed in Table 1in the Appendix.

Given these assumptions above, the Business Case tool calculated the benefits for different levels of RCM data sharing shown in Table 2 in the Appendix.

The documentation for the business case tool is in two parts, reflecting the fact that it was put together in two phases – firstly as part of RSSB research project T857-06 'Overview of the benefits and risks associated with condition monitoring'; then updated in work package T1010-04. This makes the working of the tool somewhat opaque, so we had to spend time delving into the formulae and VBA code of the tool to work out exactly how it does its calculations and therefore how best to input the details of the project.

Also notable is the way in which the tool assumes that the IT elements of data integration projects will be delivered. The options here are described in terms of local IT systems, servers and direct point-to-point data interfaces. Using current best practice, it is much more likely that services will be cloud-hosted; data will be delivered via shared

web services rather than point-to-point interfaces; and charging will be based on data volumes or the number of transactions rather than consisting of capex and fixed by-period opex. There is thus a case for re-examining the description and the charging basis for the IT components of future schemes.

D.2.2 Process Map and Template Contract

These documents, created in project T1010-02, describe the process for assessing the business case and setting up the commercial framework for a data interchange project. We reviewed the template contract from T1010-02 to determine which of its elements were relevant to the current data interchange project and to identify any areas that might need project-specific work or extension to the contract or its guidance documentation. The results are shown Table 3 in the Appendix.

The template contract is a very good fit for the requirements of the current project. The current project also requires some useful extensions to the basic contract which may be more generally applicable in future contracts, particularly around the use of the data broker.

D.2.3 Data architecture

The T1010-01 data architecture was examined to see which elements of it would add value to the data interchange project and to identify:

- Areas where the documentation needs to be reinforced, expanded or clarified.
- Areas where centralised functions such as governance, standardisation and service hosting need to be considered.

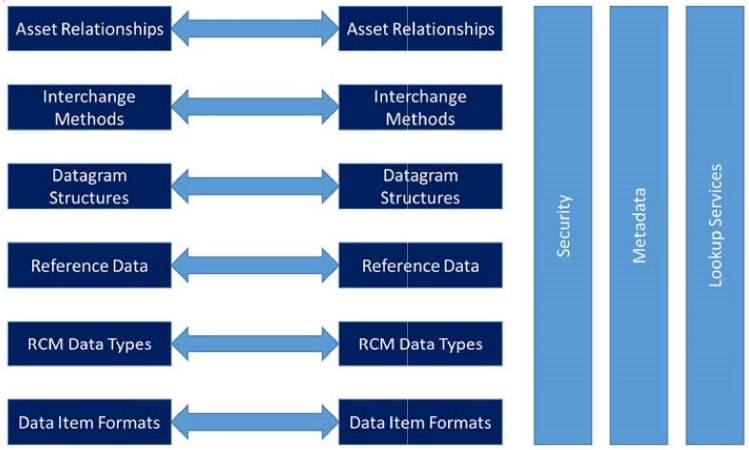


Figure 7: **Data integration levels**

Figure 7 is taken from the T1010-01 data architecture requirements document. It shows on the left the levels of integration proposed by the data architecture, together with the supporting functions that integration requires at any level. Any given data integration

project will require a certain level of integration, which then implies that the levels below it are also needed.

Figure 8: **Data architecture - integration levels and shared artefacts**

Figure 8 shows how the level of integration (along the top) implies the presence of shared capabilities in terms of data and IT infrastructure.

For the current integration project, Table 4 in the Appendix shows which elements of the data architecture must or could be applied.

The project gives an excellent opportunity to trial most of the important functions of the data architecture and thus improve the documentation and guidance for its broader use.

D.2.4 Gap analysis

Some areas of mismatch were found between the T1010 documentation and the requirements of the present project. These suggest gaps which may be addressed in the current project or by later changes to the T1010 documentation itself.

D.3 Use cases and constraints for interchanged data

In discussion with the engineers involved in the supply and use of the track condition data being interchanged through this project, we have identified the following use cases to which the data can add value and some constraints on its use.

D.3.1 Use cases

UGMS data can add value to a number of use cases related to the maintenance of track. The main ones are listed in Table 6 in the Appendix. The use cases identified help to drive the business case and also determine the data items required and their quality characteristics.

D.3.2 Constraints

The data from UTGM equipment are generally less precise and less reliable than those coming from properly-calibrated equipment on measurement trains. As such, they cannot presently be used to demonstrate compliance to mandatory track quality standards.

On the other hand, the UTGM data are gathered much more frequently (probably several times a day for each track section, as compared with once every several weeks for the measurement train data). Also, techniques are now becoming commonplace to merge the data from many readings, each of which may be imprecise or biased, to come up with composite measurements of much higher dependability.

For now, therefore, the UTGM data can only be used to support rather than replace the measurement train data. This limits the scope of benefits that can be achieved in the short run.

Looking further ahead, it is likely that work will be done with standards and regulatory authorities to change the compliance criteria for track quality inspection so that compliance can be demonstrated using lower-quality data but certified algorithms for combining them to come up with reliable and safe conclusions. This will reduce the cost of demonstrating acceptable track quality.

To support such work, the quality of metadata associated with UTGM data and the algorithms used to process them will be important. The T1010 toolkit contains mechanisms to define and record these metadata: these can be demonstrated in the current project.

D.4 Data

Balfour Beatty has supplied a provisional specification for the data file formats to be provided by the UTGM equipment fitted to 3 of the Class 345 Crossrail units. The files are in .CSV format and are transmitted from the trains in batches at a rate determined by the train manufacturer Bombardier which has control of the over-the-air channel.

There are several different .CSV file formats, each containing a different type of data. The relevant ones for this data interchange exercise are:

- location data: location events and event detail
- geometry data
- end of batch.

The content of the data feed is described in the following sections.

D.4.1 Location data

The UTGM equipment generates a record for each event for which a location is required. Information available depending on the event type are:

- Position in terms of tacho clicks. This would need to be calibrated and referenced to turn it into a track location.
- Position in terms of previous and next station or platform.
- Passing an AWS magnet or TPWS/ATP grid or balise.
- GPS position and signal quality.

D.4.2 Geometry data

The geometry data file consist of instantaneous readings of a number of track geometry parameters. Basic parameters which are directly usable are listed in Table 7 below; other parameters for which some processing or calculation would need to be done are listed in Table 8.

D.4.3 Batch end

The UTGM data are periodically batched up and transmitted to the wayside by functions within the Train Management System. Each batch has a control file which lists the files contained within it as well as some additional identification and time-stamps.

At the time of writing it is not known how big the batches are and how frequently they are transmitted.

D.4.4 Alerts

The technical specification for the Crossrail rolling stock requires that alerting thresholds can be set for track geometry parameters, and alerts transmitted as rapidly as possible if any of these are exceeded. At the time of writing it is not clear how these alerts are transmitted or whether it is possible to intercept them or pass them on via the data broker.

D.5 Commercial arrangement

D.5.1 Principles

The transfer of UTGM data from Crossrail to NR Western would take place whether the T1010 data broker concept was to be used or not. This therefore suggests the following basic principles for a commercial arrangement between Crossrail and NR Western:

- NR Western is the prime beneficiary of the data interchange so should bear the bulk of the cost of providing it.
- Crossrail do derive benefit from the improved track quality and reduction in delays expected from the data interchange. There are thus some aspects of the cost which Crossrail can be expected to bear.

- The use of the data broker approach generates some setup and ongoing support cost that would not arise if the alternative approach - direct access to the Crossrail DataMap system – were used instead. However, there are follow-on benefits to both Crossrail and Network Rail from the use of this approach, as additional data sources and users are added.
- RSSB also gain benefit on behalf of the industry at large from the use of the data broker approach – particularly in terms of having a demonstrator of the T1010 principles and an improved set of documentation (see below).
- Network Rail and RSSB should share the setup cost of the data interchange, at least to the demonstrator level.
- Network Rail should therefore bear 90%, say, of the ongoing cost of providing the basic data broker service. Crossrail should bear the remaining 10%, which would roughly cover the cost of any additional effort by its suppliers Bombardier and Balfour Beatty in making the data available.
- Further development of the data broker, for example to add data processing and data reference services or to build a richer web front end, would be funded by Network Rail.
- The Bata broker supplier would be encouraged to manage the broker in such a way that it supported the XIRCM goals of broader availability and ease of use of RCM data. This means maintaining the documentation, supporting the development of data adapters, and managing the architecture so that it abides by the principles set out in T1010-01.

D.5.2 Role of RSSB and XIRCM SG

RSSB and XIRCM SG, as representatives of the wider rail industry, have a role in using this project to trial the T1010 approach and maximise the benefits it can offer to future RCM data interchange projects. This role has several aspects:

- Trial of the elements of the T1010 toolkit identified in Section D.2 to verify that they are fit for purpose and to identify any gaps or changes necessary to them.
- Preparation of a technology demonstrator that can show wider industry the value of the T1010 approach.
- Preparation of further user documentation to complement the existing T1010 toolkit elements.

This role predominantly features in the early development and prototyping phases of the project; less so in the later production phase.

In recognition of the value of the project to RSSB, it is reasonable that RSSB should provide some of the funding to support the decision to use the T1010 approach on this project.

D.5.3 Services and centralised elements

To maximise the possible breadth and utility of future data sharing, some aspects of the data sharing architecture need to be managed centrally rather than left to individual projects to decide. These elements could include:

- The API specification.
- Usage notes, code samples, and best practice documentation.
- Reference data services such as location mappers.
- Access control, security, and data integrity layers.
- Shared data services such as location registration, aggregation, smoothing, and calibration services.
- Shared messaging facilities to support guaranteed delivery, publishing, subscription, and other enterprise integration patterns.
- Shared data models and ontologies to support cross-domain mapping and software-based reasoners.
- Shared bulk data storage.

There are options for how these shared elements can be managed and delivered:

- IT services may be hosted or provided by similar facilities operated by Network Rail.
- RSSB can host documentation and manage the code samples and best practice wiki on standard sites such as github.
- RSSB or other industry bodies can manage commercial firms to provide the services. To simplify the payment mechanisms, these may be cloud-hosted using standard micropayment mechanisms (per transaction or per volume of data) to fund themselves. This is a similar mechanism to that used by Network Rail to host the open data feeds. The role of the supplier would be similar to that adopted for example by TransportAPI in the provision of transport schedule data: acting as an integrator of many differently-formatted inbound data feeds and offering them as outputs in standard formats, charging per hit.

E Conclusions and recommendations

In carrying out this Stage 1 work, we have drawn some conclusions about the use of T1010 toolkit for the pilot project and identified some recommendations for the way the project is managed in Stage 2 and beyond.

E.1 Conclusions

E.1.1 Applicability of the T1010 toolkit to this project

All the key elements of the T1010 toolkit – the Business Case tool, the Process Map and Template Contract, and the Data Architecture – have a useful role to play in the project. The scope of the project is such that nearly all the key elements of the architecture can be used. This means that the project represents a good opportunity both to road-test the architecture and to apply experience from using it to improving it.

E.1.2 Gaps and weaknesses in the T1010 toolkit

Our gap analysis of the T1010 toolkit documentation revealed some gaps, weaknesses and areas that would benefit from an update. The most significant of these are:

- The business case tool is somewhat awkward to use for trains monitoring track, when the costs associated with service delay and cancellation need to be apportioned between train operator and infrastructure manager. It also has a somewhat outdated view of how data interchange projects would be delivered which complicates the calculation of RCM system costs.
- The template contract is specifically designed around a direct data interchange between provider and consumer. It does not recognise what is likely to be a common model: the use of a 3rd party to provide a broker service.
- The data architecture has a rigid view of how a data bus might be implemented that does not accord with current best practice. The recent development of cloud-based, service-oriented, API-driven applications has made this type of approach more common and deliverable without the use of a formal data bus.
- The data architecture in particular gives no guidance as to how its standards and principles should be applied to a specific data sharing project and how they should be reflected in the schedules of the agreement between such a project's stakeholders.

E.1.3 Business case for the data interchange

There is common agreement that there will be business benefits to the broader sharing of RCM data along the lines of this project. However, it is not clear exactly how they will be realised, what other conditions will need to be satisfied for the benefits to be gained,

or when they might accrue. Potential small benefits may follow from a reduced frequency of in-service track faults and from faster confirmation of fix and lifting of TSRs andESRs. The most significant benefits will come from changes to maintenance regimes that reduce the frequency and cost of inspections.

This uncertainty means that it is not clear which aspects of the proposed data interchange should be prioritised in Stage 2. Support for the elucidation of this requirement is included in the scope of RSSB's supplier's Stage 2 work, but because of the uncertainty of the duration of this task, it has been expressed as an expected number of supplier person-days input rather than as a task for which the supplier should quote.

E.1.4 Project readiness and timing

Whilst the project stakeholders are clear that they wish to proceed with the data interchange, there has not yet been any formal process to move to an agreement between them. This process could potentially be time-consuming once all the considerations mentioned in the T1010-02 template contract become known. These cover such matters as ownership of data and IP in additional processing (which could involve Balfour Beatty as the UTGM supplier and other parts of Network Rail as data service providers), responsibility for data provision and completeness (which could involve Bombardier as train and over-the-air channel supplier), equipment and software necessary to deliver the data, security, availability, safety of the data (which could involve the data bus supplier), the role of RSSB and its Stage 2 supplier in time-critical aspects of the project timeline. Since the duration of this phase is not known, the scope of RSSB's Stage 2 supplier input to the process has been stated as an anticipated number of supplier person-days rather than as a task whose effort the supplier should estimate.

The stakeholders have an aspiration to have this data interchange in live operation ready for the introduction of Crossrail services on the Western route in May 2018. Whilst there is time to define and build the interface according to our estimated programme shown in the Stage 2 Specification document given a prompt start, there are key external dependencies whose timing is not yet known that may introduce delays to this timeline:

- The format of data files coming from the Balfour Beatty UTGM equipment will not be finalised till the completion of the UTGM Factory Acceptance Testing. The date of this is not presently known.
- The actual flow of UTGM data will not start till the equipment is mounted on the trains and integrated with the Bombardier train management system and over-the-air link. The date for this integration is not yet known.

E.1.5 Centralised elements

The choice of the T1010 data bus approach means that there are shared elements of the solution independent of the data provider and consumer. It is presently not clear how these elements will be managed. While it is reasonable for RSSB's Stage 2 supplier to manage them for the prototype stage, there is no arrangement in place for the more significant production phase in which it is more important that the data flow is reliable and the solution well-founded and extensible.

E.2 Recommendations

From this work and the conclusions above, we can make some recommendations for the conduct of the project. These are directed at RSSB as the procurer of the Stage 2 supplier's input; they also apply to the project's main stakeholders.

E.2.1 Project board and sponsor

The number of stakeholders, the complexity of the relationships and the amount of discussion that needs to take place to achieve an agreement suggests the need for a Project Board and Project Sponsor to organise and direct activities.

E.2.2 Clarification of RSSB role and contribution

The principles set out in this document (Section D.5.1) should be used as the starting point for a discussion and agreement on the role that RSSB should play in the project and the extent to which it should contribute to funding it. This role should extend as far as successful deployment of the demonstrator data broker. The role of RSSB beyond that is to be determined.

E.2.3 Use cases and value of data

Work should start immediately on identifying the most productive use cases for the interchanged data in the short run and therefore the focus of the prototype outbound interfaces from the data broker.

E.2.4 Liaison with other data integration initiatives

There is significant interest across the industry in the T1010 work, as it has potential applicability in other data integration initiatives. Of particular note are:

- Network Rail's 'actical Data Hub work, based initially on data captured by unattended train-borne OLE monitoring equipment running on the London North Western Route.
- Network Rail's ORBIS programme, specifically the efforts within it to offer data services and to work towards an ontology-based model of the rail network.

The project should ensure that its work aligns with these initiatives and information is shared with them.

The In2SMART project is part of the Shift2Rail programme funded by the EU. Network Rail is involved in this project. The goals of at least two of the work packages in this project overlap considerably with the T1010 work. There is thus much to be gained by ensuring that In2SMART is aware of the work already done for T1010.

E.2.5 Update of the T1010 toolkit

Many of the gaps identified by the Gap Analysis in Section D.2.4 can be addressed by improvement in the guidance documentation as proposed as part of the Stage 2 scope. However, some cannot and should be considered for an update to the toolkit. This should be done once the key lessons from the current project have been learnt and documented.

Appendix

This appendix contains tables generated by the investigation and mapping exerciase carried out in Stage 1 of the work.

Table 1 Business Case Tool Inputs and Assumptions

Input	Assumptions	Notes
Implementation Timescales	2 years to see benefit of State Detection implementation; further 2 years or 3 years for each higher level of integration	These are conservative for the lower levels of integration. The higher levels will depend on process change external to the project so are harder to predict
Asset information – NR Western	60 miles of track (= 2 slow lines from Paddington to Reading)	
Asset risk - Crossrail	20 incidents per year, each causing 100 minutes of delay. Each minute's delay costing £25. 1% of services cancelled, each cancellation costing £10,000.	Costs associated with delays and cancellations caused by track related incidents impacting on Crossrail trains. Figures notional – need benchmarking against actual delays and cancellations on Great Western trains.
Asset risk- NR Western	50 incidents per year, each costing £50 per minute for 100 minutes of delay.	Impact is estimated Schedule 8 cost of delays associated with track.
RCM system costs – UTGM on trains	Failure rate 1 per year, cost £2000 to fix Additional costs for NR interface £15k per annum, borne by NR.	This is the cost that would be incurred by NR whatever solution was employed.
RCM system costs- Data Broker	£70k setup cost, split 50/50 RSSB and NR; 7-year life. £15k per annum maintenance cost, split 50/50 RSSB and NR	These are the additional costs resulting from adopting a data broker approach.
RCM costs – higher integration levels	Additional costs: Data link £3k; System link £10k; Complex system £30k; External data £10k. Opex 30% per annum Integration levels need: HA1: 3 x DL 1 x SL 1 x CS HA2: 2 x DL 1 x SL	1: difficult to predict costs of the higher levels 2: extension categories Data Link, System Link, Complex System and External Data need to be revised in the light of recent IT / data integration practice and the growing availability of data services

Input	Assumptions	Notes
	PA: 2 x DL 2 x SL 2 x ED AGL 3 x DL	
Benefits vs current	Basic to SD: 5% reduction in asset failure rate HA1: 8% reduction in failure rate; 2% reduction in maintenance activity HA2: 11% reduction in failure rate, 3% reduction in maintenance activity PA: 20% reduction in failure rate; 5% reduction in maintenance activity AG:20% reduction in failure rate; 10% reduction in maintenance activity	These are purely speculative, particularly at the higher integration levels where the benefit comes from changes to procedure and practice that require multiple inputs and agreements.

Table 2 Business Case Tool Benefits Calculated

Party / level	Benefit £ per annum	Notes
Crossrail		Crossrail benefits to a small extent from the reduction in cost associated with track fault-related delays on NR infrastructure
• SD level	£3.5k	
• HA1 level	£5.6k	
• HA2 level	£7.7k	
• PA level	£14.0k	
• AG level	£14.0k	
NR Western		NR benefits to a small extent from the reduction in cost due to track fault-related delays. Much greater benefits accrue at higher levels of integration which permit reductions in maintenance cost.
• SD level	£13k	
• HA1 level	£200k	
• HA2 level	£300k	
• PA level	£500k	
• AG level	£1m	
RSSB		RSSB (or more general industry) bears a share of the cost of providing shared data services. The figures here suggest charges that RSSB or its contractor may wish to levy for providing these services to industry.
• SD level	-£12k	
• HA1 level	-£17k	
• HA2 level	-£18k	
• PA level	-£22k	
• AG level	-£26k	

Table 3 Template Contract Applicability

Section	Applies / Details	Notes
Parties	Yes. Basic contract is between "Facilitator" (= Crossrail) and "Beneficiary" (= Network Rail Western Route). Allows for other parties as well.	Will need to work out how to include RSSB and the Data Broker Supplier in the agreement, both in the development and production phases of the work. Initially, the Data Broker Supplier will be contracted by RSSB; this may change for the production phase.
Whereas	Yes. Will need to add the special circumstances of this project: role of RSSB and the Data Broker Supplier.	May need to define role of RSSB during development phase but possibly not during later production phase.
1 Definitions	Yes.	May need to define Development and Production phases.
2 Supply and Installation	Yes. Refers to Equipment defined in Schedule 2; supply and maintenance arrangements described in that Schedule; programme defined in Schedule 3.	
3 Operation and Maintenance of Equipment	Yes. Again, refers to Equipment as defined in Schedule 2.	
4 Intellectual Property	Yes. Refers to Schedule 4.	
5 Payment	Yes. Refers to Schedule 5.	
6 Warranties, Indemnity, Liability	Yes. Presumption (unless overridden in Schedule 1 or Schedule 6) of no warranty of any kind on Equipment or Data.	

Section	Applies / Details	Notes
	General indemnity for consequences of breach of this agreement. Subject to limit of liability in Schedule 1.	
7 Insurance	Yes. Refers to Schedule 7.	
8 Force Majeure	Yes.	
9 Confidentiality	Yes. Refers to Schedule 2 for specifics of permitted use of the Data. Refers to Schedule 1 definition of whether Freedom of Information Act applies.	
10 Costs	Yes.	
11 Assignment	Yes.	
12 Termination	Yes. Voluntary termination may be permitted after a notice period defined in Schedule 1.	
13 Consequences of Termination	Yes. Requires uninstallation of Equipment unless specified in Schedule 1.	
14 Governance	Yes. Describes a Joint Management Group, specified in Schedule 8.	
15 General	Yes. Refers to possible need to transfer to a future franchise operator. States that title to Equipment is identified in Schedule 1. Allows dispute resolution mechanism to be defined in Schedule 1.	
Schedule 1 – Contract-Specific Provisions		
Asset Protection Agreement	No: this only applies to fixed equipment on Network Rail infrastructure.	
Heads of Terms date	Maybe: depends on whether Crossrail and Network Rail	

Section	Applies / Details	Notes
	Western agree heads of terms.	
Conditions Precedent	Maybe: are there circumstances under which the agreement may not come into effect, or may be modified?	
Supply and Installation	Yes. Who is the owner of any Equipment (defined in Schedule 2)?	Depends on what is counted as Equipment in this sense: does it relate to software or networking setups specific to this contract?
Operation and Maintenance	Maybe: depends on Equipment. Includes right of access for 3rd party maintainers.	
Risk-Reward Arrangement	Maybe: depends on how business case is handled.	This needs to be agreed separately between Network Rail Western, Crossrail, RSSB and Data Broker Supplier.
Limit on Liability	Yes. Note could be 0 for all parties if no liability accepted.	
Force Majeure	Yes. Define number of days / months for which Force Majeure continues before contract is terminated.	
Confidentiality	Maybe. Define any other circumstances under which confidential data may be disclosed.	
Freedom of Information Act	Maybe. Identify parties who must comply with FOIA.	
Assignment	Maybe. Impact of change of corporate structure of any of the parties.	
Termination	Yes. Definition of whether voluntary termination is	

Section	Applies / Details	Notes
	allowed, by whom and with what notice.	
Consequences of Termination	Yes. Define who should uninstall Equipment, what contract clauses survive termination, who bears costs of termination.	
Governance	Yes. Is there a Joint Management Group? If so, refer to Schedule 8.	
Franchise Agreement	Maybe. Definition of what happens if one of the parties is a franchised operator and the franchise changes hands.	
Notices	Yes	
General	Yes. Define who owns title to the Equipment	
Disputes and Governing Law	Yes. Define dispute resolution mechanism or use the default provided (RIDR rules or TECBAR).	
Other	Maybe. Any other special conditions not covered above. Note that these take precedence over the contract terms if there is a conflict.	
Schedule 2 - Specification		
Equipment	Maybe, if there is anything which counts as Equipment. Defines the equipment, who is fitting it and where, any approvals require and by whom, industry standards that apply, ownership of equipment, who can use it, when and why may it need to be removed, any safety or operational conditions to be considered.	The Data Broker may count as Equipment. Use T1010-01 (as updated) to define the Data Broker.
Data Collection and Handling	Yes. Definition of Data, data lifecycle, frequency of	Use T1010-01 for definitions of data

Section	Applies / Details	Notes
	collection and provision, reporting, storage.	standards and formats.
Data Sharing Protocol	Yes. Definition of data sharing methods, access control, integrity verification, usage restrictions, fall-back scenarios.	Use T1010-01 for definitions of data sharing mechanisms and controls.
Schedule 3 - Programme	Yes. Timetable for delivery of Equipment; for delivery, transmission of Data. May also include schedule of train operations describing coverage of track by trains. May include any other dates (e.g. when train services start; when UGMS equipment is going to be available; when FAT / SAT phases of UGNS delivery have been completed.	May impose diagramming / dispatching requirements on Crossrail to ensure coverage of NR infrastructure.
Schedule 4 – Ownership and IP; Permitted Usage	Yes. Defines ownership and usage of the Data. In particular: Defines how Data, or derived data, can be used and by whom. Defines title to the Data, with several options: 1 – title rests with party which produces the Data 2 – title of original data rests with party which produces it, except that title to improvements or modifications belongs to the party who makes them 3 – title to IPR created in course of this agreement belongs to party that created it. Defines conditions for licensing of IPR ("Permitted Purpose"), with options:	There may be other options, particularly where there are several processing / augmentation steps. Reference should be made to the principles of ISO 13374, with data being open but IPR in processing being protected. Consideration needs to be given to rights of Crossrail, of 3rd party reference data suppliers such as NR ORBIS, and of the Data Broker supplier who will be providing value-added services.

Section	Applies / Details	Notes
	1 – all purposes including improvement 2 – a specific purpose or purposes.	
Schedule 5 - Payment	Yes. Has payment for: Supply of Equipment. Milestone-based payment schedule. Operation and Maintenance of the Equipment – either by regular payment, or by some other arrangement. Payment for Data. Either: No payment Time-based payment	In terms of Data Broker, it will be the Data Broker Supplier which provides it and maintains it. Operation / Maintenance cost may be driven by data usage rather than time. Data cost may be by usage rather than time.
Schedule 6 – Service Level Agreement	Maybe, if there are to be contractual penalties for failure to deliver. Driven by Schedule 2. Should consider availability, timeliness, quality, dependability of transfer, security, fault tolerance, response time, rectification procedure, escalation procedure, compensation, disaster recovery, performance regime.	Likely to apply only in Production phase rather than Development.
Schedule 7 - Insurance	Yes	Standard insurances and conditions.
Schedule 8 – Joint Management Group	Yes. Will cover representation, frequency of meeting, quorum, how to identify issues, responsibility for workstreams, risk management, decision making process and relationship with dispute resolution.	

Table 4 Data Architecture Applicability

Architecture Element	Applies / Details	Notes
Integration Level	The architecture must be applied up to the integration level of "Datagram Formats". The scope of Stage 2 could usefully also include elements at the "Interchange Methods" level to demonstrate the value of these aspects of the architecture.	The decision on the highest level of integration to adopt depends on the budget available and the requirements of the stakeholders.
Data Item Formats	Yes. The architecture requirements on open data formats conforming to current standards can be applied in full.	The data are supplied by Crossrail in .csv format and can be presented to NR Western in a mapped version of .csv. It may also be possible to present the same data in a json format which is more usable by web services.
RCM Data Types	Yes. The data provided can be expressed using some of the standard forms proposed by the data architecture	Data are provided in two forms with different levels of processing / aggregation at the Data Manipulation level.
Reference Data	Yes, partly. There will need to be a location mapping and possibly a train / timetable mapping.	The location mapping is potentially very useful and generally applicable: from GPS location to track location. There is potential value for a train timetable mapping which will give a schedule for when the trains equipped with UTGM will pass over a given track section.
Datagram Structures	Yes. The concepts of datagrams can be applied to the data output. Some of the extension points can be used to attach useful metadata.	The data architecture defines a datagram structure as an XML snippet which can potentially be mapped into other useful formats such as .csv and json. Recent developments in IT support this type of mapping.
Interchange Methods	Yes. The raw data are transferred in batches of indeterminate size covering	There would be much value in demonstrating the principles of a variety of interchange

Architecture Element	Applies / Details	Notes
	indeterminate time bands. There is much value in batching the data so it is more usefully presented.	methods: a sample web page and a sample queryable web service, for example.
Shared Data Artefacts		
Basic Data Format Specification	Yes. Data items common to all rail RCM data interchanges	
RCM Data Format Specification	Yes. Data item definitions for the specific RCM data items used in this interchange.	These will be expressed using a general method which is extensible to other types of RCM data (such as OLE pressure / position readings)
Reference Data	Maybe	It is not presently clear whether shared data will be needed.
Web Service Definitions	Maybe. The Data Broker may offer a sample web service.	There is considerable value in offering a basic web service showing how URL schemes and a REST-based approach can make data generally available in a consistent way.
Messaging Configuration	No.	Messaging infrastructure is beyond the scope of this project but would be a possible future extension of the Data Broker.
Rail Ontology	No / Maybe.	The Rail Ontology is beyond the scope of this project. However, there may be demonstrable in making some data available via a SPARQL endpoint: this is the basic data sharing mechanism for ontologically-structured data.
Shared IT / Infrastructure Artefacts		
Documentation Repository	Yes. The interface definitions and procedures must be shared on a website / wiki.	Arrangements will need to be made for hosting this information. Spark?

Architecture Element	Applies / Details	Notes
Reference Data Store	Maybe. If shared reference data are required, they should be made readily available online.	
Unique ID Manager	Maybe. It is not currently clear whether there is a need to provide a method of generating unique IDs for assets.	This would need to be provided by the Data Broker supplier.
Reference Web Services	Maybe. There may be value in offering for example a location mapping service.	This would be in the form of an on-line API specification and test suite.
ESB Messaging	Partially. It should be possible to deliver some basic patterns such as publish / subscribe and guaranteed delivery within the project scope.	A full ESB approach is beyond the scope of the current project. However, the intended scope will be able to be enhanced later if required to add this type of functionality.
Message Orchestration	No.	Beyond the scope of this project. However, the architecture proposed will be able to be extended to add this type of functionality.
Software Reasoners	No.	Beyond the scope of this project.

Table 5 Toolkit Gap Analysis

Area of Mismatch	Description	Notes
T1010-02 Process Map and Contract		
Parties besides provider and user of data	The draft contract mentions the Facilitator (= data provider) and Beneficiary (= data user) and indicates that there may be other parties. The T1010-02 report mentions various other possible parties, such as equipment and software suppliers, data analysers, data distributors, regulatory authorities. However, there is	The current project has several other stakeholders (described in Section D.1 above) who may need to be included in a contractual agreement. Specifically, RSSB and the Data Broker supplier are party; Network Rail as reference data suppliers may also be party.

Area of Mismatch	Description	Notes
	no guidance on how the interests of these parties should be fed into the contract.	
Terminology differences	The names of the key and subsidiary parties in the Commercial Principles document do not match those in the template contract.	The commercial principles describe a Scheme Lead or Joint Scheme Leads. Facilitators are other parties. The contract has a Facilitator as the data provider and Beneficiary as the data recipient.
Arrangement	The template contract is geared to a direct transaction between a data supplier and consumer. It does not cover the case where a supplier offers a general service.	This will be adequate for the immediate project requirement. However, it will not support future service-based approaches, such as where a supplier offers a data-processing service to all parties.
Risk / Reward arrangements	The report mentions these arrangements but there is no guidance on how they should be represented in the contract.	The current project would benefit from such an arrangement. It is one way to incentivise investment in the data broker.
Data Status and IP	The commercial principles describe “raw data” and “processed information”. The principles also state that IP in the “processed information” goes to the party for whom it is processed. In fact, data go through many processing steps, so each such step has “input data” and “output data”. The template contract has 3 options for IP, none of which reflects the stated commercial principles: 1. IP in data belongs to the party that produces it 2. Ditto, but IP in improvements / modifications to the	This is a key area that needs to be clarified. The difference between IP / title / ownership of the data and that of the processes that act upon it needs to be made explicit.

Area of Mismatch	Description	Notes
	data belongs to the party that did the improvement 3. All IP in data belongs to the party that required its creation. There is no guidance in the documentation about how the complex IP issues should be addressed in the contract.	
T1010-04 Business Case Tool		
Track assets	It is not clear from the documentation how linear assets should be entered and their costs handled.	It was necessary to dig into the formulae of the spreadsheet to work out how to enter these costs in a way that calculated correctly. Documentation needs to be updated.
IT components	The IT options associated with RCM sharing schemes are out of date and restricted to single party to single party interactions.	The approach taken needs to be generalised to make it work with service-based and cloud-based solutions and pricing mechanisms.
Schedule 8 payments	With a train operator and an infrastructure manager involved, it is tricky to handle the treatment of schedule 8 performance-related costs.	For infrastructure faults, the cost of the delay is incurred by the operator but charged to the IM, but there may be additional reputational cost to the operator not covered by the schedule 8 payment.
Shared service providers	The model assumes a direct relationship between a data supplier and consumer. It is not obvious how to represent a party that is a service provider.	The documentation needs updating with some scenarios.
T1010-01 Data Architecture		
Schema	The toolkit defines datagram formats using XML Schema. This is too heavyweight an approach for the current project.	JSON-Schema is a suitable alternative which is lighter-weight and maps on to current web service technologies better.

Area of Mismatch	Description	Notes
Simple protocols	The toolkit suggests the use of an Enterprise Service Bus to provide interchange protocols. Though the principles of an ESB remain valid, other less onerous methods are now available to implement them.	Simple API-driven functions using cloud-based microservices can replace the ESB with a cheaper and more extensible alternative.
Shared data model	The toolkit suggests that a shared data model is necessary to support data sharing at anything but the most basic level. This may not be so as long as the participants have a common view of the data.	Notwithstanding this, there is value in starting development of a shared data model to support future mapping efforts.
API approach	The toolkit introduces but does not require an API-based service orientation. Since it was written, the API approach has become the mainstream method of linking heterogeneous systems.	New tools and techniques have emerged and become popular to make the development and use of APIs much easier.
Usage notes	The toolkit is not structured to help implementers or users of a data architecture.	Usage notes will be developed as part of the Stage 2 work.

Table 6 Use Cases for UGMS Data

Use Case	Description	Notes
Alerting to new faults	Issue an alert whenever a significant exceedance is discovered that would indicate a serious track fault such as a broken rail.	There is a parallel process already in place, by which Crossrail will alert NR Western immediately train-based equipment detects a fault. This means that this use case is not a priority for the current project. However, there may be value in setting up a generic alerting mechanism that can support and enhance the existing process as this will be generally useful to other customers.

Use Case	Description	Notes
		This is an “SD” function in the terms of ISO 13374.
Condition changes over time	Monitor the condition of a section of track over time to see the rate and extent of degradation, to support early intervention before failure.	This is a promising use for this data, if it can be presented and visualised in a way which makes it useful. Suggests a data store with a web interface to visualise metrics for a track section and their condition over time. This is a combination of “DM” functions in terms of ISO 13374, with the possibility of adding an “SD” alerting function for rapid or accelerating deterioration.
Verification of Fix	Verify rapidly that an intervention to correct a defect has in fact been completed.	A strong use case for this data, as it comes available in general much earlier than that from a measurement train. This requires notification that a track section that was defective is now not – i.e. it is a combination of DM stages (identify track status) and SD (detect change in state and issue an alert).
Investigation of Influences	Look for correlations between the rate of deterioration of track condition and influencing factors such as ground conditions, weather, traffic mix and quantity, vegetation.	A case where value may be added by the data because readings occur at a high frequency and can therefore detect rapid changes in track state. This requires a long-term data store and the means to integrate other data streams. Probably beyond the scope of the current project but one to consider for a later production installation.

Table 7 Basic Track Geometry Measures

Geometry Measure
Top Left 35m
Top Right 35m
Top Left 70m
Top Right 70m
Top Mean 70m
Alignment Left 35m
Alignment Right 35m
Alignment Mean 35m
Alignment Left 70m
Alignment Right 70m
Alignment Mean 70m
Gauge
Crosslevel
Twist 3m
Twist 5m
Curvature
Gradient

Table 8 Additional Track Geometry Components

Measure Component
Cant Deficiency Comp 1
Cant Deficiency Comp 2
Cant Deficiency Comp 1 HP
Cant Deficiency Comp 2 HP
Cant Deficiency Comp 1 LP
Cant Deficiency Comp 2 LP
Left Dipped Joint
Right Dipped Joint
Cyclic Right Top 4.5m
Cyclic Left Top 4.5m
Cyclic Right Top 6m
Cyclic Left Top 6m
Cyclic Right Top 9m
Cyclic Left Top 9m
Cyclic Right Top 13m

Measure Component
Cyclic Left Top 13m
Cyclic Right Top 18m
Cyclic Left Top 18m
Cyclic Left Top 4.5m Accu
Cyclic Left Top 4.5m Peak Count
Cyclic Right Top 4.5m Accu
Cyclic Right Top 4.5m Peak Count
Cyclic Left Top 6m Accu
Cyclic Left Top 6m Peak Count
Cyclic Right Top 6m Accu
Cyclic Right Top 6m Peak Count
Cyclic Left Top 9m Accu
Cyclic Left Top 9m Peak Count
Cyclic Right Top 9m Accu
Cyclic Right Top 9m Peak Count
Cyclic Left Top 13m Accu
Cyclic Left Top 13m Peak Count
Cyclic Right Top 13m Accu
Cyclic Right Top 13m Peak Count
Cyclic Left Top 18m Accu
Cyclic Left Top 18m Peak Count
Cyclic Right Top 18m Accu
Cyclic Right Top 18m Peak Count