Highways Agency

ITS Data Registry Pilot
Final Report

September 2006

Mott MacDonald
1 Atlantic Quay
Broomielaw
Glasgow
G2 8JB
UK
Tel: 44 (0)141 222 4500
Fax: 44 (0)141 221 8083

ITS Data Registry Pilot
Final Report

Issue and Revision Record

Rev	Date	Originator	Checker	Approver	Description
A	19/9/06	Ian Cornwell	Alastair Dunsmore	Ian Cornwell	First Issue

List of Contents

Page

Executive Summary

This report presents the results of a one year pilot operation of a "data registry" for intelligent transportation systems (ITS). This pilot followed from an earlier research project which constructed and trialled a data registry for the Travel Information Highway (TIH) community in order to increase the efficiency of information exchanges between the Highways Agency and its partner organisations. The objectives of the pilot project were to improve and exploit the registry service to begin to achieve the identified potential benefits, and to explore the application of the registry in other contexts within the Highways Agency.

Results of the Data Registry Pilot Project

In the year of pilot operation the registry has provided an expanded set of benefits in improving quality, in efficiently answering data questions, in joining up separate developments, and in harmonising data definitions, which should lead to cost savings in system development.

The registry facility created by the research project has provided an appropriate technical foundation, though refinements have been required to increase usability in order to achieve the potential benefits for users. In particular, harmonisation of data definitions requires something more than published registry processes, and the project developed a technique of "core components analysis" which has contributed to an improvement of quality of crucial models and the reflection of needs from across the range of registered models.

The Highways Agency's Information Directorate now see the ITS data registry as fulfilling needs that are closely related to their own requirements to improve data management and to provide business intelligence reporting facilities. The planned system architecture for Information Directorate's business intelligence framework includes a "data dictionary" component. The Roads Information Framework team and the registry team both realised that the work on the ITS Data Registry was closely related to this component. A series of workshops was held, leading to an understanding of how the registry experience and technology could be used within the overall business intelligence framework for the Highways Agency.

Recommendations

1. **Apply the data registry experience, approach and technology to the Highways Agency's internal data management needs.**

This will allow the Information Directorate to begin to catalogue data within the organisation, identify duplication of functionality and data to facilitate consolidation with the aim of increasing operational efficiency, and facilitate delivery of business intelligence solutions by providing accurate information on source systems.

2. **Operate, improve and exploit the ITS Data Registry for the Highways Agency and its partners in the Travel Information Highway community**

The Highways Agency should continue to operate and exploit the ITS data registry service to underpin discovery and harmonisation of data definitions and interfaces, resulting in improved efficiency in data exchanges between the Highways Agency and its partner organisations.

1 Introduction

1.1 Purpose

This report presents the results of a one year pilot operation of a data registry for Intelligent Transportation Systems (ITS). It summarises conclusions reached in each potential area of application, and makes recommendations for further progress of the registry as a data management tool.

The report describes how the pilot made progress building on the recommendations made by an earlier research project. This report can be read on its own, but for full background details the reader should also consult the final report from the research project [1].

1.2 Project background

Recognising the potential benefits for data management, the Highways Agency commissioned a research and development project to identify and test a suitable registry structure and mechanism that meets the needs of the Agency and other organisations, specifically for Travel Information Highway (TIH), but with the possibility of extension to the wider ITS sector. The research project completed in March 2005 and recommended a one year pilot to apply the research results and to explore the benefits in different application areas:

- Information exchanges between Highways Agency and its partner organisations, focusing on the Travel Information Highway (TIH).
- Internal information exchanges and developments, such as RIF, NMCS Site Data remodelling, and common location referencing.

This is the report from the pilot, which ran from July 2005 to August 2006.

1.3 Report structure

Sections 2 and 3 report the activity and progress of the pilot, beginning with user requirements and explaining the developments introduced to address these requirements. The final report from the research project [1] has already given a comprehensive treatment of aspects such as user needs, registry structure, and registration process. The details are therefore not repeated here, but refinements to our understanding in these areas are summarised. Harmonisation has been a technical focus in the pilot, and has a dedicated section (section 3).

Section 4 summarises progress from the viewpoint of each target business context for application of the registry.

Section 5 presents cost and benefits, and finally section 6 contains the recommendations of the project.

2 Technical progress of the pilot

The research project provided an understanding of the user requirements for a registry service, and created a registry facility to begin to meet those requirements. In the pilot project, our understanding of the requirements was refined and the registry service was improved to meet requirements more completely. The following sections describe the refined requirements and the associated technical developments.

2.1 User requirements for a registry service

A registry solution is derived from its user requirements and these in turn must be chosen to fit the business process of the user organisations. The research report [1] gives a full treatment of this subject. These requirements are still valid, but further user experiences plus reviews of the registry website suggested that the registry had to be easier to use in order to fulfil its potential. The registry website needed to be more inviting and more intuitive; it had to be easier to navigate and easier to find items of interest for people with different viewpoints. Developments to address these requirements are described in section 2.4.

2.1.1 Requirements for a Highways Agency data registry

During the pilot the registry project team held a series of workshops with Highways Agency's Information Directorate to understand how the registry experience and technology could be used in an internal data registry operating as a component within a wider business intelligence framework. The internal context brings additional requirements for directory services. The internal user does not simply want to know how data is structured and where it is defined, but also:

- Who manages / owns / supplies this data?
- What people and systems use this data?
- In what system is the data held?
- Where is data duplicated?
- What data falls within the Data Protection Act?

These requirements do not fall under the scope of the registry in the TIH context where the registry is coupled to the external TIH directory of services. Information Directorate is planning further work to address these requirements, building on the data registry experience.

2.2 Assessment of Registry Structure

The structural foundation established in the research project proved a sufficient foundation for the continued operation of the registry service in the pilot. The foundation is the use of UML for all registered items and also for additional structures for indexing and for registry administration. A mapping is observed between UML and the ISO 14817 ITS data registry standard. XML Schemas are supported through translation to UML.

The main structural advances of the pilot were the addition of a subject matter index (see section 2.4) and the refinement of the "core components" area (see section 3) to encourage improvement and harmonisation of registry contents.

The pilot project also assessed how the ITS data registry fits with the general purpose standards in this area: ISO/IEC 11179 (metadata registries) and ISO/IEC 19763 (a metamodel framework for registry interoperability). The ISO 14817 ITS standard was based on a much earlier version of ISO/IEC 11179 and the latter has developed significantly since that time. The developers of these standards are addressing many of the same requirements that are important in the UK ITS context. Registry administration attributes in ISO/IEC 11179 are more extensive and better normalised than in the ITS counterpart. However, ISO/IEC 11179 is currently lacking in support for structural modelling, missing some of the features that we have found useful in UML. This is one reason for the development of ISO/IEC 19763, but that draft standard has taken a technically ambitious approach that makes it harder to use existing commercial tools, and furthermore it is almost impossible to understand in detail due to the writing by non-native English speakers. The conclusion of the registry project team is to continue with the existing structural foundation.

2.3 Assessment of Registration Process

The research project established a registration process in which roles defined in ISO 14817 were mapped to existing bodies in the Travel Information Highway context. The process is fully described in the research report [1]. The pilot continued to follow this process but with the following refinements:

- It became clear that unfunded volunteers will not in general contribute reviews as active stewards. Typically the registrar was in a position to conduct an independent review which was sufficient. Additional reviews would be required if (i) the registrar was not independent, having been involved with the production of the items, or (ii) the review was for the highest level of status ("preferred"). In this case the managing organisation must arrange reviews using a funded source. In the case of TIH, the reviews would be conducted by members of what is now called the "TIH Technical Board". This group therefore replaces the TIH working group as the registry steward.

- Harmonisation requires an additional catalyst – published registry processes are not sufficient to encourage significant progress. Progress on harmonisation is described in section 3**.**

2.4 Enhancements to the registry facility

As the registry moved from research trial into operational pilot, it became important to enhance the usability of the registry web site, which is the principal way in which people interact with the registry to achieve benefits.

Enhancements included:

(i) Registry URL

The registry URL address was changed at the start of the pilot. A new URL, www.itsregistry.org.uk was established as the home of the registry.

(ii) News and content overview

A new top level news page was added and maintained to highlight the continuing work carried out by the registry and to show to users that the ITS Registry is an actively maintained resource.

(iii) High-level pictorial navigation

To help users understand the layout of the ITS Data Registry, a hyperlinked diagram was added at the top level. It shows how the branches of the registry are inter-related, and highlights options for navigating and searching registry content. It provides a more reassuring and stimulating welcome than the previous list of folders – first impressions are important!

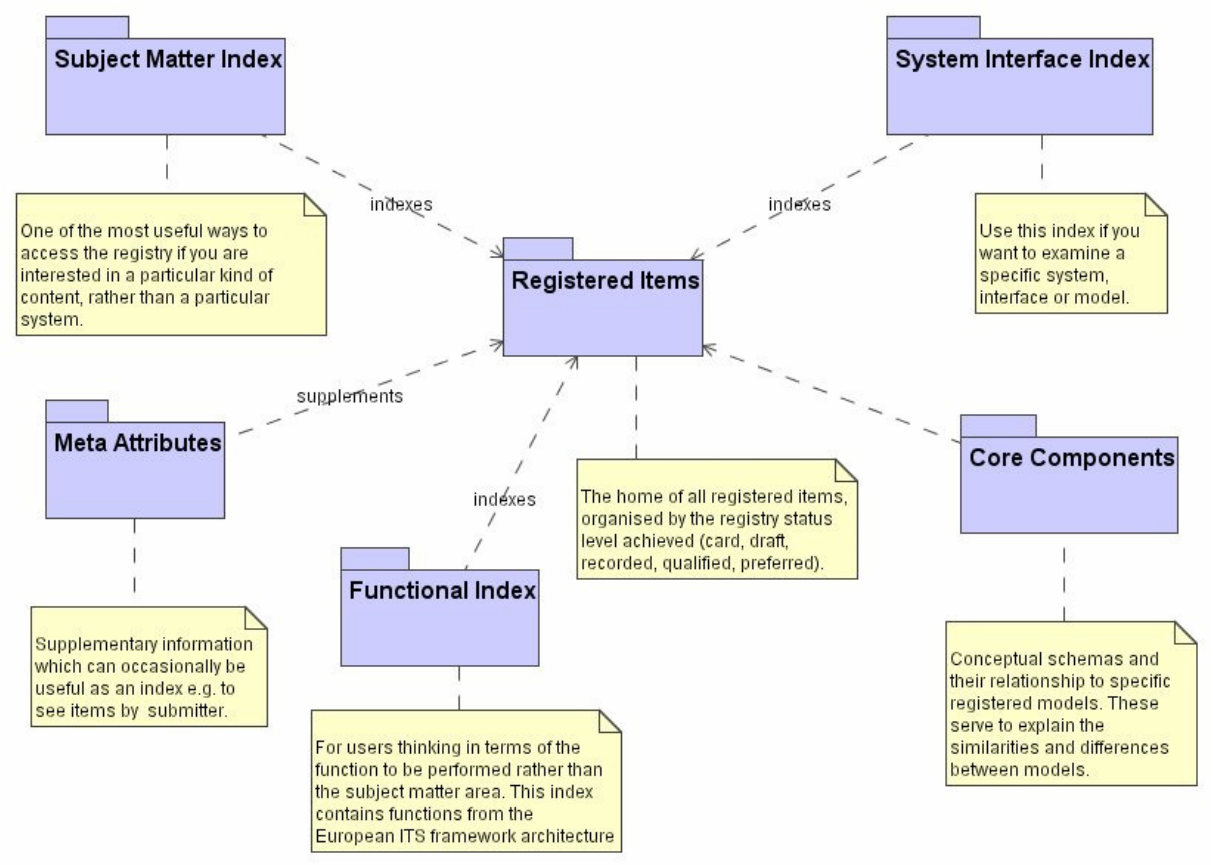

Figure 1 Entry level diagram

(iv) Content navigation

The previous method to allow a user to navigate through content had two limitations: the controls took up valuable screen space that could be used to show diagrams, and the lack of a familiar tree view caused some users to lose track of their location within the registry. Two alternative navigation mechanisms were added to the registry website to counter these problems, and the user can choose between them.

(v) Subject matter index

Experience suggests that registry users searching for content often think in terms of the real subject matter entities that they wish to represent. A new subject matter index therefore provides a classification of model elements based on the real world concept being modelled. This classification does not follow any existing standard, but is chosen by the registrar to provide an effective index of the registry content. The classification will be adapted to usefully index the registry content as it expands over time.

(vi) Search

The registry already offered logical search mechanisms, but today's web users expect a simple text search as offered in popular search engines. A text search facility was developed and added to a new page that also explains all the various search facilities of the registry.

(vii) Improved imports of XML Schema

The research project developed software to translate XML Schema into platform-independent UML models for inclusion in the registry. The software was sufficient for the XML Schema imported at that time, but the pilot encountered submissions with additional XML Schema constructs ("choice" and "restriction"). The tool was improved to be able to successfully translate these Schemas, and this enabled the registration of further important XML Schema such as the public transport standard "SIRI".

2.5 Content of the Registry

During the pilot period the content of the registry increased to cover a significant proportion of the information used in exchanges between the Highways Agency and partner organisations. The following functional areas are now covered:

Highways Agency

- Traffic Operation structures – Scheduled Road Works, RCC Command and Control, and HALOGEN information services
- Asset Management structures - HAPMS Location Referencing
- Travel information services – DATEX II, NTCC TIH, Unified OTAP, Journey Time Database, Video Information Highway

Urban traffic

- Traffic management - UTMC v2 and the specific example of MATTISSE
- Local streetworks - SDEP

Public transport

- Real-time information – RTIG, SIRI

- Public transport planning and registration - TransXchange.

General travel-related

- TPEG, the broadcasters' standard for travel information distribution
- Aspects of Transport Direct.

Additionally the registry contains preferred data types which are building blocks, and "core components" which are derived constructs used to encourage harmonisation (see section 3).

The full range of content can be seen online at www.itsregistry.org.uk.

2.5.1 Source Formats

Submissions were made in a variety of formats. Proportions are illustrated in Figure 2. The most common format was UML submitted as an XMI file. The most common electronic source requiring transformation to UML was XML Schema, although CORBA IDL and entity-relationship models were also submitted electronically and transformed. A number of submissions were synthesised by the registrar from source documents in various forms including XML DTD and excel spreadsheets; in these cases the formal definitions were supplemented by significant information that had been expressed only in informal text.

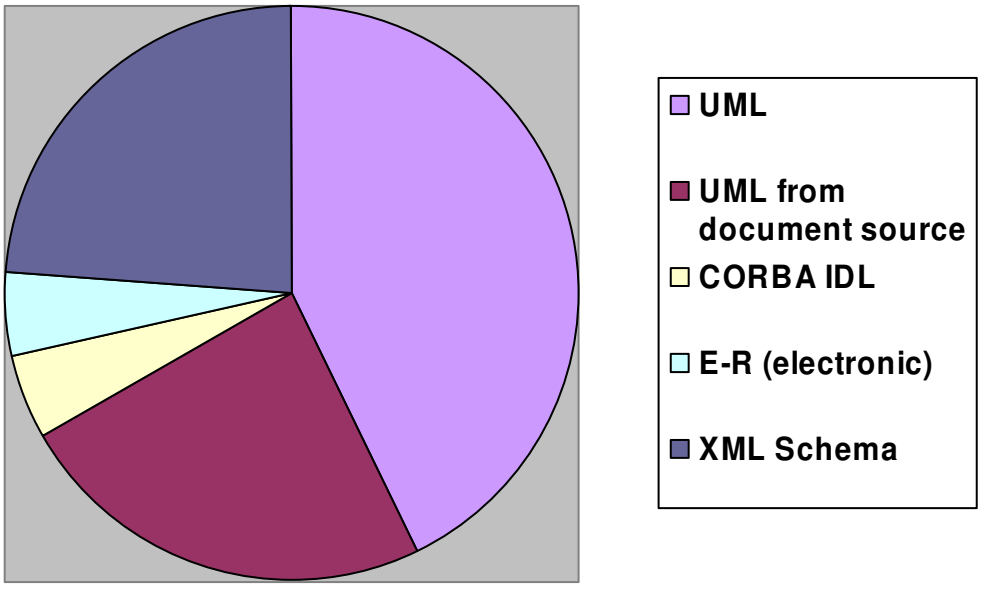

Figure 2 Source formats of registry submissions

The overall pattern of submissions justifies the choice of UML and support of XML Schema through transformation.

2.6 International feedback for ITS data registry

The ITS data registry has gained international plaudits through standards development channels. The largest and most advanced registry community in the world, grouped around ISO/IEC JTC 1 SC32 WG 2 “metadata”, invited the project to present at two annual Open Forums, in Xi’an, China in 2004, and Kobe, Japan in 2006. These presentations led to widespread interest in our work, endorsement of approach and an invitation to participate in ISO/IEC standardisation as UK representative. In the ITS context the registry is now well known to ISO TC 204 (ITS) WG 1 “architecture”, and it is considered to be the most advanced ITS data registry implementation in the world. Australian and US participants in particular are actively involved in discussions on how best to take this work forward in standards and for their own contexts.

3 Harmonisation

"Harmonisation" is the process producing convergence of data definitions used across different systems, leading to greater interoperability and reuse. Harmonisation is an important function of a registry process, one that is essential if the largest benefits are to be achieved. Literature in the metadata registry field often assumes that use of a registry will lead to harmonisation. However no known source satisfactorily explains how to achieve harmonisation in a mature domain where there are already established implementations and standards, and (as in TIH) where there is no single controlling authority who can enforce the use of one particular standard even if this means costly changes to existing implementations.

The ITS Data Registry required a method of encouraging harmonisation across a diverse domain without relying on the enforcement of a single solution. The registry already had multiple overlapping submissions in many functional areas.

Methods of harmonisation can be characterised as "top down", "bottom up" or "middle out". The ITS Data Registry uses these approaches in varying degrees.

- "Top down" standardisation would flow from an agreed UK national ITS architecture and associated standardisation programme. The ITS Data Registry is not in a position to enforce its own standard, but makes a limited contribution by providing top-down indices of registry contents according to subject matter and function. These also signal the overlap across registered models.

- "Bottom up" harmonisation begins with agreeing common primitive data types, proceeding to specific types and further up towards ITS-specific data definitions and structures. The ITS Data Registry has an agreed set of "preferred" data types that can be used as a foundation for further agreements.

- The most important harmonisation tactic used in the Pilot can be viewed as a "middle out" approach – the technique of "core components analysis".

3.1 Core components background

"ebXML Core Components" is a UN/CEFACT specification used in e-business standardisation. The key idea that supports harmonisation is the separation of "core components", which have no specific business context, from "business information entities", which apply in specific business contexts. In terms of our registry, the registered models all have their individual business contexts and they can be seen as context-specific instantiations of a single set of underlying concepts, the core components, as shown in the example of Figure 3.

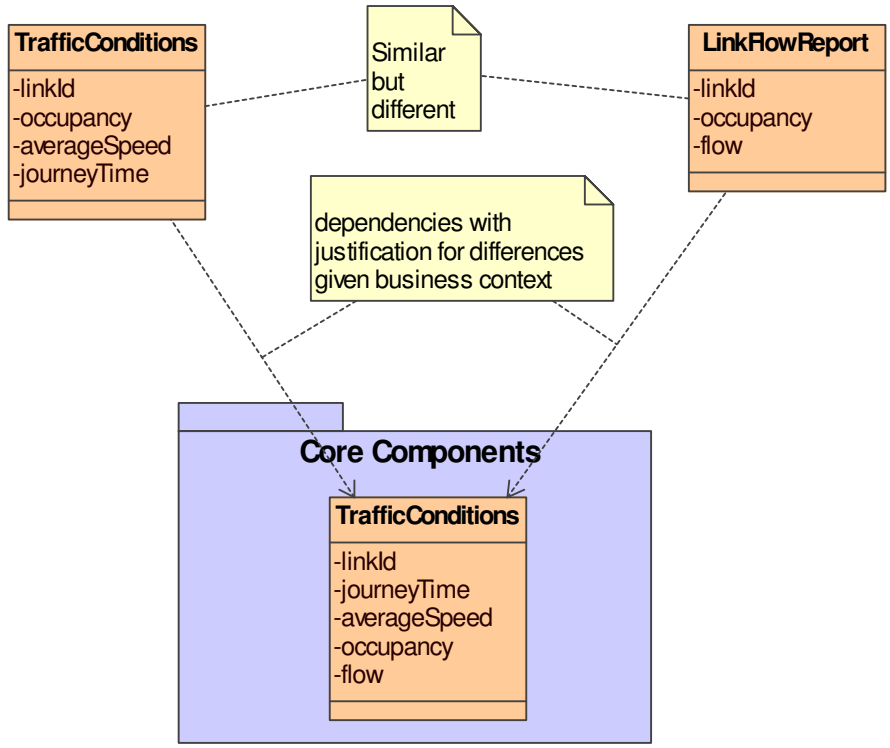

Figure 3 Core component example

The registry research project suggested that the core components could effectively demonstrate the similarities and differences between alternative representations of the same underlying concepts, and could express the justifications for differences due to business context.

3.2 Core components analysis

The pilot project put this idea into practice on a larger scale and devised the process of "core components analysis". In this process a set of core components is derived through analysis of the existing registered models. The process is as objective as possible to avoid the possibility of the core components being yet another competing model. The core components are not introduced unless justified by existing models. The engineer resists the temptation of "fixing" undesirable qualities in registered models unless there are other registered models to justify the improvement.

The outputs of the analysis process are the set of core components plus the mappings from registered models to core components. Each dependency drawn between a registered item and a core component must have a justification for any differences. These justifications are then agreed with the submitters before publication on the registry website. The exercise of agreeing a mapping helps to expose the quality of the specific model. If there are legitimate justifications for the shape of the model, those will be declared, but if the model contains poor design or sloppy thinking, these should also be exposed because a good business justification cannot be agreed. This gives a powerful way to provide objective review feedback and improve the quality of submitted models, since there is often no genuine justification and the submitter can understand and agree this. Furthermore the improvements made through this process are in line with requirements brought by other models, and often the specific change brings the model closer to the others.

Core components have now been derived in the areas of location referencing, traffic measurements and VMS. The process has led to significant improvements of DATEX II in particular.

However, core components analysis can be time consuming, and in particular registration of the mappings can produce high maintenance if the registered models are not stable.

The recommendation is to proceed with core components analysis, not in every area, but only in subject matter areas where further modelling or translation work of relevance to the Highways Agency is imminent or ongoing. If the registered models are relatively stable then the mappings shall be registered. If the models are undergoing active development then the mappings should be explored with the submitters, but not registered until the model achieves greater stability.

3.3 Core components can point out the most suitable models

The core components analysis process should also lead to identification of "preferred" items within the registry. The scope of the "preference" would be the business context declared for the specific model, in the subject matter area under study. For an item to reach "preferred" status, the justifications should all be sound.

At present, with many important models under development, "preferred" status cannot be awarded to any significant portions of registered model because this would condone some remaining deficiencies. It is common for 95% of an area to make sense and be well modelled, but for the other 5% to have flaws that cannot be cleanly separated from the good parts. The ideal situation would be that the model was fixed, and was then promoted to "preferred" status. However, models often have existing implementations so change is difficult and slow.

In this situation the core component mappings play an interim role of pointing out the available choices and making the deficiencies apparent. It is recommended that the prominence of the core components in the registry should be increased, so that users consult the core components and their mappings to discover which model is best for their specific business context.

3.4 Summary of value of core components analysis

- Makes similarities and differences explicit
- Mappings process distinguishes justified design from flawed design
- Generates objective feedback to submitters
- Can use understanding when building translators
- Signals recommended models for a specific business context.
- All the thinking is exposed to future system designers.

Technical details of the core components analysis are available in technical notes [2,3,4,5].

4 Progress of the registry in specific target contexts

The research trials recommended a number of contexts in which the registry could be beneficially applied. This section summarises the progress of the registry pilot in each of these areas.

4.1 Travel Information Highway – Highways Agency and partner organisations

The pilot confirmed the registry as a valuable resource for TIH. The registry continues to provide service definitions for the TIH directory service. More importantly the registry was used as a focus for reviews of models leading to significant improvements of DATEX II. The registry was also used to prepare for TIH standardisation workshops on street works, VMS and journey time.

The urban/inter-urban link is supported through the inclusion of the UTMC common data objects in the registry. During the pilot the registered UTMC model was brought up to date with recent updates in UTMC, and our registry provides the only public source of information on UTMC data objects where the relationships are explicit. Further progress on alignment of HA and UTMC registries and contents was slower than hoped: during the pilot period the UTMC Development Group (UDG) seemed to be in throws of re-organisation and the opportunities for discussion did not materialize.

4.2 Highways Agency – RIF and Information Directorate

The research project indicated the potential benefits of a data registry for the Highways Agency's systems, and pointed out the relevance to the Roads Information Framework (RIF), then under active development by the Information Directorate.

During the time of the registry pilot project, the initial phase of RIF development was completed and the Information Directorate began to consider plans for a wider business intelligence framework to deliver their Information Strategy. This framework would:

- Improve the ability to access corporate information
- Ability to provide reports in a flexible manner to a wider audience
- Ability to rationalise data and processes

As part of this framework, Information Directorate foresaw a need for a "Data Dictionary" component. The RIF team and the registry team both realised that the work on the ITS Data Registry was closely related to this Data Dictionary Component. A series of workshops was held, leading to an understanding of how the registry experience and technology could be used in the "Data Dictionary" component within the overall business intelligence framework. Unlike the public-facing ITS Data Registry, this new deployment would be internal to the Highways Agency. The RIF team now require a prototype development and deployment, which will be delivered outside the scope of the registry pilot project.

4.3 Highways Agency – HATMS Community

The research project also indicated potential benefits for the Regional Control Centre (RCC) Working Group on "Site Data and Initialisation", who planned to rationalise the current site data in order to support the more demanding requirements of RCCs. However, the Highways Agency took a different route to achieving similar benefits – a single supplier was contracted to produce a new improved data model. The registry project provided technical advice in workshops with this group, and the approach eventually used was technically similar to that of the registry project in that normalised UML models are developed in a UML tool and distributed as navigable HTML although not online on the web. The Highways Agency was able to take this approach in this case because:

- There was already a coherent group that included all the necessary parties.
- The scope was sufficiently limited that it was possible for a single supplier to grasp the details of the whole area and successfully remodel it.
- There is a desire not to share this information with any outside parties due to security sensitivity.

The registry project therefore stepped back from this effort after seeing that it was on track and using best practices.

5 Benefits and costs

Further experience of registry operation in the pilot has allowed us to refine the measurement of essential costs and also the estimate of benefits.

5.1 Costs

In this section only costs incurred during the pilot project are analysed, since these best reflect the current registry business processes.

5.1.1 Submitter costs

Several submitters provided a summary of costs incurred in making the submission. The required effort depends on the availability of a satisfactory electronic definition and its format.

Seven submissions were based on existing definitions in formats already supported by the registry facility. Here the cost was trivial for the submitter, who merely had to email existing files to the registrar. The cost in these submissions resides with the registrar who transformed the data for inclusion within the registry.

There were also three cases in which UML models were hand-built especially for submission due to shortcomings in the existing electronic definitions. Effort ranged from to 3.5 hours to 1.5 days.

No submitter incurred any software purchase costs.

5.1.2 Registrar costs

- Once all technical facilities were in place, the effort required for the mechanical stages of registration varied from 30 minutes to 3 days per submission, with an average of 5 ¾ hours.
- The production and validation of each web site update requires around 2-3 hours of effort.
- Subject matter indexing adds in the order of 30 minutes to each submission.
- Each area of full-scale core components analysis has required in the order of one week of effort to derive the analysis, create mappings, agree these with submitters and publish the results. Total effort is harder to quantify because the full analysis is only performed to support areas of special interest to the Highways Agency.

These costs do not reflect the role of the registrar in continuous improvement of the registry facility. Each new submission may potentially raise a new issue, and although the frequency of occurrence is decreasing with time, the pattern is expected to continue to some extent even beyond the pilot project. While the registrar role could be separated from the provision of detailed technical direction for the registry, experience suggests that it is efficient for these aspects to be combined. The registrar also shepherds the stewards. For these reasons an effective registrar service would typically incur 3-5 days of effort in total per major submission, with occasional higher peaks when technical developments are required.

Including re-submissions there were 12 submissions during the pilot in a period of around 14 months. Looking at the new contexts for registry application and the known metadata sources which have not yet been registered, we expect a slight slowing of submissions. On that basis, one person spending approximately 30% of working time (around 7 days/month) as registrar should provide an efficient service.

There also may be software licence costs. UML tools are relatively inexpensive – for example the annual cost of subscription for the tool employed in the registry at time of writing is $259 / licence / year.

5.1.3 Steward costs

Evidence from the research trial phases suggested that a feasible overall steward cost in an active and successful registry would be 5-6 days per month. This was based on a higher rate of submissions than is now expected in the future. Based on pilot projections, the corresponding figure would be 3-4 days per month. The research trials also suggested that 25% of reviews would be contributed by non-funded parties with an interest in the area. This level of support did not materialise in the pilot, and we must conclude that all stewarding must be funded by the central funding organisation – in this case the Highways Agency.

5.1.4 Total cost to the funding body

The pilot suggests that, using feasible rates for roles required and adding appropriate project management and administration, the total cost to the funding body for the operation of an active and successful registry could be in the order of £5,000 per month. However, this figure includes costs that would be incurred for activities required even if a registry was not used to facilitate them: discovery, dissemination and review of data definitions, data models and interface specifications.

5.2 Benefits

Benefits from a metadata registry lag behind investment, possibly by up to 2 years. However, the pilot is now demonstrating a number of qualitative benefits and has yielded enough experience to clearly envisage further expected benefits and to make approximate quantitative estimates in some areas.

The following discussion assumes a Highways Agency business context, but the benefits listed should occur in any business context in which the registry is applied.

5.2.1 Registry increases the quality of ITS models and their documentation

The pilot continued to produce valuable review points for several models. In the research trials the improvements caused by the registry had mostly been in quality of documentation. In the pilot the core components analysis also encouraged significant improvements to designs and features of models. In particular significant improvements to DATEX II followed shortly after the registry reviews, increasing the chance of DATEX II being usable for different applications in the future.

An improvement in quality could help avoid a misinterpretation of the data definition in the construction of a system. An individual mistake of interpretation and the resulting rework at integration time could cost many hours of effort. In some cases the improvements in quality will simply remove the need for further research by the user, saving up to a few hours of effort.

The improvements in features and design are more significant. They potentially make the difference between a standard being usable and being unusable for particular applications – and this can save weeks of future integration effort.

5.2.2 Registry increases the users' ability to answer questions

By using the information in the registry, requests for information can be managed more effectively. For example, in a specific case in 2003, Highways Agency staff needed data on incidents on the network but did not know where to find the data. Eventually, after chasing round a number of data sources over a period of days, the answer was found. Had the registry been fully established within the Highways Agency, the answer would have been identified within minutes. Amongst the TIH technical community known to the registrar, the registry is often used as a resource to answer technical questions.

The registry is also a convenient way for public sector organisations to provide information to external stakeholders, in line with the open government code of practice.

5.2.3 Registry increases the efficiency of system integration and rationalisation

The Highways Agency is already engaged in significant integration and rationalisation tasks. The registry provides an efficient support mechanism for this activity. During the pilot the registry allowed the analysis of streetworks, VMS and journey time models to support negotiations between NTCC and partners to agree formats for future information exchanges. The registry made the analysis more efficient and was able to introduce requirements and techniques from other areas.

The registry was used to support the re-use of DATEX II in at least two developments: the HALOGEN information service, and the combination of DATEX II and SIRI in a development for Leicester City Council.

During the earlier research phases, the registry played a facilitating role in the integration of the TPEG-Loc location referencing method into the Unified CENTRICO OTAP model.

5.2.4 Registry promotes re-use of data definitions, enabling integration

In addition to increasing the efficiency of integration, the registry also plays an earlier role in creating awareness of the existing specifications in the first place, and promoting their re-use. It is difficult to know with certainty the effect that the registry has had in this area, and what would have happened without it, but it at least seems certain that the registry has strengthened the presence of the various models in the minds of its users, and it may even have directly led to some of the integrations mentioned above.

A beneficial side-effect of the registry is the ability to provide electronic copies of definitions for re-use in external tools, saving translation or re-keying. On the single occasion this feature has been used, the registry helped the OTAP community migrate from Rational Rose to Enterprise Architect with no significant costs.

5.2.5 Registry underpins directory services

The registry supports the TIH directory of publishers by describing the data content supplied by each TIH publisher. Similar directory services can be created as the registry is applied to other areas. As the number of data sources increases, so will the complexity of finding, comparing and understanding the data that the user requires. An integrated directory service promises to mitigate the increase in complexity and provide a catalyst for uptake. Potential users can easily find out the content offered by the sources. Days of effort can be saved per user. Users who would otherwise give up plans of ITS data consumption may be encouraged to continue because they can see valuable content. The directory service is therefore important in strengthening the business case for submitters, who gain direct advertising.

5.2.6 Benefits in External Data Registry Experience

An interesting parallel may be seen with the introduction of a Data Registry in the US Environmental Protection Agency (EPA). The EPA organisational structure in the 1990s created a culture fostering specific systems, resulting in a lack of information sharing, a lack of ability to aggregate information across systems, and an inability to retrieve data to answer questions. An external review identified the need for EPA to improve information management, characterising EPA as "data rich and information poor".

Their data registry has been crucial in improving this situation. The EPA finds that organised metadata promotes better data management. The EPA registry effort began in 1996, so they now have a substantial body of experience. They have managed to gradually change culture, building the data registry into their business processes. The EPA has not been able to measure the amount of money that the registry has saved, but it has brought significant benefits including:

- reduction in cost from reinvention of information components
- ensures greater interoperability of systems and improved information exchange
- supports EPA's enterprise architecture
- reduces cost of making information available to the public
- reduces cost of maintaining inventories in multiple locations
- improves reporting consistency

The EPA continues to actively sponsor registry research and standardisation because of the benefits achieved. The EPA registries can be viewed at www.epa.gov/sor.

5.3 Outcomes

The individual benefits discussed should lead to a significant outcome for the business context in which the registry is applied. For the Highways Agency, the cumulative effect should be to allow work across various development "silos" to lower costs of system development and maintenance, and improve information services and management reporting.

For the wider ITS community, the cumulative effect should be to support seamless door-to-door services through integration of open systems. Without a registry this goal will be achieved later and at greater cost, as various organisations slowly find out how to integrate fragments of the overall service.

5.4 Cost / Benefit

5.4.1 Sources of funding

It remains unlikely that alternative sources of registry funding would be found in the short-to-medium term. The registry web usage statistics and trends are encouraging for the achievement of the benefits listed above, but are not at the level which would attract any significant advertising revenue. Charge per use is thought to be counter-productive in cost-benefit terms, at least until the registry is very firmly established in the business process of the user organisations.

The TIH Business Working Group were invited to contribute in this area but were not able to respond.

5.4.2 Feasible cost/benefit scenarios

(i) Submitter cost/benefit

A typical submission requires very little effort because the registrar performs the necessary translation and registry annotation work. The value of new use of the data should certainly be more significant – either in revenue for private information providers, or indirect benefit for public sector systems. In addition it is feasible that the quality improvement saves several hours of response to users' queries.

(ii) Central cost/benefit

The central funding required for the submission, registration and stewarding of a major submission was noted to be typically a few days of effort. The avoidance of even a single misinterpretation of data model could save a similar amount. The cost can also be recouped simply through a number of users each saving hours of further research into data models. The value of new users attracted through the directory service may be less direct, but may have a significant overall economic worth. The biggest benefits will be when integration is eased by adherence to registered models that have been made fit for new purposes through participation in the registry process. Even a single re-use and subsequent integration has the potential to save the cost of a submission and its subsequent upgrades.

Each of these benefits has the potential to balance costs on their own. When considered together, the business case is strong.

6 Recommendation

The ITS data registry has established benefits in improving quality, in efficiently answering data questions, in joining up separate developments, and in harmonising data definitions, which should lead to cost savings in system development. The Highways Agency's Information Directorate see the ITS data registry as fulfilling needs that are closely related to their own requirements to improve data management and to provide business intelligence reporting facilities. The planned system architecture for Information Directorate's business intelligence framework includes a metadata repository component that can re-use the approach and technology of the ITS data registry.

This report therefore makes two recommendations:

1. **Apply the data registry experience, approach and technology to the Highways Agency's internal data management needs.**

This will allow the Information Directorate to begin to catalogue data within the organisation, identify duplication of functionality and data to facilitate consolidation with the aim of increasing operational efficiency, and facilitate delivery of business intelligence solutions by providing accurate information on source systems.

This effort should begin with the creation of a prototype metadata repository implementing Information Directorate's specific requirements. Where appropriate the prototype should re-use the approach and the technology of the ITS data registry. The Highways Agency should then use the experience to evaluate the next steps for the metadata repository within an overall business intelligence framework.

2. **Operate, improve and exploit the ITS Data Registry for the Highways Agency and its partners in the Travel Information Highway community**

The Highways Agency should continue to operate and exploit the ITS data registry service to underpin discovery and harmonisation of data definitions and interfaces, resulting in improved efficiency in data exchanges between the Highways Agency and its partner organisations. The registry should be used as a resource to support data standardisation in areas of interest to the Highways Agency. The registrar should also continue to share knowledge with other registry implementers and researchers in order to improve the registry service and to protect investment by aligning with standards.

A further operational period of one year is recommended. During this period the results of the internal application of registry technology should be analysed and exploited for the TIH context where appropriate. At the end of the period, the project should deliver a management report with a summary of the benefits achieved and a recommendation on how the ITS registry should be managed in future given the progress achieved in the same period with the internal repository and business intelligence framework.

Appendix A References

1. "ITS Data Registry: Report from Operational Trials", March 2005, MM ref 203429/TD/03. Delivered to Highways Agency, March 2005, also available from www.itsregistry.org.uk.

2. Technical note "Legacy Harmonisation", Ian Cornwell, 22/12/05, MM ref 221939/TN/022v2

3. Technical note "Traffic Measurement Models", Ian Cornwell, 20/1/06, MM ref 221939/TN/027

4. Technical note "Exploiting Core Components", Ian Cornwell, 9/03/06, MM ref 221939/TN/29.

5. Technical note "Representing VMS – Registry Analysis", Ian Cornwell, 30/06/06, MM ref 221939/TN/33.