A new data, document and geospatial repository: Knowledge base and project support for a major international railway project

Marcus. R. Wigan[1]
Professorial Fellow, Australasian Centre of Governance and Management of Urban Transport, Faculty of Architecture, The University of Melbourne, Victoria Australia
Phone +61 3 94599671
Fax +61 3 9459 8663
Email oxsys@optusnet.com.au

Robert Kukla
Transport Research Institute
Napier University,
Sighthill,
Edinburgh EH11 4BN Scotland
Phone +44 131 455 2738
Email r.kukla@napier.ac.uk

Michiel Benjamins
Demis NL, Rotterdamse weg 183C
2629 HD Delft, The Netherlands
Phone: +31-15-2682555
Fax: +31-15-2682507
Email: michielb@demis.nl

Poul Grashoff
Demis NL, Rotterdamse weg 183C
2629 HD Delft, The Netherlands
Phone: +31-15-2682510
Fax: +31-15-2682507
Email: poulg@demis.nl

ABSTRACT
This paper presents a knowledge management, knowledge base and group support tool system (the Reorient Knowledge Base or RKB) developed by Napier University and Demis NL to provide both support during the Reorient project and knowledge transfer capacities at the end. REORIENT is a major multinational European Commission-funded railway integration project supported by the RKB through a website (www.reorient.org.uk). It comprises a full document repository, data repository, analysis, communication and geospatial exploration system.

The relevance to a repository conference is that this was developed entirely by the research endusers due to the dearth of library science tools for this integrated project management, data, document and mapping framework. At the present conference it is hoped that this will encourage libraray science practitioners to involve researchers in a more meaningful manner, and avoid the need for heavy handed 'mandates' etc, as the RKB has shown that the RKB approach has led to difficult decisions as to what to leave OUT of the final repository. rather than any difficulties in getting it in..

The design objectives, and the rationale for the inclusion of specific capabilities are

---

[1] Emeritus Professor Napier University

discussed, and some of the novel aspects presented in greater detail. The key components are document management and searching, where a full SGML/Z39.50 text database is used; data management and manipulation, where the NESSTAR (www.nesstar.com) federated data base publishing system is integrated, thematic map and multilayer exploration of datasets using both NESSTAR facilities and tailored extension to the Demis global Mapserver (www.demis.nl), and Marratech (www.marratech.com), a commercial video conferencing system. The latter is the only resource not accessed through the REORIENT website, but shares the same consistent layered security management. The RKB has proved to be progressively more useful as the project proceeds, and has already been used to deliver an initial managed knowledge base to the European Commission and several other visually different simultaneous KBs, including a full document repository. Knowledge Bases such as the RKB provide an effective means of supporting international and global cooperation within projects as well as an effective and usable means of knowledge transfer at the end of the project. However, due to its design and generalisability, its capacity already far exceeds the needs of a single project and offers a basis for a considerably extended range of data.

The simplicity of the basic design has proved to be robust to the changing and varied needs of different 'clients', but the complexity of the intellectual property issues engendered by the mix of commercial, open source and hand rolled software has proved to be an insuperable barrier to its direct continuation. It has however demonstrated that such a wide ranging repository and communication system, can be developed in a short time at economic cost (1.6 man years of technical input, 1.0 of project management including Eu project communication, participation and management overheads.

INTRODUCTION

Knowledge is a key part of all research projects, but is not always delivered in a format that can be used other than a report, and sometimes a dataset, in some form. Different disciplines have different conventions. In psychology it is essential to retain the original data to permit later reassessment or revalidation, albeit in confidence, but in many areas involved in transportation the data is still not regarded as an intrinsic component of required archiving and reporting. The necessity to keep detailed documentation on the data collection, definitions, instructions, processing and other factors are also less than universally acknowledged or undertaken – and the emergent requirement to formally specify the details on each item of data (metadata) is still a tentative and less than universally recognized need. This is spite of decades of experience attempting to make use of secondary data collected by others, where such information has neither been stored not kept collected with the data itself.

This (in a document sense) is where the document repository movement has been aiming, and the simultaneous development of Open Source and Open Access Green and Gold has married the need to make such systems possible, shareable and the contents legally available, in at least a metadata format. The rapid growth in the use of Z30.50 clients by end users (notably EndNote) has simplified the data discovery task and the barriers to entry of addressing an unfamiliar interface, while making all materials harvested via EndNote(for example) immediately usable in referencing, bibliography creation and library sharing.

In 2004 the European Commission (EC) Directorate General of Transport and Energy (DGTREN) initiated a triumvirate of projects on the issues of European railway integration under the Framework 6 series of research programs. The consortia were TREND, New Opera and REORIENT. The Knowledge Base described in the present paper is the one proposed by Napier University Edinburgh in the successful bid by the REORIENT consortium. REOREINT is aimed at determining and assessing the barriers to European railway integration along two major international movement corridors, and building the basis for business cases for subsequent operational projects. The REORIENT project was completed in

May 2007, and the Knowledge Base system has been in operation within the project since early 2006, as it developed. It will provide a flexible and sophisticated basis for knowledge transfer at the end of the project, and subsequently to encompass a much wider range of projects if supported in some form as a live system. Requests to consider constructing this system in some form as a commercial service subsequent to the completion of REORIENT were received from the Commission as a result of the mid-project seminar in Brussels in June 2006 showing where the development had got to by that time, with only 1500 documents and about a Gigabyte of data in the RKB at that point.

By the time the project was completed several times this initial amount was held, and after the review of the materials held a small selection was delivered as a complete project repository at the end of the project, with all publishable documents and data included. The much larger body of materials collated, used, updated and integrated during the project progress were also dumped into a browser-navigable complete reinstatement binary and readable dvd series with all the open source software created to develop the system. This did not form part of the released materials but is held in a readily accessible format in confidence. Consequently both the formal document repository and the security backups were simply an output of the RKB, with no special action required or mandate necessary

The RKB includes several major software systems and was based on a multiserver architecture and strict interface standards. The key components are document management and searching, where a full SGML/Z39.50 text database is used; data management and manipulation, where the NESSTAR (www.nesstar.com) federated data base publishing system is integrated, thematic map and multilayer exploration of datasets using both NESSTAR facilities and tailored extensions to the Demis global Mapserver and interactive network data collection system (Internetter) (www.demis.nl), and Marratech (www.marratech.com), a commercial video conferencing system. The latter is the only resource not accessed directly through the REORIENT website, but shares the same consistent layered security management.

The RKB has proved to be progressively more useful as the project proceeds, and has already been used to deliver an initial managed knowledge base to the European Commission in a static format on DVD. A compete document repository for evaluation by the Association for European Transport (www.aet.reorient.org.uk), and a full range of facilities for the EC Framework project, Worldnet (www.worldnetproject.eu) where the web-2 style on line editing of vector data of freight movement was critical, as was the thematic mapping tools and interactive access provided in the ReOrient project itself. All these are running as demonstration exercises on the same ReOrient software infrastructure base (usually referred to as the Napier Knowledge Based System Framework) base, through agreements with the SAIC Teratext supplier.

.   Knowledge Bases such as the RKB provide an effective means of supporting international and global cooperation within projects as well as an effective and usable means of knowledge transfer at the end of the project. However, due to its design and generalisability, its capacity already far exceeds the needs of a single project and offers a basis for a considerably extended range of data

OBJECTIVES OF THE KNOWLEDGE BASE

The RKB was designed to bring together documents, data group support tools under a single framework. It was designed to support the operations and communication needed with in the REORIENT project as it developed, as well as to deliver a fully populated final Knowledge Base at the end of the project. It was always envisaged that provision would probably be required for continuing access to the live RKB software system and that support for this would probably be forthcoming if this part of REORIENT was found to be effective and

meeting a real EC or related project need.

The framework of the RKB was designed to support a wide range of projects, not simply REORIENT, and to ensure that the knowledge created was both accessible and usable. This aspect was attractive to the authors at Napier University Edinburgh as a possible key networked research support and communication tool for the University when populated with very different data and documents.

DESIGN PRINCIPLES

A well tested principle in software systems is that if users want a system to succeed it generally will, and if they do not there is a hightened probability of failure. The RKB was therefore designed to offer the simplest possible interface to users, and to require no special software at the client end. This proved to be possible for all aspects other than the commercial video conferencing system, which has to be run separately.

The range of facilities offered are summarised in Figure 1, with clips from the sidebar menus illustrating the full range of facility areas. The key item to note on Figure 1 is the coloured dots against each facility access point: they are coded for the security level of the materials.

These drop down menus start as minimal lists of one or two items for public access to ww.reorient.org.uk (the website hosting the RKB), up to the full range shown here for the administrators. The colour code is shown as a band across the top of the website screen as a subtle confirmation of the security level at which one is logged in. This colour code is also used to flag the security level of any documents retrieved by the system when searches are done.

Considerable effort has been spent on making the power of the underlying software as nonthreatening and uncomplicated as possible, and the feedback from the REOREINT team has been invaluable.

"Documents"

The document aspects of the RKB permit extensive metadata handling and searches at a very high level of complexity (the underlying engine, SAIC's TeraText, is in extensive use by intelligence agencies across the planet, as well as to drive major legal and other sites), but these facilities are simplified in presentation by several levels of search, including menu driven metadata searches, to keep the access process straightforward.

The document system does not include a 'folder' model (and indeed modern operating systems are increasingly moving to a 'search' model instead of the of folder paradigm), but users still find this unfamiliar, so a pseudo folder view has been built in both to the uploading system (using WebDav for security) and to review the document database and the results of any search.

The 'document' system also includes access to large databases (usually in a zip format), accessed via the metadata (which include files names sources etc). The results of any search can be downloaded to the users machine.

The document system also includes all draft documents, and acts as a document management system as security level access to each level of user permits drafts, secure, and public documents to reside in the same document database.

Data

The data is held in a federated database, NESSTAR (www.nesstar.com), and delivered in the webpublisher component of NESSTAR through the RKB website interface. NESSTAR is contains both data - and the related documents as metadata, as well as formal metadata for each data element. The important aspect here is that data may be held – and the same data –

in three different ways in the RKB.\

1. As a binary object (.zip file), retrieved in bulk as a single unit from the document database – and multiple versions can be held and retrieved in that framework
2. As a fully examinable set of tables (a datacube) in Nesstar, with access to individual elements, and potential linkage across the different databases held in NESSATR, which may be held in completely different geographical locations. Data analysis is also possible in NESSTAR (table generation, regressions etc), and thematic mapping of the layer being examined at any specific point, and downloads of tables and other views of the data.
3. As the basis of a Demis Mapserver layered thematic map display, where the layers can be switched on and off, the map zoomed etc

This illustrates several of the multiple aspects of the RKB: archival access to bulk data, full access to all versions of documents, access to data with analysis and download capacity, and exploration using visual and mapping tools.

The RKB also contains an online network oriented data capture system, Internetter from Demis, which by implication enforces some data collection standards

Collaboration

Collaboration is a key aspect of all projects, and the RKB provided Bulletin Boards, WIKIs, mailing lists and remailers within the website framework, and a third party collaboration support video/audio conferencing tool.

The Bulletin Board was simply not used at all, and has been dropped from the RKB. However the WIKI (a user-editable website) linked to a calendar has proved to be effective. This was initially specified in the original proposal in 2002, and was then regarded as a marginal element. Since then the recognition of WIKIS as a key component of the far more interactive Web-2 family of approaches has made its inclusion more easily recognisable.

The mailing management system is a small but extremely useful aspect of the overall system. The mailing lists are drawn from the user/registered user/partner/etc hierarchy of RKB users, and form a central reliable communication point whereby mailing the appropriate listname @ reorient.org.uk assures accurate and updated communications via email. The use of such a remailer also has ensured that spam has been kept to an extremely low level in all such communications.

Communication

Knowledge base systems need not only to hold data and documents, provide extensive search facilities, but also need to support exploration and understanding of the information held.

This aspect of data exploration is not always addressed in transportation systems, but is slowly becoming important in Data Observatories (Wigan 2006) where large administrative units valiantly try to integrate their data holdings to make them mutually accessible across many different agencies with in their domain. - and also make them available to their increasingly sophisticated publics, although still often as an afterthought.

The document facilities are not restricted to the simple 'google' approaches, and include full Boolean searches and more. But the key area where confusion often reigns is when an item of interest such as a project consortium name is a common word. Here we have TREND as a consortium name. Or an html component of a filename of a document pointing to a .zipped database in the search engine? How does one extract that from a document search engine?

The RKB answer is to add metadata elements to the construction of the Boolean search (see Figure 2). TREND is a collective author or keyword in the document metadata

and so can be picked out easily. This is a single simple example of the power of a more sophisticated search engine on document databases.

INTEGRATING SOFTWARE COMPONENTS INTO THE RKB WEB PORTAL

One of the technical challenges of the REORIENT project was to integrate highly specialised software applications (e.g. GIS software) with more off-the-shelf products and open source software under a uniform look and feel to make them all accessible to the target audience.

Hardware Platform

The software products targeted for use within the project use very different technologies and programming languages and individual components would only run on either MS-Windows or Linux platforms which made it necessary to have at least two server machines running different OS. Unfortunately this incurs additional maintenance cost, which however can be partly offset by running the most suitable and best priced software.

It is important to have reliable hardware base and provisions for hot-standby in case of a failure. Two branded server machines where purchased for deployment at the main site in Edinburgh, while an additional machine was situated at an off-site location for mirroring purposes. Each server is a 3.6GHz Xeon with 4GB of RAM and 2x250GB of hard disk space (RAID1). A fast connection to the internet is established by a direct (redundant) 1Gigabyte link.

Software Framework

The main front end is written with in php. It generates all web pages dynamically based on data in a MySql [http://www.mysql.com/] database and store on the local file system. It handles user authentication (see below) and acts as glue between the different applications.

The generated HTML follows the standard "v4.01 strict" of the W3 consortium for maximum compatibility with existing web browsers. Extensive utilisation of cascading style sheets (CSS) allows easy re-formatting of pages e.g. for printing or the improved accessibility. URL-rewriting is used to hide parameter passing and create virtual directories for URLs in order to "encourage" search engine providers to index the site. Menu items are kept in database and can point to internal and external resources.

Architecture

Fig. 4 shows an overview on how the software components are distributed between the to server machines. In addition to the web services via secure SSL (port 443 https) and the encrypted multi media streams for Marratech, there is also secure communication with a webserver. All internal communication is secured behind a firewall.

Security Requirements

The REORIENT website has two related tasks: support knowledge accusation and dissemination. It is to act as a publicity portal to inform interested parties about the goals, the progress and the result of the project and the possibly to get involved and use results of the work. But there is also a requirement to stimulate the co-operation between the partners by providing a platform for communication, data exchange and management – things typically more associated with an intranet environment.

To strictly separate confidential and public data a hierarchical system of access control based on "access level" was implemented. The access level is assigned to a user on

account creation and can only be changed by the Sysadmin. Conversely web Pages, data sets and documents have a "page level", which is assigned at creation by creator, can be change by owner. If a user with an access level < page level tries to access a resource, then access is denied by the server. Currently 6 levels exist (expandable on requirement): general public, registered public, subcontractors, partners, work package leader council (WPC), Sysadmin).

Access is inclusive of lower levels, so partners will be able to get to subcontractor related data but not vice versa. The MySql database maintains the access restrictions for users and resources. Colour coding is used to visually represent the different access levels. All traffic is encrypted via SSL (secure socket layer) and Cookie based authentication is used to make man in the middle attacks impossible and eavesdrop attacks impossible. Comprehensive logging is implemented and activated for intruder detection and activity review purposes.

Integration of Open Source software

Whenever possible open source software was used; mainly to keep the cost down, but also because due to its nature (availability of the source code) it makes it easy to adapt and integrate. The *Erfurt Wiki* [http://erfurtwiki.sourceforge.net/ewiki] was used as the WIKI software as it is implemented in php and only needed a few adjustments with regards to calling conventions and to utilise the central user account database.

Webalizer [http://www.mrunix.net/webalizer/] is flexible web log analyser. It works off-line in regular intervals and the resulting tables and diagrams can seamlessly integrated into the portal.

The *JavaScript Form* Validator [http://www.javascript-coder.com/] is used for client side form input validation. Embedding the required rules in dynamically generated forms, guarantees the user entry errors can be caught instantaneously and be corrected by the user.

Marratech and Nesstar

As both are closed source applications complete seamless integration was impossible. However by establishing a semi-automated way to transfer user-authentication information it is possible to keep the user databases in sync and use the same passwords across the different applications.

Nesstar [http://www.nesstar.com/] provides limited integration restricted to "branding". Using the standard REORIENT header means the user will be confident that he is still navigating the REORIENT website, even though the layout differs slightly.

Marratech [http://www.marratech.com/] provide a stand alone client which can be downloaded from their website. The REORIENT website provides a short tutorial about its use and specific project related aspects.

Demis Mapserver and interactive Maps

The Demis [http://www.demis.nl] GIS system consists of a (map-) server and individually downloadable flash clients for different application areas. The web server runs independently on the windows server machine. Flash can be seamlessly embedded into html pages and consequently the integration consists of generating the required tags and parameters to display the maps as integral part of the REORIENT website. Once the Flash client is displayed, further interaction takes place in the client area.

Teratext
  – High performance XML native database

- Implements the Z39.50 protocol (open systems ISO and ANSI protocol for information retrieval)
- Support for different document formats via comprehensive import filters
- Incorporates a high performance, multi-threaded web server
- Powerful Ace language can be used to build Web applications
- Flexibility comes at a price – as a front end needs to be written
- Hybrid of php for authentication & embedding into portal and Ace for database access.

The general functions of Teratext from a user point of view are covered earlier in this paper

INTERNETTER: ONLINE COLLABORATIVE NETWORK EDITING TOOL

While the various elements of the RKB are all of some real interest to transportation professionals, the example chosen for more detailed discussion as an illustration of the RKB facilities is restricted to the tools tailored by one of the REORIENT partners (Demis) specifically for this system.

For the network data collection process in REORIENT an innovative approach was chosen. It consists of an online collaboration model and tool that allows registered users to edit the REORIENT multimodal road, rail and sea link network and attribute data. The result can be described as a sort of WIKI for online GIS network editing. The background for this approach comes from the fact that in the REORIENT project subcontractors in many countries need to enter network related data and edit the network topology.

A classic approach with stand-alone GIS applications might lead to problems in combining the country-specific partial networks, as well as to the problem of not having an integrated complete network available until the last subcontractor finished its work.

A collaborative online approach as used in REORIENT has many advantages, such as:
- Using a web based tool ensures that all locally edited data is stored in one central database with a consistent network topology for the whole network of Europe.
- Simultaneous editing from distributed locations is possible.
- The latest complete network version is always available by way of regular exports to standard GIS formats
- No special client software is needed to edit the network, a simple browser with a Macromedia Flash player installed is enough.
- No (high) license fees for GIS software are needed.

This approach in REORIENT was made possible as Demis already had developed an earlier version of Internetter, used in large volunteer projects for editing bicycle network data for the Dutch cyclists union ENFB. Based on the collected bicycle network data the ENFB and Demis launched door-to-door route finders for bicycles for the provinces Utrecht and South Holland in the Netherlands.

For the REORIENT project Internetter was expanded to cover multimodal network editing and advanced editing and user management options. It has since been adjusted to provide worldwide interactive data input for a major freight project in the EC (worldnet)., but on this later development the mapserver resides on the Dutch demsi base server, and not on the Reorient server in Scotland. This change is invisible, and is possible die to the essential multiserver architecture of the basic reorient KB design.

The client-server approach

Internetter is based on a client-server approach. The Internetter Server consists of a knowledge layer for storing and manipulating multimodal network vector and scalar data, a user manager and a Dijkstra routing algorithm for route finding. The Internetter Server was

developed in VB.NET and ASP.NET. The Macromedia Flash 8 development environment was chosen for building the client application user interface. An overview of the Client-Server approach is given in Figure 5.

When the web browser requests via the HyperText Transfer Protocol (HTTP) the Flash file (in SWF format), it will be played on the Flash client application by the Flash player. The first step then is the log-in procedure, which is used to ensure that only registered users can have access to the network data. The second step is to set-up a persistent two-way communication between server and the Flash client application using the Flash Real-Time Message Protocol (RTMP). This allows an uninterrupted two-way data stream to flow between client and server. This data flow is used to send network data from the server to the client application and when the user has edited the network topology or the data, then the updated network is send back to the server.

As can be seen in Figure 5 the background topographical maps are retrieved using the OpenGIS protocol using the Demis Web Map Server (WMS). The Demis WMS is an OpenGIS certified WMS. Of course any other certified WMS can also be used instead.

Viewing network data using Internetter

The first main functionality of Internetter is to be able to visualize and navigate through the network data. What can be visualized depends on how far the user has zoomed in. The methods of visualization that Internetter offers are:

1. Thematic coloring of network links for a chosen link attribute
2. Varying the thickness of links based on the value of a chosen link attribute
3. Add text to a network links that shows the value of a chosen link attribute
4. Hide network links based on the value of a chosen link attribute
5. Thematic coloring of network nodes for a chosen node attribute
6. Varying size of network nodes for a chosen node attribute
7. Hide network links based on the value of a chosen link attribute

As combinations of the above mentioned functionality are also possible, this gives the user many possibilities in effectively presenting network data. Fig. 6 shows an example within the RKB website.

Editing network data using Internetter

Internetter uses an Edit toolbar to expose the edit functionality to the user (see the yellow toolbar in Fig. 6). The Edit toolbar contains fourteen buttons that all have a specific functionality. The following general editing functions are present (buttons from left to right):

1. **Add a new network link**: this allows the user to add a new link in the rail, road, water or sea network. The network link can be added between existing nodes but can also be created between link points or unattached to the current network. In the latter two cases the network nodes are automatically created;
2. **Add a new network node**: this allows the user to add a new node. When clicking on a network link point, it will add a node to the link by replacing the link point by a node;
3. **Add a new network link point**: this allows the user to add new points on a network link so the trajectory of the network link can be adjusted;
4. **Move a network node- or link point**: this allows the user to move a node or link point by clicking, moving and releasing it;
5. **Modify attributes of a network link- or node**: this allows the user to modify all attributes of a single network node- or link;
6. **Modify one specific attribute of one or more network links** : this allows the user to modify one specific attributes for multiple network links;

7. **Modify one specific attribute of one or more network nodes**: this allows the user to modify one specific attributes for multiple network nodes;
8. **Remove a network node and join the remaining network links**: this allows the user to safely remove a node without harming the network links that are attached. This can only be done if the attributes of the links are identical;
9. **Remove a network link, -node or –link point**: this allows the user to remove a link, node or link point. In the case of a node, the adjoining links are deleted as well!
10. **Copy attribute values**: this allows the user to copy (and later paste) the attributes a single link or node;
11. **Paste attribute values**: this allows the user to paste the attributes of a single node or link from which the values were copied.
12. **Undo** last change
13. **Redo** last change that was undone
14. **Unlock** network objects with changes, so other editors can work on these objects

These functions mean that for all practical purposes Internetter is an online geographic Information System (GIS) for editing point and link data in a network topology. For ease of use in editing larger networks for a long list of attributes special functions were added that allow the user to copy and paste attribute data from one network object to another, as well as setting a value for a specific attribute for all selected network objects at once. And Undo / Redo functions were added to handle accidental mistakes. Fig. 6 also shows how the attribute data can be edited. Adding a new network link is a two step process, first a new link is made then the trajectory is added. This is shown in Fig. 7.

User management
In a collaborative tool user management is important to ensure high quality work. First of all in REORIENT only registered users are given editing rights. Secondly each action by the user is logged. In cases where the action leads to obvious errors, the system administrator can roll back the change(s) made by that user. In the user management overviews (thematic maps) of who last changed any object for a specific attribute can be generated. This makes it easier to spot erroneous data and who's responsible for that.

DEMIS WMS: A THEMATIC MAPPING TOOL FOR TIME SERIES DATA
For Internet based GIS and mapping applications, Demis has developed the Demis Web Map Server (Demis WMS, see <u>also http://www.demis.nl/home/pages/wms/demiswms.htm</u>) in compliance with standards and protocols defined by the OpenGIS consortium. The OpenGIS Web Map Server (WMS) protocol defines a simple interface for web based mapping applications. The WMS protocol is based on a simple query syntax for posting a request for the desired layers and zoom window to the server, which returns a map as a standard picture (GIF, PNG or other format).

The Demis Web Map Server runs on Microsoft Windows NT/2000/2003 server editions with IIS 4 and higher. The Demis WMS service components are Microsoft Component Object Model (COM) based, WMS connector scripts are provided for integration with ASP 2.0 or ASP.NET web programming environments

A unique feature of the Demis WMS is its support for time series data, in the form of animated thematic maps and times series charts. In contrast with popular Geographic Information Systems (GIS), the underlying map engine of the Demis WMS has been designed with time as an important feature, since simulation models often run over long time series (multiple years). For example: it is possible to create animations in the magnitude of the transport on a network link is shown using a color and/or thickness of the branch.

In REORIENT the Demis WMS is used for a variety of thematic maps, including the network data in the REORIENT corridor and EUROSTAT time series data per country or per point of interest. Fig.8 is an example of one of these maps. The implementation is based on Flash client application and the Demis Web Map Server. The Flash client application allows the user to select a data time series, to animate the time series and to navigate in the map.

The ReOrient Kb contains many more facilities (2) (such as open journal publishing and automatic archiving and version updating) but these are the primary ones for a repository focus, as distinct from one of research support.

METADATA GENERATION

End user input of metadata- even with extensive support within the system – proved to be inadequate in the ReOrient system. When one of the two demonstration/extension trials were added (www.aet.reorient.org.uk) the near complete holdings of the European Transport Conference comprising over 2000 papers and 2.7Gb of Acrobat files, it was clear that there was a dire need for some form of automated support for metadata generation and input. The dteaisl are the subject of a separate paper to OA08 (1), but in summary the document engine used in the RKB (SAICs TeraText) was used to test the contents of two complete professional Thesaurai (one Australian/UK English, the other American) against the document set, and after pruning single entries and universal entries, this was then used to map keywords from the contents of each document in turn. This produced large numbers of keywords, which worked well in domain expert searching, but fell apart on methodology papers and searches as Transport Thesauri are acknowledged to be weak in this area. This technique offers opportunities for iterative use to improve specialised Thesauri.

LESSONS LEARNED RELEVANT TO REPOSITORY STRATEGIES

It is mildly surprising that, given the regular calls for mandated entry to repositories that so little was (and has yet) been done to work with the end users and creators of the materials, as documented here, as the problems reverse and become ones of trying to handle the volume of materials built up, rather than their delayed capture later on. It is also unfortunate to have to report that no cooperation or joint creation or adaptation of extant software resources from library science groups or practitioners proved to be possible, and so the RKB system had to be built from scratch in enforced isolation. Yet the RKB included OAI-PH, Z39.50, Endnote plugins RSS, WKI supports and integrated communications that were expected by the end users and overlap the technical requirements for mainstream library-based document repositories.

The simplicity of the basic design has proved to be robust to the changing and varied needs of different 'clients', but the complexity of the intellectual property issues engendered by the mix of commercial, open source and hand rolled software has proved to be an insuperable barrier to its direct continuation. It has however demonstrated that such a wide ranging repository and communication system, can be developed in a short time at economic cost (1.6 man years of technical input, 1.0 man years of project management including Eu project communication, participation and management overheads.

Our hope is that having built such an extensive and capable system, that this may encourage the participation of the repository library science groups to begin to work with the many such sophisticated end users directly. Not only will this banish the "issue" of mandates deemed essential by so many libarary personnel (uneccessary if the approach encompasses the

research and communication processes rather than the archeological procedures often required to exhume materials after a project is dead and buried), but also will lead to far more rapid development of data and geospatial data observatories and live repositories. It should be noted that efforts by research staff to secure help in setting up their own specialized document set repositories are often rebuffed due to the cultural differences between these groups. Such cultural differences must not be discounted, as they are a key determinant in successful joint projects; It is an emergent requirement from experiences to date that it is essential that end users 'own' repositories in a real sense, rather than being subject to management *fiat* to use a system in which they have no influence or control or design stake.

The other stream from which inspiration can be drawn is the data observatory movement Examples include http://www.norfolkdata.net/ and a detailed discussion is given in (2). All these have a single major characteristic in common: they all aim to make data accessible in a structured as well as an unstructured (simple retrieval) mode.

This is but one step along an extremely important road, one not normally associated with document repositories, that is the direct enhancement of governance and accountability of government in the community. By making structured information as readily available between professional groups (the main aim of repositories to date) and the communities involved or affected (a target only barely discerned as yet) then the role of information in transparency and accountability is enhanced and plays an active and important part. These issues are addressed by the main author elsewhere, but are part of the process of not only scholarly communication (the primary library document repository focus), but also as an active part in community and governmental linkages and processes (3) . Once data and mapping forms part of a repository system, and one these links are live and readily accessible, then these potentials are realisable.

SUMMARY
Taking document and data management seriously is necessary in multinational projects, this paper describes the Reorient Knowledge Base, which addresses these increasingly common issues encountered by projects involving multiple organizations.

The relevance to a repository conference is that this was developed entirely by the research end users due to the dearth of library science tools for this integrated project management, data, document and mapping framework. At the present conference it is hoped that this will encourage library science practitioners to involve researchers in a more meaningful manner, and avoid the need for heavy handed 'mandates' etc, as the RKB has shown that the RKB approach has led to difficult decisions as to what to leave OUT of the final repository, rather than any difficulties in getting it in.

The RKB approach as described swiftly proved to be progressively more useful as the project proceeds, and has already been used to deliver an initial managed knowledge base to the European Commission and several other visually different simultaneous KBs, including a full document repository. Knowledge Bases such as the RKB provide an effective means of supporting international and global cooperation within projects as well as an effective and usable means of knowledge transfer at the end of the project. However, due to its design and generalisability, its capacity already far exceeds the needs of a single project and offers a basis for a considerably extended range of data.

ACKNOWLEDGEMENTS

REFERENCES

.
1. Wigan, M.R. and R. Kukla (2008 *The roles of thesauri, metadata and automated keyword generation in document repositories*, [This conference]
2. Wigan, M.R. (2005) *Presentation to the Data Stewardship session.* Transportation Research Board Annual General Meeting, Washington DC January.
3. Wigan, M.R. (2007) *Specification of the Reorient knowledge base system.* Reorient Final Conference, Brussels 31 May 2007. http://www.reorient.no/documents/07%2005-31%2ORKB_NKBS%20Confhandout.pdf
4. Wigan, M.R. (2007*) Contestable Evidence Based policy- the role of information sharing* . Australasian Centre for Governance and Management of Urban Transport. The University of Melbourne, Australia. (www.gamutcentre.org)

FIGURES

Fig. 1 Knowledge base functions available via the website
Fig. 2. Using metadata as one Boolean element to extract from a document database
Fig. 3 Some data analysis can be done within NESSTAR on the RKB website
Fig. 4 Software server architecture of the Reorient Knowledge Base
Fig. 5 The Internetter Client – Server approach
Fig. 6 Editing the attributes of a single node, the station of Platanos in Greece
Fig. 7 Two processes to adding a network link and its trajectory
Figure 8 Employees of Railway undertaking with time series chart for Poland

Fig. 1 Knowledge base functions available via the website

Fig. 2. Using metadata as one Boolean element to extract from a document database

Fig.3 Some data analysis can be done within NESSTAR on the RKB website

■ Each entry in this datacube display has a scroll bar

Fig. 4 Software server architecture of the Reorient Knowledge base

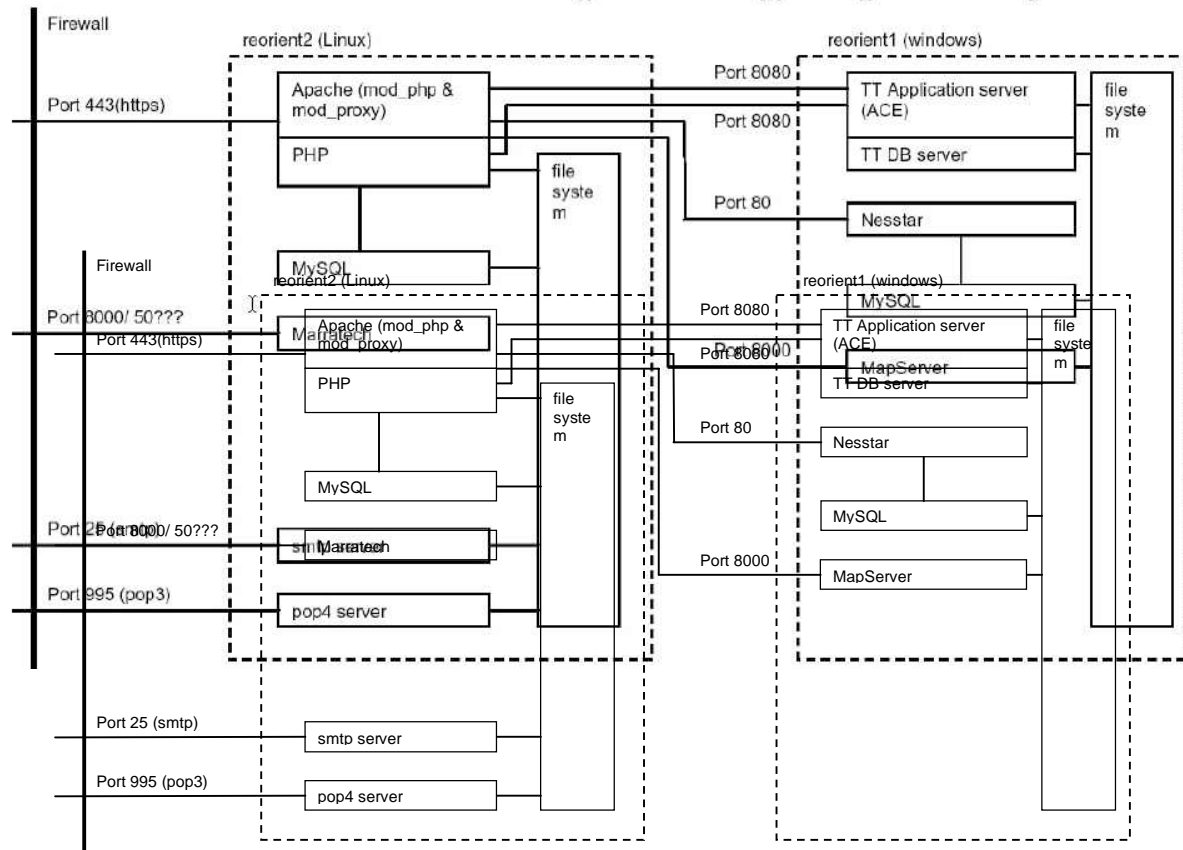*General back end server and hardware organisation supporting the R.KB system*
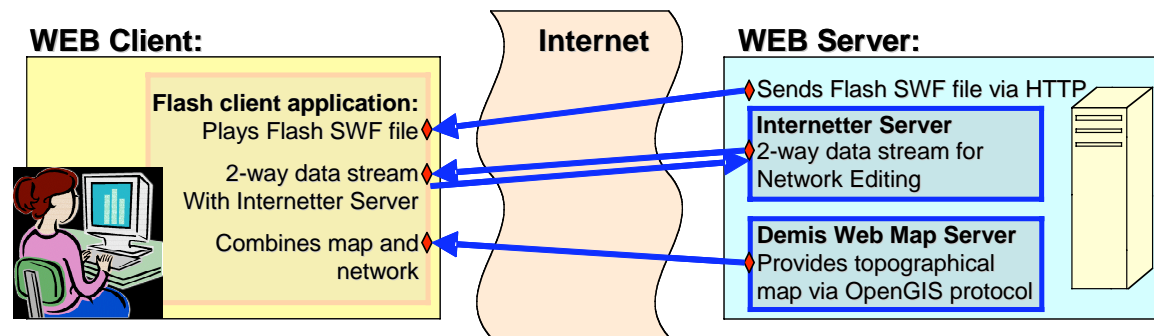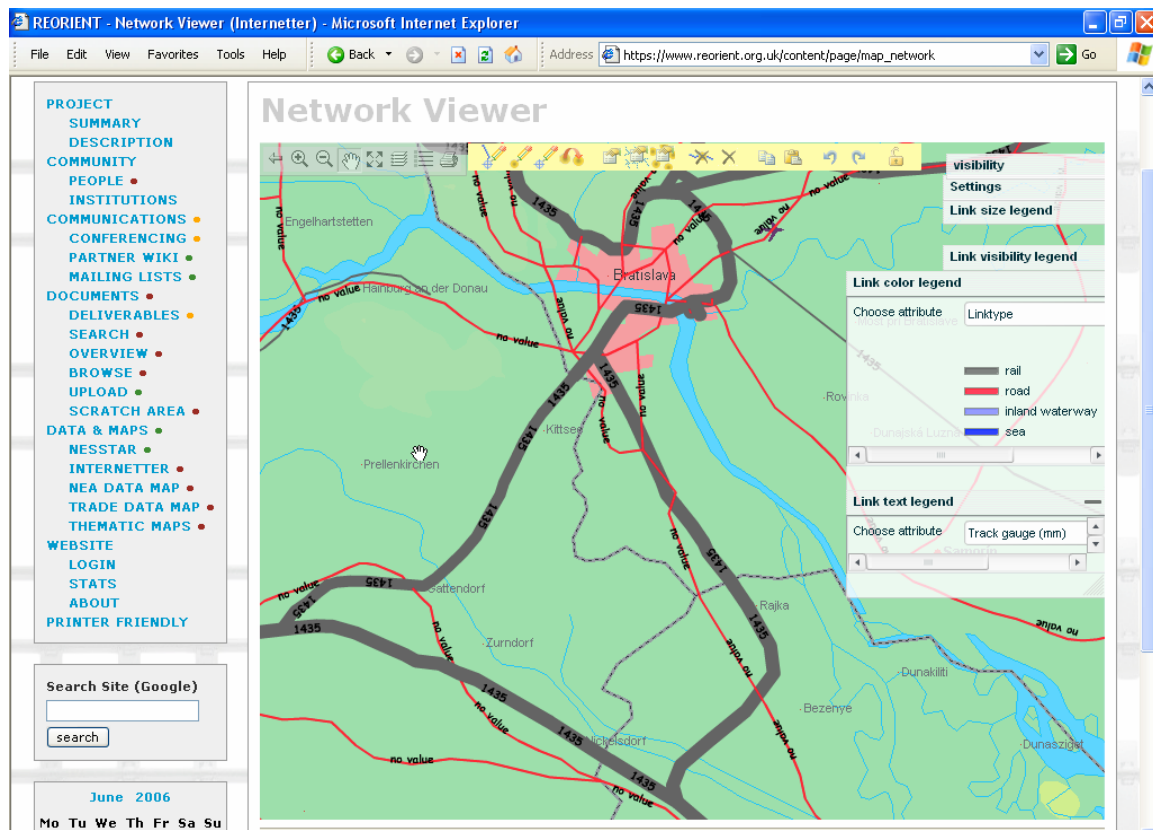
Fig. 5 The Internetter Client – Server approach

Fig. 6 Example where color shows link type, thickness corridor and text track gauge

Fig. 6 Internetter: editing the attributes of a single node, the station of Platanos in Greece
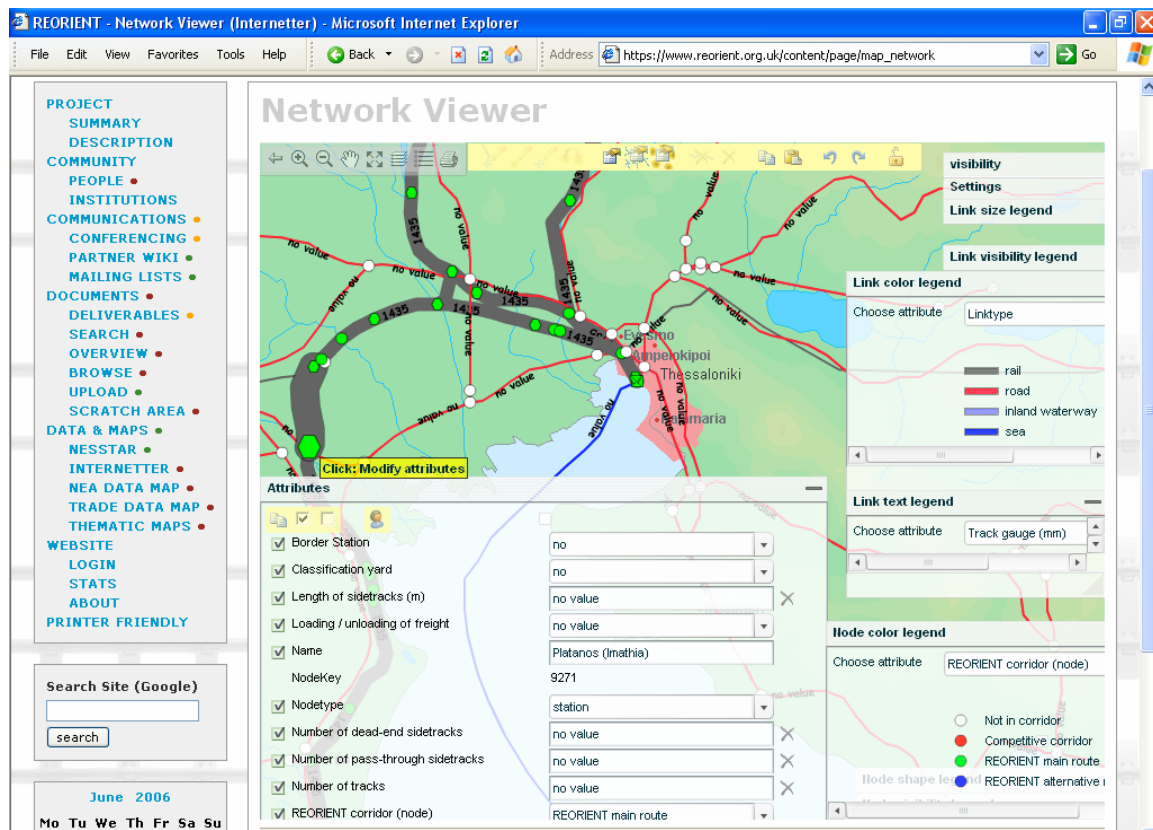
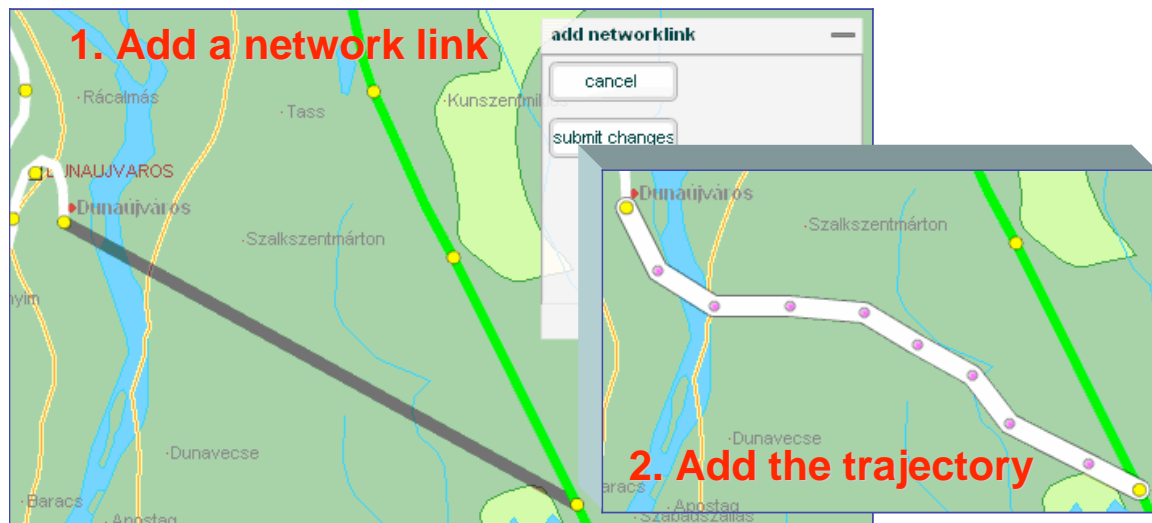Fig. 7 Two processes to adding a network link and its trajectory

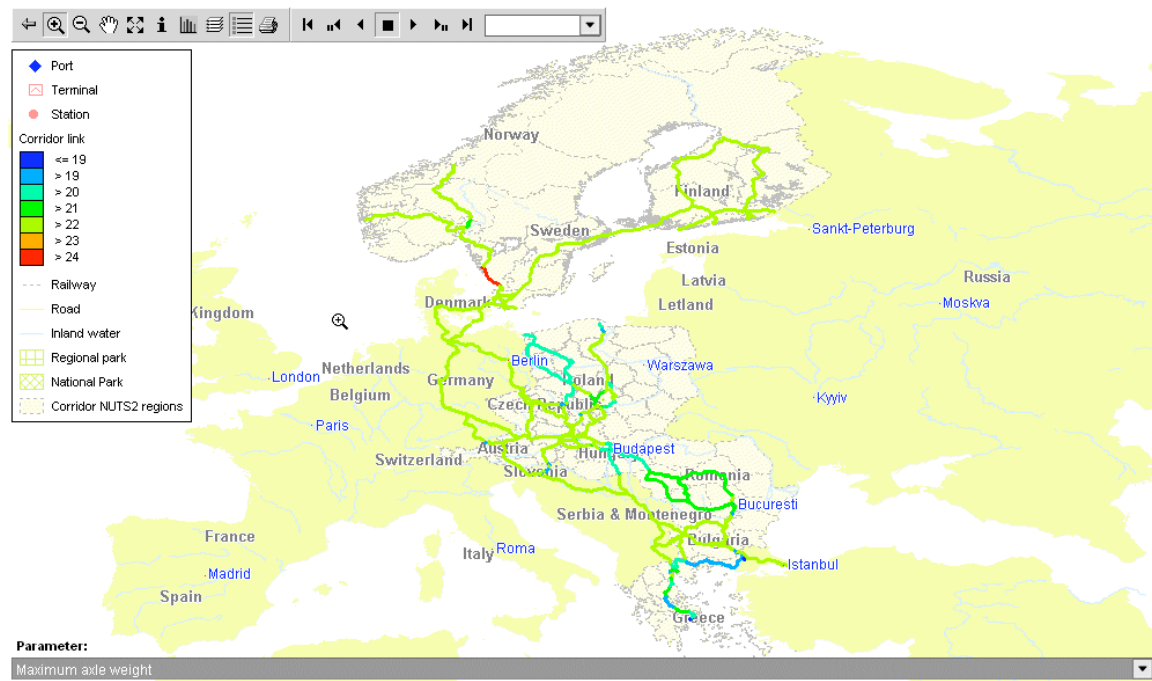Figure 10 Thematic map of axle weights on the REORIENT corridor

Figure 11 Employees of Railway undertaking with time series chart for Poland