CONNET - DESIGN ISSUES[1]

Issues in Designing a European Technology Transfer Network for the Construction Industry

ZIGA TURK, TOMO CEROVSEK
University of Ljubljana, Slovenia

AND

ROBERT AMOR
Building Research Establishment, United Kingdom.

ABSTRACT: Integrated information systems about construction products and information will become an important catalyst of the global construction market. In the frame of European Union's Technology Transfer Network such a one-stop-shop for the construction industry of Europe has been developed. During 1999, partners from the UK, Finland and Slovenia have been working on a pan-European virtual technology park on the Internet - CONNET. CONNET is designed as a set loosely coupled nodes, operated by the different partners in the project. The nodes are integrated on both a semantic and technical levels. The first integrates the information offered by the nodes by providing the mapping and translations between the meta-data and the classification systems. On the technical level, the services are integrated by providing user management and cross-node searching enabled by a publicly defined application interfaces. This paper presents aspects in the design and development of the system. We learned that in integrated, industry-wide information systems, several information schemata and classification systems must coexist; that they will not necessarily be based on standardized building product models and existing building product classification systems but on metadata standards developed outside the construction sector. We present an architecture of loosely coupled,

---

[1] published in: Heng Li, Shen, Q., Love, P.E.D. (editors). Implementing IT to obtain a competitive advantage in the 21st Century, INCITE 2000, The Hing Kong Polytechnic University, ISBN-962-367-272-1.

CONNET - design issuesINCITE 2000 / edited by Heng Li ... [et al.]. - Hong Kong : The Hong Kong Polytechnic University, 2000. - ISBN 962-367-272-1. - Pg. 416-428Ilustr.

federated services on the Web, which are flexible enough to provide a unified interface to various construction related data. The use of open standards for meta-information, Internet technology and object oriented design and analysis approach proved vital for the success of the project.

1. Introduction

Information technology (IT) is acknowledged to be a potential driving force behind Europe-wide improvements in quality, productivity and turnover for all sectors of the European economy. In particular, as common modes of operation and cross-border contracting develop in the construction industry, the need for easy access to Community-wide information is rising. Construction industry is generally regarded as being behind in its uptake of beneficial IT. CONNET therefore focuses on approaches and techniques that enable technology transfer to SMEs as well as provide new business opportunities and models.

The CONNET project provides the construction industry with an essential source of such information, by creating a "virtual technology park", accessible to the whole industry, regardless of national boundaries. CONNET forms an element of the European Union's Electronic Technology Transfer Network (ETTN) and focuses on the information needs of the construction industry.

Additionally, CONET is developing or customizing the technologies to provide virtual "park grounds" for the integration of new nodes, for example to provide a nationally specialized version of an existing services, or to introduce new services on a national or international levels.

Expected to officially launch by the end of 1999, the CONNET project has been building a virtual technology park using a number of services that already exist (though are not electronically accessible), or which are currently being developed by organizations located in EU member or associate member states. The services include (Figure 1):

- Technical Information Centre.
- Waste Exchange Centre.
- Manufactured Product Service.
- Calculation and Software Centre.
- Newspaper Service.

2. Requirements

The technology park envisioned in CONNET has a broad and diverse audience across Europe and beyond. The types of information that are being handled by CONNET as well as the information users and providers are extremely diversified. Therefore, the network had to be designed as a set of very loosely coupled nodes. The requirements can be broken down into several categories:

a) Common features of the nodes. Primary means of access to the node is the Internet. Public and free access to a substantial part of node's functionality are desirable.
b) Features that enable the nodes to function as a part of a wider technology transfer network and integrate them with central CONNET services.
c) Features which are node specific and do not influence node's cooperation with the rest of CONNET.

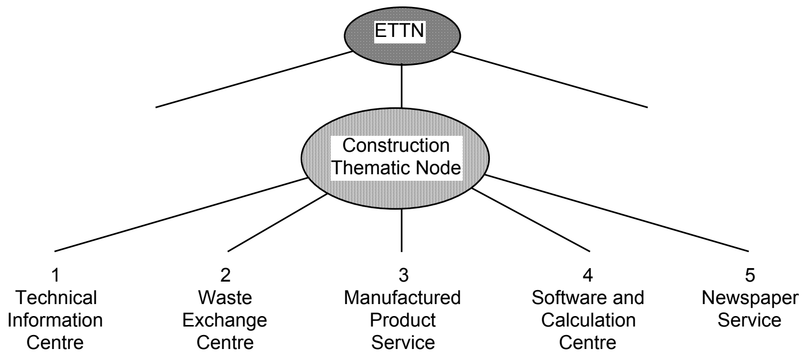

Figure 1. CONNET services are a part of the construction thematic node.

3. Business models

CONNET is not a unified company with a common business plan. Neither can one expect that all information needs for construction can be supplied by one company. Instead integrating several informaton providers with different business plans has been planned for. Each of the nodes can operate rather independently and follow its own business plan. On a very general level, however, the business models of the nodes are similar, because they have a function of a pre-sale information broker.

Typically, information offered by a CONNET node would be provided by a manufacturer of a construction product, a publisher of books, reports, or software or a provider of a specialized service on the Internet. The added value of a CONNET node is that it provides an efficient search mechanism, a uniform access to a rather complete set of information coming from different providers across Europe, some basic quality control, and, in some cases, offers electronic commerce services for finalizing the transaction between a customer and provider. Most CONNET nodes are therefore financed by the providers, either by giving commission of the sales made through CONNET, by advertising on CONNET sites or by reimbursing for the fact that the information abut their products is available on CONNET nodes. The added value of CONNET as a whole is that it provides a single point of entry for the different types of information, provides technology for the nodes and some central services, such as user profiling, help desks etc.

Alternative business models are also possible, for example, that CONNET nodes charge for the information they provide, as well as the mix of both models. Since there is very little experience in the e-commerce of construction related products and information, CONNET will be able to provide valuable insight into the future opportunities of the construction sector in this rapidly evolving market.

4. Baselines for interoperability

This section introduces the developed data models, classification systems and application protocol interfaces (APIs) for the CONNET thematic node as well as the initial five services. These data models and APIs provide the central representation required to implement the CONNET services. Through the replication of these data models and APIs new services of the same type as these five could be established anywhere in Europe, and work seamlessly with the CONNET central services.

To provide interoperability of the services, possibilities of some level of compatibility in the following areas were studied:

- **Product and meta-data**. Most of CONNET services provide information about information or about a product: about software, publication, manufactured product etc.
- **Classification**. Classifications ease the browsing in databases and enable establishing the relations between different types of information, for example there are books, software, reports and product data about reinforced concrete.
- **APIs**. Application interfaces of the services should be public so that the integration with other services is possible.

4.1 RELEVANT META-DATA STANDARDS

For information as diversified as CONNET, several standards for the product- and meta-data have been considered.

ISO 10303 (STEP). STEP is the abbreviation for the Standard for the Exchange of Product Model Data. It is defined by the International Standards Organisation (ISO) Technical Committee 184, Sub-Committee 4.

IAI-IFC. The purpose of the IAI is to define Industry Foundation Classes (IFCs) which enable the development of information exchange not only by means of file based exchange and sharing of database repositories as in STEP but also by promoting the development of interoperable software applications which use the newer technology of client/server interfacing. This is expressed as an ideal in the IFC Specification Development Guide.

Dublin Core. "The Dublin Core is a 15-element set of descriptors that has emerged from the Dublin Core Metadata Workshop Series that brought together librarians, digital library researchers, content experts, and text-markup experts to promote better discovery standards for electronic resources (Weibel et al, 1998). The consensus was reached on the semantics of 15 metadata elements. The metadata elements fall into three groups which roughly indicate the class or scope of information stored in them: (1) elements related mainly to the Content of the resource, (2) elements related mainly to the resource when viewed as Intellectual Property, and (3) elements related mainly to the Instantiation of the resource. Dublin Core is designed as a means for 'publishers' and authors to provide metadata at the point of mounting information on the Web. It is in the interest of these publishers to make metadata available that can be harvested by commercial and selective search services as a means to ensure their publications become publicised.

Linux Software Map (LSM) templates. The Linux software archives at the SunSITE addressed their need for a metadata standard with structured templates that contain 12 attributes appropriate to the archive needs.

IAFA templates. The Internet Engineering Task Force (IETF) Working Group on Internet Anonymous FTP Archives (IAFA), later called IIIR, have produced the IAFA templates Internet Draft (Becket, 1995). This defines a range of indexing information that can be used to describe the contents and services provided by anonymous FTP archives. The draft has a rich range of templates, attributes and values that can be used to describe common and useful elements. The goal is that these are to be used to index archives, made available publicly in them to allow searching, indexing and sharing of information on the archive contents, services and administrative data.

SOIF / RDM. The SOIF (Summary Object Interchange Format) is a record format used by the Harvest software. Harvest software was developed at the University of Colorado at Boulder, and is distributed by them as shareware. It is documented at http://harvest.cs.colorado.edu/Harvest/. Most SOIF records are generated by robots, although as they are based on simple attribute:value pairs they can easily be generated by hand. SOIF records can also be used as an aid to creation of other metadata formats. A broker can support different attributes, depending on the data it holds. Often brokers will hold the full text of documents as well as metadata.

Harvest has been widely taken up within the academic community as a basis for search services. Of significant importance has been the recent adoption of Harvest technologies by Netscape. In 1996 Netscape announced they would use SOIF as a basis for their Catalog Server product. In a significant extension to the Harvest architecture, Netscape are working on 'Resource Description Messages' which provide a framework for the creation and communication of metadata. Resource Description Messaging (RDM) is a messaging format, which can be used as the basis of a query syntax. RDM evolved into RDF, which is addressed in the next paragraph.

RDF. RDF is the Resource Description Framework proposed by Netscape as an open industry standard for describing how metadata for content is defined in web documents (http://www.w3.org/RDF/). This metadata is descriptive information about the structure and content of information in a document. RDF is an application of XML. Many companies that provide information, like ABC News, CNN and Time Inc support the RDF proposal. The search engines like AltaVista, Yahoo and Webcrawler also support RDF.

ISO 7200:1984 Technical documentation - Document headers and title blocks. This standard covers drawings, specifications, technical drawings, and title blocks in CAD drawings.

ISO 11442:1993 Technical documentation. Looks at the handling of technical based computer information.

IEC 82 045 Management data (Metadata) associated with technical documents.

In CONNET a subset of the Dublin Core is being used as the lowest common denominator with which meta-data can be exchanged between services.

4.2 RELEVANT CLASSIFICATION SYSTEMS

Classification of information has been an efficient way of ordering and structuring it at least since Aristotle. Systems like the Dewy's or the decimal classification have been one of the only ways of organizing materials in a library. With the introduction of metadata and full text searching, the importance of classifications has lessened. Surfers on the Web do not access Amazon.com's bookstore or Yahoo's catalog of the Internet using these classifications. Classification systems, however, remain important:

- They continue to provide navigation and access structure, which is particularly important for users who are not quite sure what to search for.
- They enable the relating of different types of information, as described in the introduction to this chapter.
- Classifications also have explanatory and descriptive value. They extend the keyword set with which information has been tagged.

The following classification systems appeared useful in developing the initial set of services:

EI classification (by Engineering Information, Inc): It is very well structured, starting with Civil, Mining, Mechanical, Electrical, Chemical and Engineering General at the top level. These top-level categories are labeled with code X00. Each sub-category level has up to ten items and code fills next right 0. For example sub-categories of top-level category 400 Civil Engineering starts with 410: Bridges and Tunnels and ranges to 480: Structural Design and each category ends with general sub-category 490: Civil Engineering, General. It is reasonably structured and quite widely used. Considering Calculation and Software Center the problem was that existing classifications are usually developed to cover engineering activities or building products and they can not be directly used to classify engineering software or its purpose. This one (EI classification) was the only one, which was appropriate for this service.

CSC categories: Beside EI classification, we have used our own classification for Calculation and Software Center. Goal of this classification scheme was to cover engineering programs by disciplines and by purpose. It is quite self explainable and we believe that it will be most commonly used. The scheme is oriented towards three most important users of our service - A/E/C. These are also three of file top-level categories and divided to arbitrary number of subcategories. For example the top-level starts with Architecture, Engineering, Construction, Management and General-purpose software, Engineering is divided to Civil, Chemical and so on. So far we didn't find any reasons for establishing of code labels in the scheme, due to the fact that scheme is used just for browsing. Software is also classified by license as well as platforms, which are typical classifications for software directories.

Uniclass: BRE Classification Tool, also used in CONNET, helps to find appropriate classification code either by navigation through the Uniclass classification scheme or text search. Uniclass top-level categories codes with letters A to Q. This classification is very extensive and covers from Forms of information, Construction elements, products to material and their properties. It also includes Universal decimal classification. Sub-category code label are formed so that a digit is added right form last code label. For example L: Construction products has sub-categories from L1: Ground treatment and retention products to L8: Fixtures and furnishings. So each subcategory can have up to ten sub-categories, except the top level that can have more entries.

Building90: This classification has the same structure as Uniclass except that top-level codes start with numbers. It is partly covered by Uniclass classification and it starts with code label 1: Site equipment to 9: Operational devices. This scheme is specialized for building product and on site activities. It is used in Manufactured Product Service and is widely used in engineering community.

The following issues emerged related to the classification:

- Automated classification of metadata that is gathered by automated means, for example by the robots and agents of the Calculation center. In the calculation center this is currently implemented in a way that uses the classifications on the source data. Human control is discovering AN ERROR MARGIN OF ABOUT 2-3%. In the future, the 3500 classified items will be used as a training set for machine learning of classification skills.
- Automated translations and conversions between various classification systems. This too is done by a conversion program that does 1:many mapping between classifications. In the future, however, the machine-learned tool will be used to classify information directly, with a limited aid of the original classification.

4.3 APPLICATION INTERFACES

All interfaces to the CONNET central node and the related services are implemented using HTTP. However, a language specific API is then provided to enable a more easily implementable interface to any of the services. This may not be as efficient as using some other protocols for the exchange of data on the Internet, such as the Z39.50, but it is robust, does not require and extra server and can be implemented with tools and software already in place to provide HTML pages for the end users. The API therefore contains a of number of agreed-upon URLs. Since the location of the URLs are not defined and in the future will not be fixed, the location must be parameterized, for example after defining:

```
$LOC_CSC = 'www.fagg.uni-lj.si/connet/scs.cgi';
```

a function how to search library for software is defined like:

http://$LOC_CSC/Search?search=searchterm&format=formats

The API therefore defines the last part of the URL, the parameters passed to the service and the results returned. All AIPs have been documented using the Java language.

4.4 OVERALL CONNET SCHEMA

CONNET was designed using the Rational approach (Rational, 1999); some documentation, however, was also done using the EXPRESS-G language.

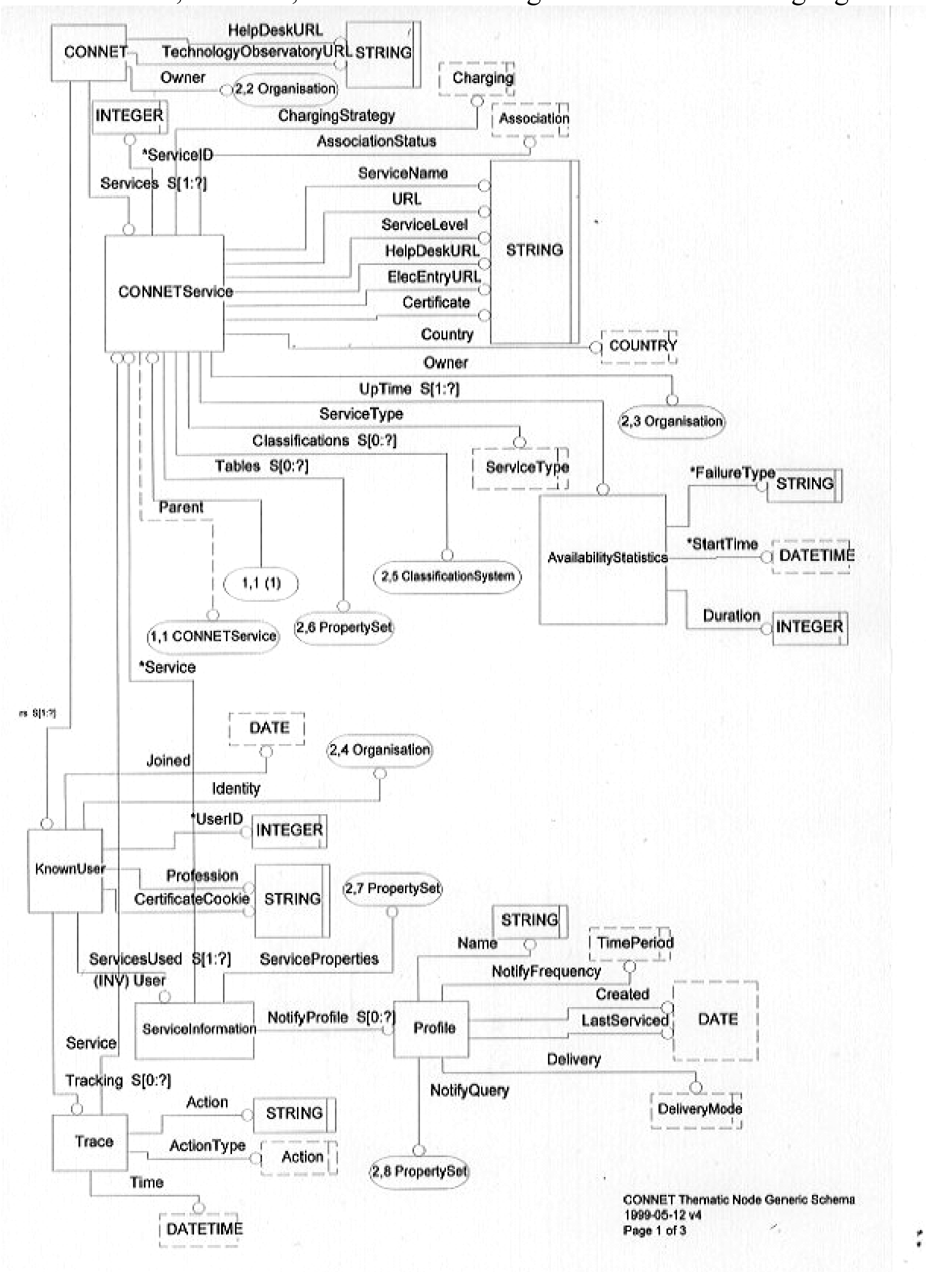

Figure 2. CONET thematic node schema presenting the concept of a CONNET node (in EXPRESS-G notation).

Figures 2 and 3 show two such diagrams, the first showing the information schema of CONNET as a whole and the second a detailed model of the Dublin Core metadata which is used as the lowest common denominator in CONNET.

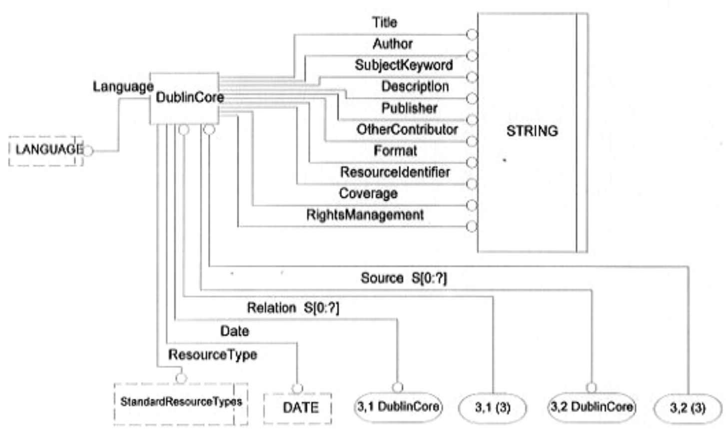

Figure 3. The concept of Dublin Core metadata - the lowest common denominator of CONNET metadata (in EXPRESS-G notation).

5. CONNET Implementation

Currently there are five services to be offered by CONNET (Figure 4). They are all implemented on the Web. They have a publicly specified API - a set of URLs that are available not only to the end users browsing the services, but to other nodes and talk to each other. Particularly user profiling, authentication, certificate distribution and the help desk are the services offered by the CONNET core. User profiling, for example, maintains information about a user and transfers it from service to service - e.g. if a user showed interest in concrete in the Technical Information Centre, she could be reminded of related software when she visits the Software Centre.

Services run on different hardware platforms, use different Web server and different database software. During 1999 an increasing part of their functionality has been made available on the Web. Since large parts of CONNET will be on line at the time of publication of this paper, we do not describe the surface of CONNET here - please visit www.connet.org. Technical details of individual services will be described in the follow-up papers.

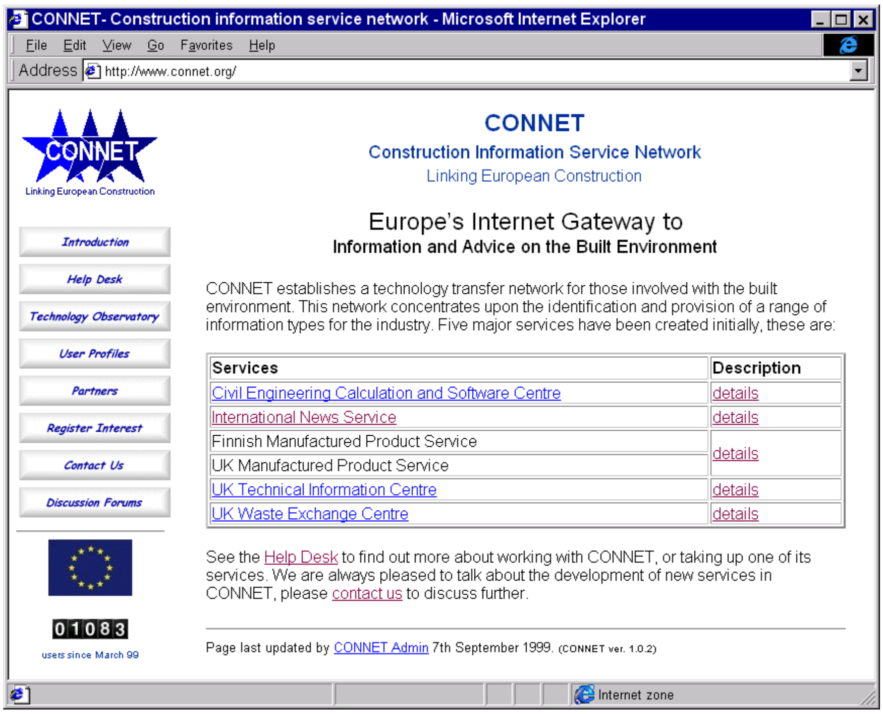

Figure 4. Main entry point of CONNET.

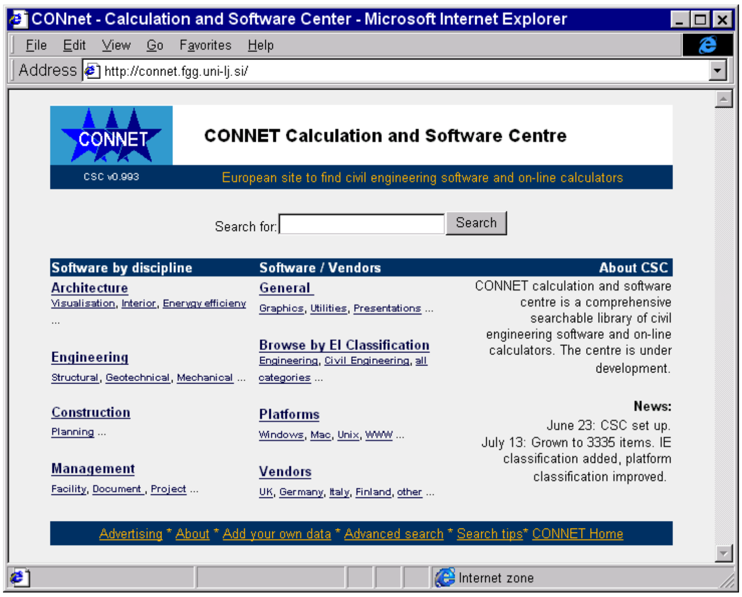

Figure 5. Title page of the CONNET Calculation and Software Centre.

6. Conclusions

CONNET has been a one-year project with a clear technical focus and little time for in-depth scientific research. The experiences we gained, however, are valuable for the current research directions in construction IT. Questions have arisen, that require a more detailed future studies.

There are several ways at which one could approach the design and development of an information system for the construction industry. At one extreme there is an integrated, coherent database into which any information the construction industry could ever be interested in, fits. At the other there are incompatible services on the Web. The overall architecture of CONNET, consisting of a loose federation of services is proving, is a compromise between the two. We acknowledged from the start that there is no right, unique, central way of structuring or classifying the information. Instead we felt it was important that each service's schema is transparent. It is this transparency that allows for the interoperability of the services. The areas in which a minimal coordination is required are:

- transparent information models
- defined application interfaces
- compatible user interfaces.

Information models. There are several metadata standards around but it does not seem that a consensus can be reached quickly with the vendors of the construction-related information. The diversity of the metadata should be acknowledged. A subset of Dublin Core is a reasonable lowest common denominator. The much hyped XML could be a very neat way of adding metadata to the Web or to exchange data on the Web - but for the first use it is too complicated and for the second there are a number of neutral data formats that can be applied as well.

In an ideal world there will be a "common" classification system for similar types of information. In the real world, however, information providers and information users see the classification system as their own, personal way of looking at the world. The classification systems are different and will remain so. Automated resource discovery also requires automated classification. A satisfactory solution to this problem also solves the problem of mapping between different classifications. In the mean time, however, looking at classes as fields that supply keywords only, is sufficient to provide functional browsing across different classification trees.

API. The lowest common denominator of the different services is that they use the Web. Some may include SQL databases or CORBA services, but to the end users, all will be showing HTML pages. It is therefore practical to use http as the lowest common denominator of the different services. If services standardize or freeze some URLs this is sufficient for the services to talk to each other.

User interface. The user interface can be made functional and aesthetic simply by using conventions of the Web, when it comes to navigation and browsing, and a usual search syntax. CONNET design is fairly conservative, not because we would not know how to use latest HTML gadgets, but because CONNET is more than "screen deep".

In the future, the existing CONNET services will be used as templates to provide new ones - both in terms of content as well as in terms of CONNET's geographic reach. A growing number of nodes will result in a growing number of information models and classifications. This will bring the "federated" approach to a harder test. Because of the distributed and fragmented nature of construction, we believe that technologies that enable the mutiple-schema and multiple classifications.

7. Acknowledgements

This paper would have not been possible without the contribution to the CONNET project by M. Hann**us, A. Hutchison , J. Hyvarinen, M. Salmi** and many others.

8. References

Weibel, S., J. Kunze, C. Lagoze, M. Wolf (1998). Dublin Core Metadata for Resource Discovery, RFC 2413, http://sunsite.cnlab-switch.ch/ftp/doc/standard/rfc/24xx/2413

Beckett, D. (1995). IAFA Templates in Use as Internet Metadata. David Beckett, Computing Laboratory, University of Kent, Canterbury, CT2 7NF, England http:// www.hensa.ac.uk/ tools/www/iafatools/paper/paper.html

Rational (1999). Website. www.rational.com.