ARCHITECTURAL FOUNDATIONS OF A CONSTRUCTION INFORMATION NETWORK

Z. Turk[1] and R. Amor[2]

ABSTRACT: Integrated information systems for construction products and information will become an important catalyst of the global construction market. In the frame of the 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 (www.connet.org). CONNET is designed as a set of loosely coupled nodes, operated by the different partners in the project. The nodes are integrated on both semantic and technical levels. The first integrates the information offered by the nodes by providing the mapping and translations between the metadata and the classification systems. On the technical level, the services are integrated by providing user management and cross-node searching enabled by publicly defined application interfaces. This paper presents aspects of 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 standardised 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, 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 an object oriented design and analysis approach proved vital for the success of the project.

KEYWORDS: information systems, construction, classification, metadata, world wide web

INTRODUCTION

Information technology (IT) is acknowledged to be a potent 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. The construction industry is generally regarded as being behind in its uptake of beneficial IT (Brandon and Betts, 1997). CONNET therefore focuses on approaches and techniques that enable technology transfer to small and medium enterprises (SMEs) as well as providing 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, 1999) and focuses on the information needs of the construction industry.

Additionally, CONNET is developing or customising the technologies to provide virtual "park grounds" for the integration of new nodes, for example to provide a nationally specialised version of an existing service, or to introduce new services on a national or international level.

---

[1] Ziga Turk, Assoc. Prof., IKPIR, University of Ljubljana, Jamova 2, Ljubljana, Slovenia; zturk@fagg.uni-lj.si

[2] Rober Amor, Ph.D., Building Research Establishment, Garston, Watford WD2 7JR, UK; trebor@bre.co.uk

Architectural foundations of a construction information network IN PUBLICATION.......... : Information technology for effective project management and integration / Qiping Shen, Heng Li, Peter Love (ed.). - Salford : University, 1999. - (International journal of construction information technology, ISSN 0968-0365 ; vol. 7/2). - Pg. 85-97Ilustr.

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 organisations 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.
- News Service.

Related work

The potential for the use of the Web in construction has been analysed since the mid 1990s (Turk and Vanier, 1994). Work on the possibilities for a comprehensive construction information network, which culminated in this project started in 1995 (Amor at al, 1996). A number of larger scale technology transfer initiatives were launched in the European Union. Esprit SCENIC project (SCENIC, 1999) aimed at "encouraging the adoption of advanced IT in the daily practice of the construction industry" and developed several databases on the Internet. GENIAL (1999) tackled some of the research issues for engineering networks, such as the problems of a linking a network of distributed value-added service providers to enable global engineering.

In several areas, commercial and industry sponsored activities developed concurrently with research efforts. An example of a big, commercial centre for construction information, is AECinfo (1999). Building associations or centres in several countries have been creating information nodes for the construction industry. For example, the Slovenian Building Centre launched its TIGRA service in 1996 (Turk, 1996). Building centres across Europe are implementing or planning to implement their services on the Web (UICB, 1999).

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 a node's functionality is desirable.
b) Features that enable the nodes to function as part of a wider technology transfer network and to integrate them with the central CONNET services.
c) Features which are node specific and do not influence the node's co-operation with the rest of CONNET.

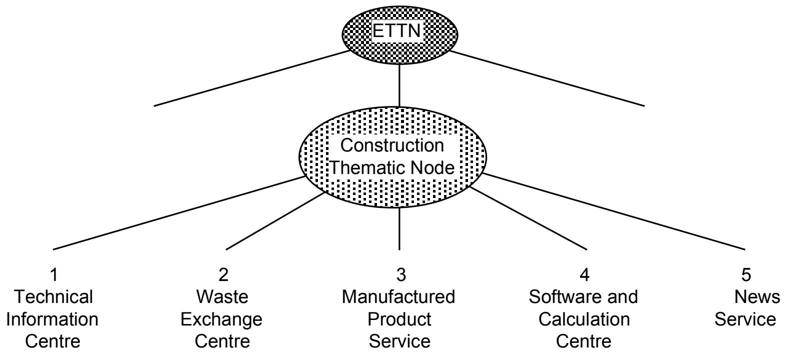

Figure 1. CONNET services are a part of the construction thematic node.

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 information 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-sales 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 specialised service on the Internet. The added value of a CONNET node is that it provides an efficient search mechanism, uniform access to a potentially complete set of information coming from different providers across Europe, some basic quality control, and, in some cases, offers electronic commerce services for finalising the transaction between a customer and provider. Most CONNET nodes are therefore financed by the providers, either by giving commission on the sales made through CONNET, by advertising on CONNET services, or by reimbursing for the fact that the information about 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 a mix of both models. Since there is very little experience in e-commerce for 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.

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, the possibilities of some level of compatibility in the following areas were studied:

- **Product and metadata.** Most of CONNET's services provide information about information or a product (e.g. software, publications, manufactured products, etc).
- **Classification.** Classifications ease the browsing in databases, provide more focused results, and enable the establishment of relations between different types of information. For example there are books, software, reports and product data about reinforced concrete.
- **APIs.** Application interfaces for the services should be public so that the integration with other services is possible.

Relevant metadata standards

For information as diversified as CONNET's, several standards for the product- and metadata have been considered.

ISO 10303 (STEP). STEP is the abbreviation for the Standard for the Exchange of Product Model Data (ISO, 1994). It is defined by the International Standards Organisation (ISO). The purpose of STEP is 'to specify a form for the unambiguous representation and exchange of computer interpretable information throughout the life of a product, enabling consistent implementations across multiple applications and systems'. It is concerned with the transfer of product data and also with the storage, access and archiving of such data and with ensuring that implementations can be tested for conformance. There is currently only one part directly relevant to building construction requirements in STEP. This is part 225 (Building Elements Using Explicit Shape Representation).

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 (IAI, 1996). 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 are locatable.

Linux Software Map (LSM) templates. The Linux software archives at the SunSITE address the need for a metadata standard with structured templates that contain 12 attributes appropriate to the archive's needs of describing software components.

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, are to be made publicly available 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. The Harvest software was developed at the University of Colorado at Boulder, and is distributed by them as shareware (Hardy et al, 1996). 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.

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 (RDF, 1999). This metadata is descriptive information about the structure and content of information in a Web 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. Covers the handling of technical based computer information.

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

In CONNET the Dublin Core model was selected and is being used as the lowest common denominator through which metadata can be exchanged between services. It is planned to take this representation forward into RDF/XML structures to integrate with existing search engines.

Relevant classification systems

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

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

The following classification systems were used in developing the initial set of services:

EI classification: Developed by Engineering Information, Inc., this is a highly structured system, starting with Civil, Mining, Mechanical, Electrical, Chemical and Engineering General at the top level. These top-level categories are labelled with the code X00. Each sub-category level has up to ten items and the code fills out towards the right-hand 0. For example sub-categories of top-level category 400 Civil Engineering starts with 410: Bridges and Tunnels and ranges to 480: Structural Design. Each category ends with a general sub-category 490: Civil Engineering, General. It is reasonably structured and quite widely used. Considering the Calculation and Software Centre, 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. The EI classification was the only identified classification, which was appropriate for this service.

Uniclass: This UK classification was developed by NBS Services (NBS, 1996). BRE has implemented it in its Generic Classification Tool (BRE, 1999) that helps to find appropriate classification codes, either by navigation through a classification scheme, or by text search of the code titles and descriptions. For Uniclass it shows top-level category codes with the letters A to Q. This classification is very extensive and covers Forms of information, Construction elements, products to material and their properties. It also includes the Universal decimal classification. Sub-category code labels are formed so that a digit is added to the right of the 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 alphabetic level that can have more entries.

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

CONNET service categories: Beside the major classifications, CONNET has defined its own classification sets specific to a service, but in many cases able to be utilised by several services. For example, in the Calculation and Software Centre a classification schema was defined to cover engineering programs by disciplines and by purpose. It is quite self explanatory and we believe that it will be the most commonly used classification mechanism in the service. The scheme is oriented towards the three most important users of the service: architects, engineers, and constructors. These are also three of the top-level categories and are divided into an arbitrary number of subcategories. For example the top-level starts with Architecture, Engineering, Construction, Management and General-purpose software. Engineering is divided into Civil, Chemical and so on. Software is also classified by licensing mechanism as well as platform, which are typical classifications for software directories.

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 centre. In the Calculation centre this is currently implemented in a way that uses classifications of the source data. Human checking is discovering an error margin of between 2-3%. In the future, the 3500 classified items will be used as a training set for machine learning of classifications for software items.
- Automated translations and conversions between various classification systems. This too is done by a conversion program that performs 1:many mappings between classifications. In the future, however, the machine-learned tool will be used to classify information directly, with limited aid from the original classification.

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 an 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 parameterised, for example after defining:

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

a function to search the 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 APIs have been documented using the Java language.

Overall CONNET schema

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

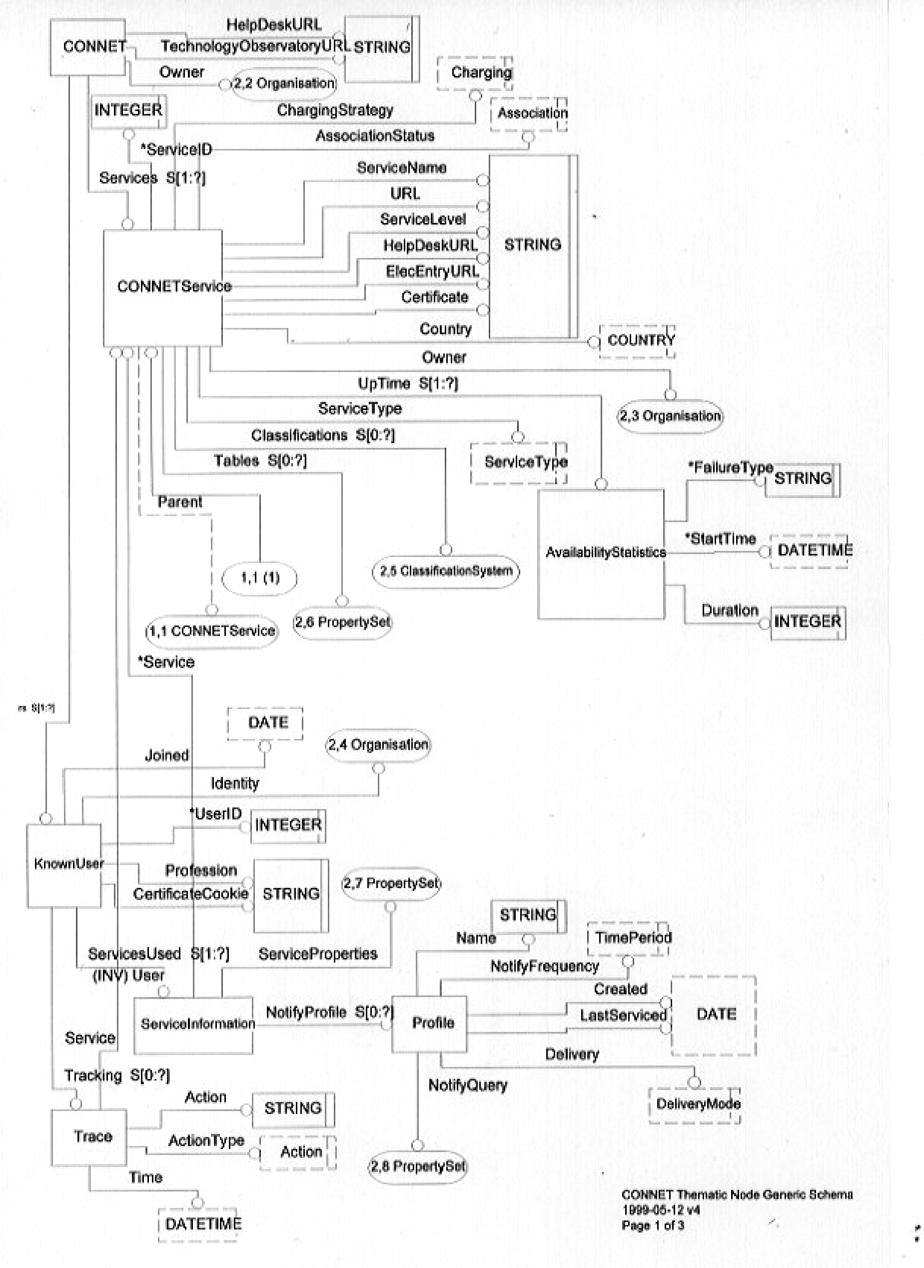

Figure 2. A portion of the CONET thematic node (in EXPRESS-G notation).

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

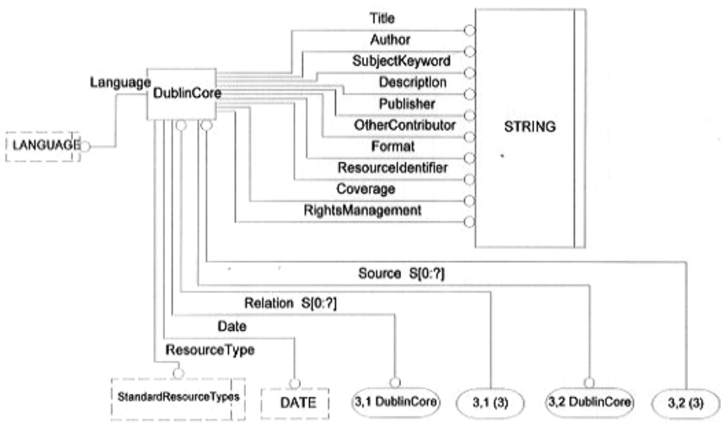

Figure 3. The Dublin Core metadata schema - the lowest common denominator of CONNET metadata (in EXPRESS-G notation).

CONNET IMPLEMENTATION

Currently there are five major information services offered through 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 to enable them to communicate with each other. User profiling, authentication, certificate distribution, and a help desk are some of 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. It also manages user requests for notification centrally for all services.

Services run on different hardware platforms, use different Web servers and different database software. During 1999 an increasing part of the 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 follow-up papers.

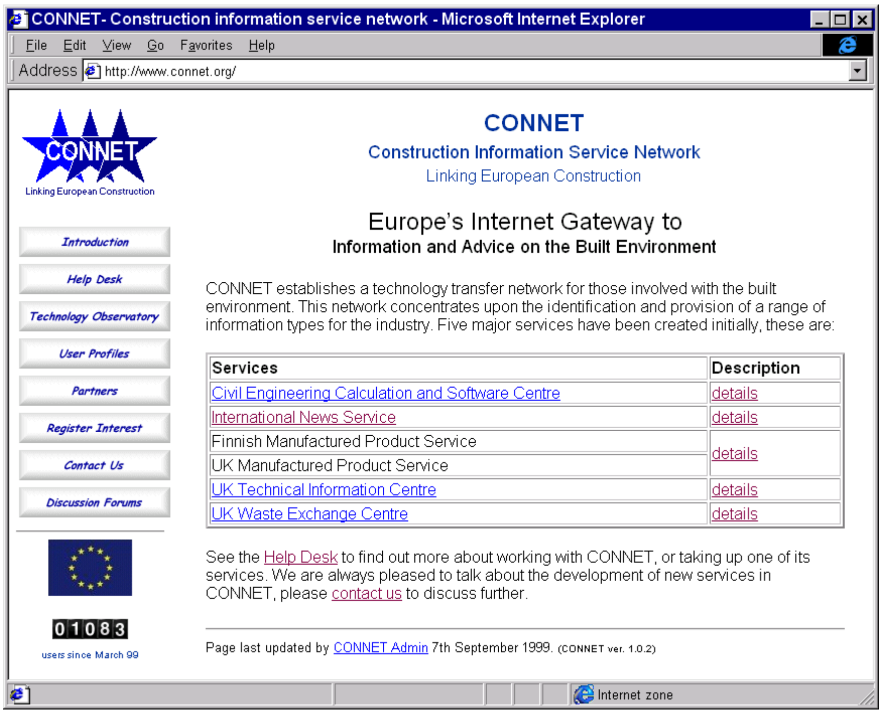

Figure 4. Main entry point of CONNET.

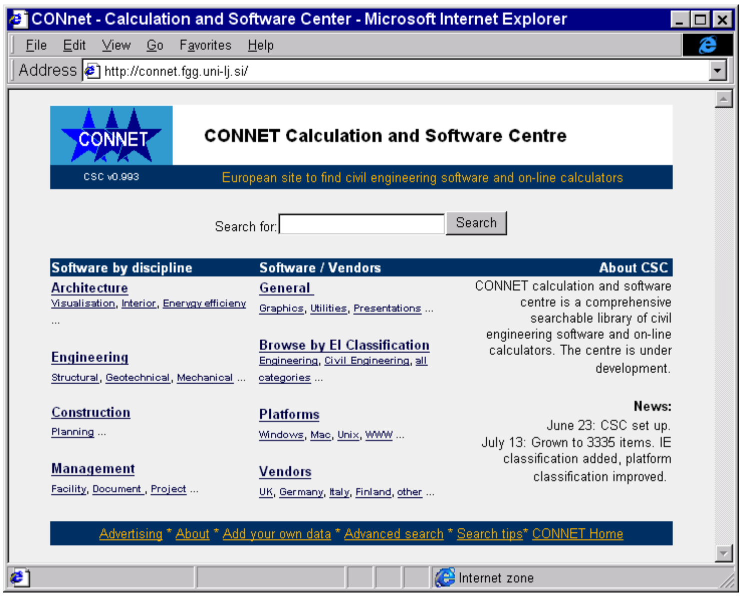

Figure 5. Title page of the CONNET Calculation and Software Centre.

CONCLUSIONS

CONNET was a one-year EC 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 detailed future studies.

There are several ways from 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, as are currently appearing on the Web. The overall architecture of CONNET, consisting of a loose federation of services, is proving that there 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 minimal co-ordination 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 construction-related information. The diversity of metadata should be acknowledged. A subset of Dublin Core is a reasonable lowest common denominator. The much-hyped XML could prove to be a very good way of adding metadata to the Web or to exchange data on the Web - but for current use it is too complicated and for now there are a number of neutral data formats that can be applied as effectively.

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 their 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 meantime, 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 standardise 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 standard search syntax. CONNET's 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 multiple schemas and multiple classifications are the most appropriate way forward.

ACKNOWLEDGEMENTS

The work presented in this paper has been funded by the EC through their European Technology Transfer Initiative. This paper would have not been possible without the contribution to the CONNET project by T. Cerovsek, C. Finne, M. Hannus, A. Hutchison , J. Hyvarinen, M. Salmi, and many others.

REFERENCES

AECInfo, 1999, Website. http://www.aecinfo.com/.

Amor, R., Langham, M., Fortmann, M. and Bloomfield, D., 1996, The UK Industry Knowledge Base Feasibility Study, CIB Proceedings Publication 198, ISBN 961-6167-11-1.

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

Brandon P. and Betts, M., 1997, Veni Vidi Vici, The Armathwaite Initiative, University of Salford.

BRE, 1999, Website. http://cig.bre.co.uk/connet/classifications/.

ETTN, 1999, Website. http://www.jrc.org/etti/ettn.htm.

GENIAL, 1999, Website. http://wwwgen.uni-paderborn.de/GENIAL/.

Hardy, D.R., Scwartz, M.F., Wessels, D., 1996, Harvest Users Manual, Technical report CU-CS-743-94, University of Colorado at Boulder.

IAI, 1996, End User Guide to Industry Foundation Classes, Enabling Interoperability in the AEC/FM Industry", International Alliance for Interoperability IAI.

ISO, 1994, Industry Automation Systems-Product Information Representation and Exchange - Part 1, Overview and Fundamental Principles", ISO TC184/SC4 Project Management Advisory Group.

NBS, 1996, Uniclass United Classification for the Construction Industry, NBS Services, UK.

Rational, 1999, Website. http://www.rational.com/.

RDF, 1999, Website. http://www.w3.org/RDF/

SCENIC, 1999, Website. http://scenic-europe.cstb.fr/

Turk, Ž. and Vanier, D., 1994, WAN Opportunities for Distributed Construction Information, Preproceedings CIB W78 Workshop on Computer Integrated Construction, VTT, Helsinki, Finland.

Turk, Ž., 1996, TIGRA - Tehnical information system for the building industry In Slovene, MEGRA 1996, Gornja Radgona, GI-ZRMK, Ljubljana, page 19-34.

UICB, 1999, Website. http://www.uicb.org/.

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/.