The Evolving CONNET Gateway to European Construction Resources

D. Bloomfield[1], R. Amor[2] and M. Groosman[3]

[1]BRE Ltd., Bucknalls Lane, Garston, Watford WD2 7JR, UK, bloomfieldd@bre.co.uk
[2]University of Auckland, Private Bag 92019, Auckland, New Zealand, trebor@cs.auckland.ac.nz
[3]TNO, Delft, The Netherlands

ABSTRACT: For the construction industries to move into the knowledge society and knowledge economy they need to be able to build upon their existing information base. This information base is unique within individual countries (though often with significant overlap between countries, for example, with Eurocodes utilised across Europe) and usually widely dispersed. Drawing together the information resources within nations, and then connecting them with each other to form transnational resources enables a more effective, informed, and intelligent industry. I-SEEC is a collaborative project funded by the European Union with the overall goal of creating an infrastructure to enable and link high quality commercial electronic information services throughout its member countries. This project started in March 2000 and finishes in April 2001. It builds upon a previous EU project - CONNET (CONstruction information service NETwork). This paper provides a description of the final state of the infrastructure, services, and business models available through I-SEEC. The countries participating in I-SEEC are Belgium, Finland, Germany, Iceland, Italy, the Netherlands, Slovenia, Spain, and the United Kingdom.

KEYWORDS: European portal; Information base; Internet gateway; Knowledge society

INTRODUCTION

The construction industry plays a very important role in the economic well-being of the European Union, representing approximately 11% of Community GDP in 1996. In spite of a slow-down in the rate of increase of productivity, the construction industry remains the largest industrial sector in the Community, ahead of the foodstuffs and chemicals industries, and it is crucial in producing investment goods. Construction is the largest sector in terms of employment, providing jobs for 8.8 million people (7% of the working population), and gives rise to 2.5 million construction-related jobs, and 14.3 million other service sector jobs.

However, the structure of the construction sector provides unique challenges. Chief among these is the size of the industry. The construction industry is also perhaps the most geographically diverse industry, and one which involves a very large number of small to medium sized enterprises (SMEs). EC figures show that of the 2 million construction companies in Europe, 97% have less than 20 employees, and 93% have less than 10.

The industry is also very challenging, due to the diverse nature of participants. The range of professionals on a project can easily cover a dozen disciplines, from architecture, engineering, construction through to facility management and demolition, all with very different information requirements. The training received by industry participants ranges from professional degrees, through to a very limited amount of formal training for many construction workers. Yet any change in the industry can impact on a wide range of these industry participants. For example,

the introduction of a new material may require changes in the construction process which need to be communicated to architects, engineers, facility managers, construction labourers responsible for the practical work, and those who finally demolish and dispose of the building.

Background to the project
The construction industry is perceived as being slow in its uptake of new technology and new processes. Certainly in the IT world this is borne out by recent surveys where IT expenditure per employee per year in construction (£453, ComputerWeekly 1999) is the lowest of any industry, and is considerably lower than the average (£2016). However, the use of IT is fairly high despite this, with only 1% of firms in the industry not using computers (Business and IT Survey 1999); 68% of architectural practices using CAD (AJ 1999); and 57% of firms using the Internet for business (Business and IT Survey 1999).

The Internet has been identified as a major form of dissemination for the majority of research and publishing organisations in the construction industry. This results in, for example, the majority of the 220 construction-related publishers in the UK having their own web sites. However, there is no unifying system to tie these sites together and offer their resources to the industry. The use of global Internet search engines provides little help, with extraneous and low quality information being returned along with the important information. Recent analysis also shows that the best search engine only covers 16% of the Internet's estimated 800 million publicly indexable web pages (Lawrence and Giles 1999). The top 11 search engines together cover approximately 42% of the total, so are in no way comprehensive.

These aspects make the construction industry ideally suited to the provision of Internet-based, quality, information gateways.

Related initiatives and parallel developments

Commercial Construction Portals
Over the period of the I-SEEC project there has been intense activity both at commercial and research levels to develop portals for A/E/C and the infrastructure that would support them. In this section we review the work in this area, compare to the I-SEEC initiative, and show where synergies exist between the differing initiatives.

Many construction portals have been launched in Europe over the last two years, in line with similar trends across all industries. Several of these portals failed within a short period of time indicating the tight marketplace within which they are competing, and the lack of known business models for successful systems. The portals that have been launched generally fall into two categories - those based around manufactured products, and those that support project participants. Portals in these two categories are discussed below.

Portals based around manufactured products form the majority of existing portals. In general they tend to duplicate paper-based catalogues, in several cases to the extent that they scan paper-based catalogues to provide their service. The benefit of this approach is fairly low; the only added benefit over a paper system (assuming that speed of access to information is fairly comparable) is that updates are immediately visible to all users. Some portals are enabling e-

commerce for purchase of selected products, and a very small number are providing for selection of products based on their performance attributes. To augment these services the portals often tie in a selected set of related information services, for example, links to Standards documents, industry news feeds, and databases of selected professionals in the industry. A major criticism of these sites is that they lack comprehensiveness. This is often true even for their major information content (i.e. manufactured products), but more especially true for their associated services, which tend to have a minute portion of the information available to the industry. Exemplar systems in this area include:

- Bricsnet (http://www.uk.bricsnet.com/) provides a portal to manufactured products which has been established in six European countries, they offer project management tools, CAD tools, real-estate life-cycle management, and bidding and procurement modules
- BuildOnline (http://www.buildonline.com/) provides a portal to manufactured products operating in seven European countries, they offer project management tools, news feeds, tendering, and procurement modules
- b2build (http://www.b2build.com/) provides for trading, consulting, and catalogue management within three European countries
- BuildingWork (http://www.buildingwork.com/) provides a portal to tradesmen, construction industry professionals, tendering, estimating, procurement, planning, and project management
- HouseBuilderXL (http://www.housebuilderxl.co.uk/) provides house builders access to merchants, software, advice, quotes, as well as manufactured product information
- Interbat (http://www.interbat.com/) provides access to CSTB publications and information resources, construction library, along with manufactured products
- BatiWeb (http://www.batiweb.com/) provides access to industry professionals along with the manufactured products
- ASCWebIndex (http://www.ascwebindex.co.uk/) purely provides manufactured product information

Portals based around project information management provide a service to industry by enabling virtual teams to be formed from all participants in a project where project documents are shared between participants in a very immediate fashion. These portals also tend to offer associated services along the same lines as the manufactured product portals and with the same deficiencies. Exemplar systems in this area include:

- services from IntraBouw e.g. BouwDesk (http://www.bouwonline.nl) a virtual office web site
- Mercadium (http://www.mercadium.com/) a portal for project management in five European countries, along with procurement, catalogues, auctions, ask the expert, and legislation and regulations
- Citadon (http://www.citadon.com/) a suite of integrated services allowing bidding, reprographics, and some financial services on top of its project management suite
- the BIW (http://www.thebiw.com/) news and industry information alongside the project management services
- Buzzsaw (http://www.buzzsaw.com/) virtual meetings, bookstore, design tools, reprographics, and quote management on top of its project management suite
- Cadweb (http://www.cadweb.co.uk/) and I-Scraper (http://www.i-scraper.com/) project management

CONNET aims to demonstrate how such services can conveniently be linked together and mixed and matched to suit requirements and preferences of individual users.

CONNET is significantly different from the above because it is not one company; rather it is an infrastructure to support collaborating portals. It adds value at a low budget. In fact, the I-SEEC project has allowed new services to be created using the CONNET infrastructure at much reduced cost.

EC Projects

Currently, there are very few EC projects underway which impact A/E/C and the Internet. The recently completed project GENIAL demonstrated a framework for service interoperability. The commercialisation of their results for multi-marketplace developments is still awaited to enable decisions to be made of their applicability to the CONNET system. Aspects which have been commercialised (e.g., database-driven web service enabling) are more appropriate for stand-alone services than for the gateways that CONNET supports.

One currently running project is eConstruct (Electronic Business in the Building and Construction Industry: Preparing for the New Internet) which:

> " aims to develop, implement, demonstrate and disseminate a new Communication Technology for the European Building-Construction industry, called Building-Construction eXtensible Mark-up Language (bcXML). This Communication Technology will provide the European Building and Construction industry with a powerful but low cost communication infrastructure that:
> - Supports electronic business between Clients, Architects and Engineers, Suppliers (of components, systems and services), Contractors and Subcontractors,
> - Is integrated with eCommerce and Design/Engineering applications, and
> - Supports virtual construction enterprises over the borders of the individual European member states."

Their bcXML and bcTaxonomy development that is based on the GENIAL approach could be linked with some of the CONNET services as a further classification item. eConstruct and I-SEEC are aware of each other and some preliminary discussions have been held to discuss the use of bcXML in I-SEEC and I-SEEC as a demonstration of the utility of bcXML.

Ongoing Standards Development

A range of Standards activities are ongoing in the A/E/C domains at the present time. Of relevance to this project are the data models under development by the IAI (which are now appearing in commercial CAD systems) as well as the XML initiatives. These data models allow for the transfer of detailed building specification and configuration information along with particular processes in the domain. For the majority of services in the CONNET gateway these structures have little bearing. The main exception being in the manufactured product area where the IAI's IFC data models provide the basic structures required for the representation of this information. However, the structures used in this project to represent documents, organisations, cost models, etc are closely aligned with those of the IAI and XML developments. The models utilised are usually derived from non-construction-specific standards organisations such as W3C

(e.g., the Dublin Core definition which has been utilised for the technical information centres). The XML developments (e.g., aecXML, bcXML, designXML) basically provide for process representation and categorisation of information from the A/E/C domains. This is currently supported by the classification structures within the CONNET data models. The API invocation mechanism within CONNET is specified in such a way that these standard representations can be supported in future versions. For example, whilst a response mechanism of 'text' is usually specified in the API system this is extensible to response types of XML or IFC which would enable structured data to be returned from these services.

E-commerce initiatives
There is a wide range of e-commerce initiatives attempting to provide the winning framework for the e-commerce revolution. Currently it is uncertain which, if any, of these frameworks will be successful, and I-SEEC along with many other projects awaits a clear leader prior to committing to a particular framework. During this time we remain aware of progress in the area and attempt to learn from the developments being made. For example, in the eCo framework (http://eco.commerce.net/) there is provision for resource discovery through a protocol by which e-commerce systems can describe themselves and their interoperability requirements. approaches.
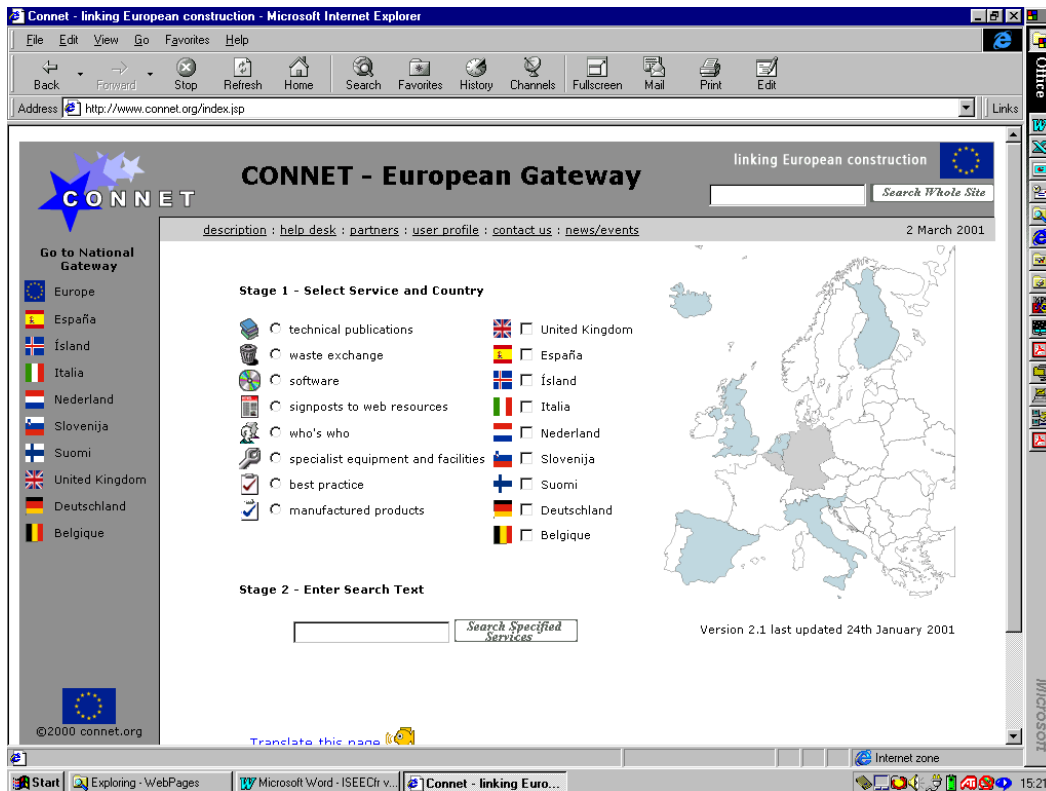

Figure 1 Entry Point to CONNET Europe

I-SEEC Introduction
I-SEEC is a collaborative project funded by the European Union with the overall goal of creating an infrastructure to enable and link high quality commercial electronic information services throughout its member countries. This project started in March 2000 and finishes in April 2001.

Nine countries are represented in the systems extant at end of February 2001. Over the course of I-SEEC contacts have been made with a wider number of countries and expressions of interest received.

I-SEEC has created a set of inter-linked national Internet gateways to quality information for the European construction industry (see http://www.connet.org/). It builds upon the CONNET initiative of DG-III ETTN (CONstruction information service NETwork, 1999) to provide the construction industry with an essential source of such information. It has created a "virtual technology park", accessible to the whole industry, regardless of national boundaries. The services that can be accessed from the CONNET gateways include examples both of:
- existing commercial systems (e.g. Finland's BII bookshop, Netherlands BouwOnline, UK BRE bookshop, German IRB bookshop and other services),
- newly developed systems using CONNET data structures and conventions (e.g. Belgium, Iceland, Italy, Spain) as well as the newly developed Netherlands web thesaurus.

CONNET provides access to a range of high quality Internet-based services for the construction industry in Europe. It provides both a European entry point to identify resources and national entry points for localised service delivery. The European CONNET entry point provides a range of technology park services as well as industry-specific services. These services include:
- Management of security services, including installation and monitoring of security systems
- Help desk, providing a point of contact for potential service providers and for problem resolution
- Information broker role, enabling transparent access to information in the CONNET services
- Technology observatory service, including leading edge, current and best-practice technologies
- Provision of user profiles, allowing personalised delivery of updates in areas of interest
- Multi-classification support, permitting handling of national systems used across the EC.
- Inter-service communication services, allowing all comparable services to be identified and a query to be passed from one service to another service to answer.
- Multi-language support, enabling EC languages to be handled correctly and to provide basic translations between them.

It is important to understand that the CONNET philosophy is NOT to create a single portal and attempt to attract all users to it. This would be unrealistic, anti-competitive, difficult to manage and would make tailoring and branding to suit local user preferences and contexts more difficult. Rather, the aim has been to allow separate portals to add value for local users by implementing a certain level of open interoperability between them.

A further important consideration has been to encourage long-term viability of the systems. It has been an objective of the project to set up new services and to test and implement business models for each service so that both operators and users can continue to benefit from them once the I-SEEC project has been completed.

The UK's Building Research Establishment (BRE) is the main contractor for the I-SEEC project with partners: BII (Finland), VTT (Finland), BCRM (UK), IBRI (Iceland), IBIC (Iceland), GCS (Slovenia), IKPIR (Slovenia), IETcc - CSIC (Spain), ICITE (Italy), and TNO (the Netherlands).

Additional services were created for Minstère de l'Equipement et des Transports (Belgium) and IRB (Germany) through informal co-operation. This demonstrated the practicality of adding new national versions of the CONNET services or further national gateways.

I-SEEC SERVICES

There are eight different types of service offered by I-SEEC, each dealing with different information types:

Technical Publications	Waste Exchange
Signposts to Web resources	Calculation and Software
Who's Who in Construction	Specialist Equipment and Facilities
Best Practice Information services	Manufactured Products

They include both existing legacy, commercial systems, and new systems developed from scratch within the project. More than 35 services can be accessed from the CONNET sites.
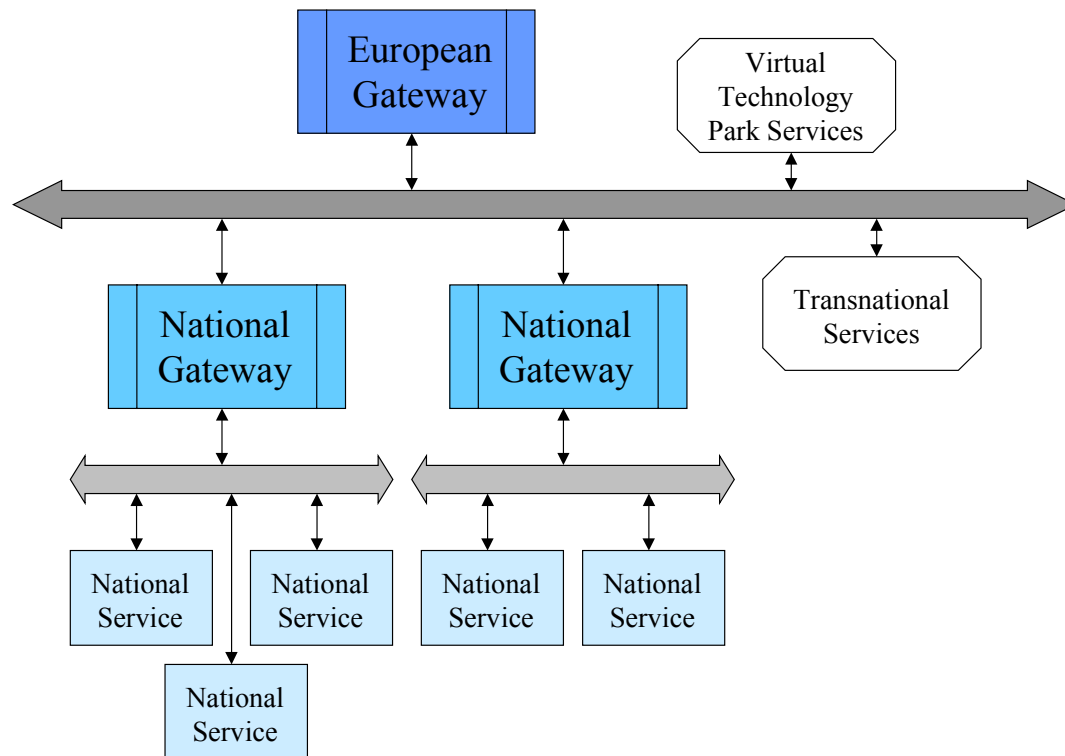

Figure 2 CONNET Gateway and Services Infrastructure

Infrastructure

The concept underlying CONNET is that advantage can be obtained by establishing certain standard ways of describing information and tailoring searches to take account of these. Data models (DM) and interfaces (API) have been devised for all the services, using existing standards where possible. These are documented in a separate report (Hutchison 2001) that is available from the FTP site for the project (http://www.connet.org/). A European gateway containing utilities to enable these searches has been implemented at BRE and this offers an

entry point to other national Gateways (see Figure 2 for relationships between infrastructure components). The latter may be hosted on the same machine as the European Gateway or by organisations in the host countries. In the current set of systems there are examples of both

Some of the systems enable e-commerce transactions to be executed and appropriate security systems have to be implemented. For example, BRE allows online purchase of technical publications, from BRE and other UK publishers. Credit card transactions are handled through arrangements with Netbanx who verify and guarantee security. Other models can be implemented by equivalent systems. For example, some only allow referral of the purchaser to the provider of the goods. In either case an extremely valuable service is provided by the CONNET systems in enabling a comprehensive search to be conducted to identify the existence of appropriate technical information from one entry point.

A national Gateway can be set up to provide information on only its national products and services but it is more useful and more interesting to allow searches to be extended to the other collaborating services (see Figure 3 for a generic national infrastructure).
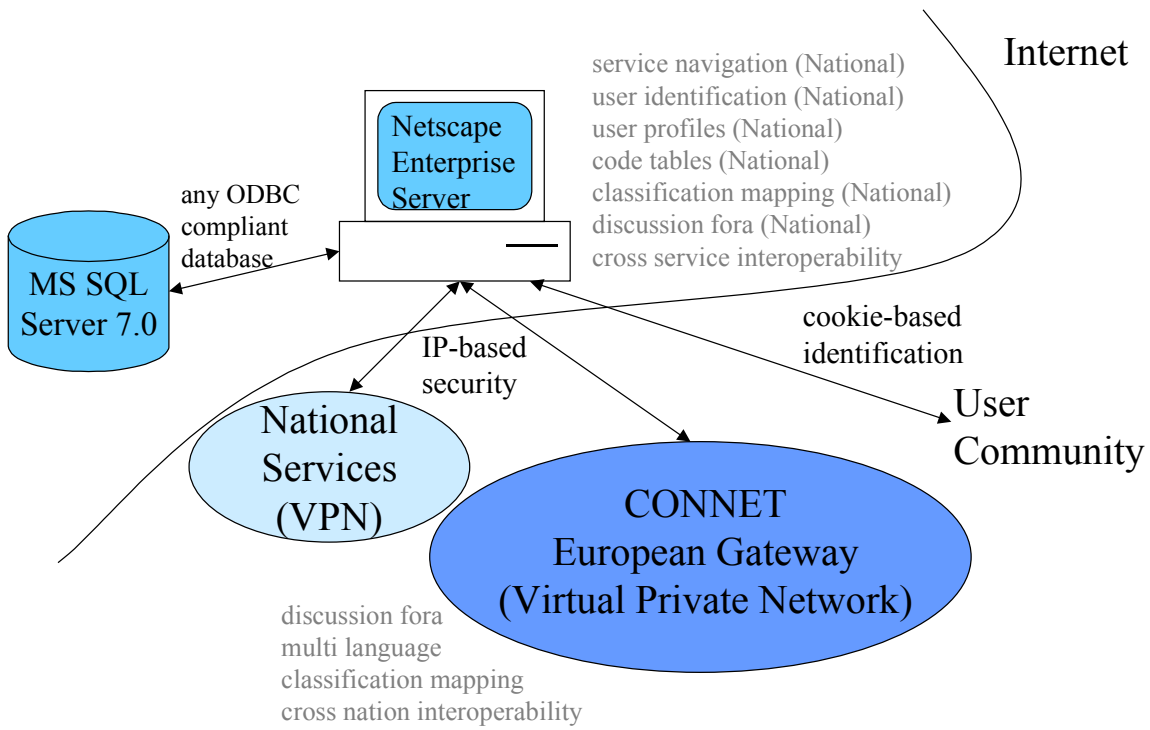

Figure 3 Generic Infrastructure for a National Gateway

I-SEEC aimed to develop a set of IT tools that were consistent across services and national gateways. It is important to note that there are two main classes of users. The first, and largest, class will wish to look for information pertaining to their own national context and language(s). The second class consists of more "European" users that wish to find information for countries other than their own and that are likely to be produced in the local language(s). The current systems allow CONNET/I-SEEC partners to decide which services would be of interest to foreign users, and the former were responsible for developing the most appropriate language interfaces and content.

An important development that has taken place within I-SEEC is the creation of an online thesaurus tool that can work behind the scenes to add synonyms and to provide translations, both for natural language words and for classification terms. It is expected that this will be a significant help for the "European" users described above. In order to cater for the first class of users, restrictions on the language of the original material can be placed by the user in setting up the search conditions. Searches can also be limited to a subset of the available national services.
National Services

It is worth repeating that the CONNET model is to encourage separate gateways/portals and add value for the user by enabling interoperability between them. This is in contrast to the more normal approach of trying to persuade all users to enter the virtual knowledge base through one entry point. The ease with which new services could be set up for new partners (Belgium and Germany) showed the advantage of defining data structures and APIs.
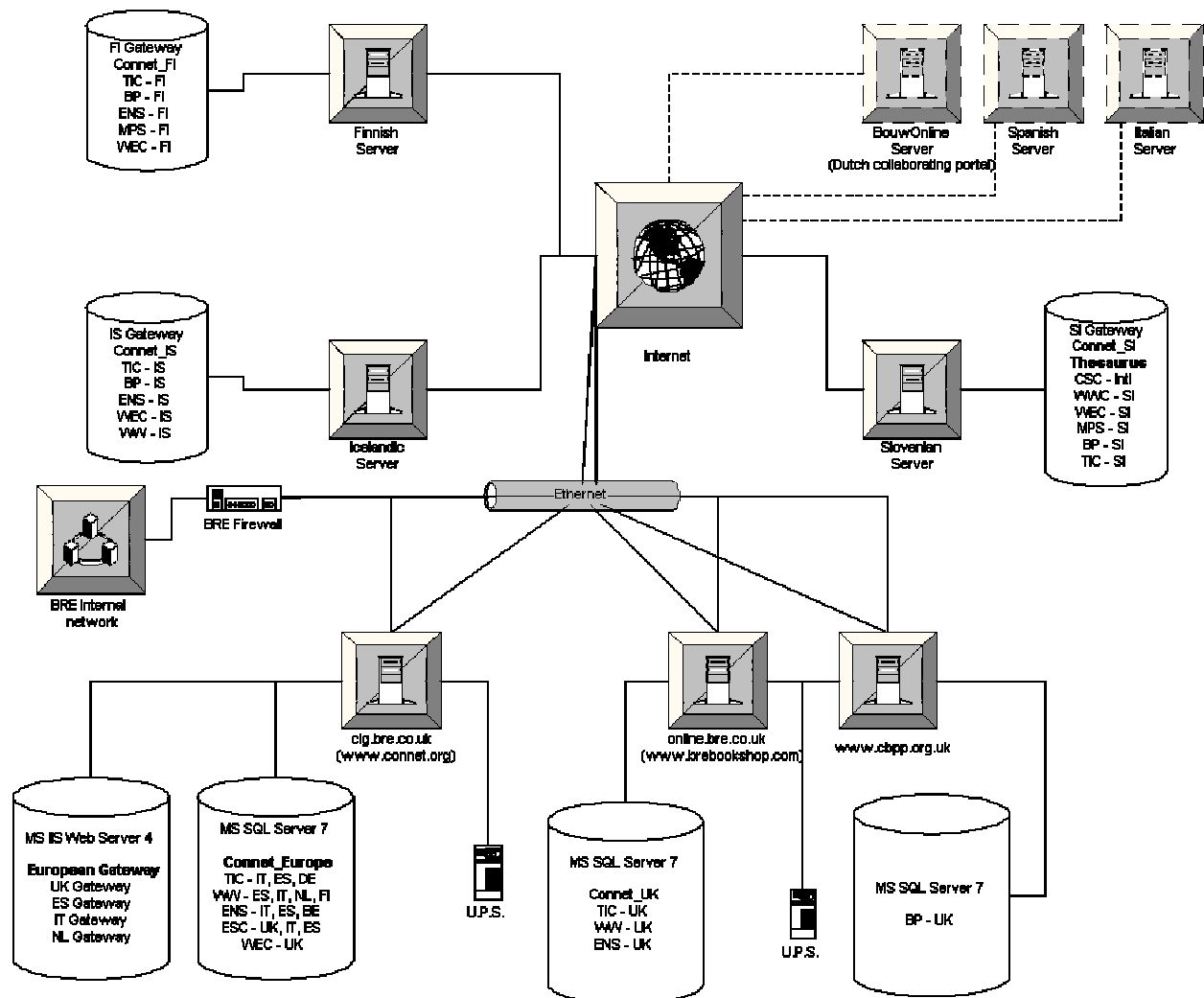

Figure 4 CONNET's Current Infrastructure

At the start of the I-SEEC project, national gateways for construction existed in very few countries. As part of their input to I-SEEC, the partners undertook to establish national gateways where none existed. Their experiences and motivation for these gateways are likely to be similar for other nations undertaking to develop their own national gateway.

Overview of systems
The diagram below shows the locations of the various servers and databases used by the CONNET system - the section that follows gives details of the web sites and services hosted on these servers.

The thesaurus, CONNET_Europe database, and the European gateway can all be used by other services to enhance their searches, and are marked in bold.

The minimum levels of functionality for a National gateway
This section describes the minimum level of functionality required from a National gateway i.e it should be able to provide details of the services that it offers. To do this each National gateway must provide the European gateway with an URL that can respond to the following queries:
- List services: which provides information on all services accessible from the gateway and the type of information they can serve
- Get service details: which provides details of an individual service including its starting URL, helpdesk, etc.

There is a more detailed document DM/API (Hutchison 2001) that describes a fuller set of functions that were originally envisaged.

The minimum level of functionality for a service
The minimum level of functionality required from a service is that it must be able to accept a query from another service or a gateway (national or European). To do this each service must provide its national gateway (and to the European gateway) with an URL that can respond to the following queries:
- Submit query: which allows a query in a structured format to be submitted and provides a structured return of information regarding the outcome of the search

Passing queries between services
There are two ways a service can widen its search
1. It can redirect/provide a link to the 'Search All' page on the European gateway, passing the query and parameters to say which services and languages to search. This will produce a page with the European gateway design i.e. the user will have left the national gateway or service they were in before. Note that the speed of this search is dependent on the slowest service being searched.
2. It can use the API to get details of the services it wants to query from the European gateway. These details will contain the URL to the API of the service. The service can then be queried through its API and the results displayed to the user in the format that the original service wants to use.

If any of the servers that a query is passed to is unavailable the request should timeout and the user will simply get no results from the service that failed - this will not lead to any delay or error message for the user.

THESAURUS, CLASSIFICATION AND LANGUAGE SUPPORT

A range of infrastructure developments were trialed within I-SEEC. This included classification, thesaurus, and language translation tools as discussed below.

Classification
The aim was to help users to find objects more reliably on the net. Classification tables were added and displayed in a search window. This is shown in the Finnish manufactured products service and also in the search facility for the Equipment Service where Uniclass and Talo 90 were implemented. This offers some improvement for users but more is needed.

The next step was to make translations of a classification system to other European languages. Under the current I-SEEC project the Finnish Talo 90 (Building 90 classification system) was made available under a Linux-like policy for its use and was translated into Dutch and Italian. Iceland is currently translating the Talo90 classification and has developed tools for tree representation and control. The description of the system is made available in PDF format.

Translation
A commercial automatic translation application "Babelfish" has been incorporated into the CONNET front page and in the Equipment page by BRE. This allows these and subsequent pages to be translated. A limited set of languages is available, for example Dutch is missing. The actual result of a translation can vary in quality, but an impression of the content presented is given. This has a different function to the search tool – it is for aiding a user's understanding of the text in the services' web pages.

These approaches, while helpful, have a limited functionality and efforts have been made to develop some other ideas as described below.

Multiple Language and Classification support in I-SEEC
As the Internet grows, so does the amount of information available through CONNET type services, and consequently the difficulty of finding what a user requires. This is the main infrastructure topic tackled under I-SEEC.

Simple string-search systems rely heavily on the user for knowledge of the domain being searching and on the information provider for proper domain-specific indexing. An open technical terminology would help both parties. This problem is certainly clear for systems providing cross-language services.

The use of a structured terminology is one of the ways to tackle this problem but in practice the classes defined in Construction Industry classifications are limited to the intended use, i.e. mostly organisation of information like specifications, budgets, planning etc. On the other hand there is not yet an exhaustive internationally accepted classification available. Classification codes are currently mainly used within manufactured product information systems. They can, however,

only be used within such a system and often the classification systems are not openly available without copyright conditions.

Another approach to improved information searching is through the use of language tools that assist in broadening or narrowing (generalise or specialise) a string search, by integrating (this object is part of …) or decomposing (this object is composed of the parts…), by referring to related terms or expressions, by using synonyms, preferred terms and explanations, definitions or scope notes.

Recently applications have been developed like Plumbnet by ThinkMap in the USA and the so-called Aquabrowser in Holland, both relying on an extensive thesaurus and visualisation of the relations between terms. However, there is no general lexical database or specialist vocabulary in construction available that could have been used directly in the I-SEEC project.

The necessity to have a common terminology available has been widely recognised and resulted recently in the approval of an ISO "Framework for object oriented information exchange" with anticipated deliverable "Framework for Terminology". (ISO/ PAS 12006-3). The development of the content of this framework is estimated in Holland to require some 25 person years' work.

WEB Thesaurus
So the approach taken in I-SEEC, based on earlier experience of TNO (using Datawares BRS in an on line pilot) and IKPIR (Geodetic Thesaurus, Web version of Canadian Thesaurus of Construction Terms) was to implement a relatively simple language toolset, based on thesaurus technology and the possibilities of a web-oriented database.

To develop content within I-SEEC's time frame, the strategy chosen was to collect as much existing material as possible (including freely available material from the Internet). Co-operation was sought with the Dutch BAS Association that integrates all the Dutch classification efforts of all sectors and links directly to the international development platforms. Extra financial input was obtained from STABU, the Dutch National Specifications Institute. The terminologies collected in I-SEEC will be made available to assist in the development of the WEB Lexicon (based on the ISO framework).

The combination of thesaurus technology, lexicography, multi-language terminology etc. were combined in a tool that was called the Web Thesaurus. Basically, languages are treated as synonyms, classification classes are treated as synonyms, and terms and expressions are explained or defined in the Scope Note when there is a formal description available. Thus an extensive construction-oriented terminology has been built that can be used for multi-language and multi-classification search and indexing.

The Web Thesaurus is a database of terms and concepts used in the AEC industry. Thesaurus relations have been added and terms in other languages and various classification classes they describe are represented as synonyms. Definitions and explanatory notes are given with a graphical illustration where useful. The web-thesaurus aims at being a helper in indexing and retrieval of AEC "objects". A prototype/demonstration model is to be found under http://thesaurus.forAEC.com/.

The Talo 90 (Manufactured Products table) was made available in 3 languages (Finnish, English, Dutch, Italian) with inputs from ICITE and TNO, and it is hoped to add Icelandic and Slovenian. The European Common Procurement Vocabulary (CPV) was reduced to terminology relevant to construction in eight languages. A Dutch-English version of the Nl-Sfb classification tables 1 and 2 were made available by TNO, together with a construction Equipment and Tools classification.   The Torroja Institute (IETcc) provided a Spanish-English construction terminology. The EPIC vocabulary was made available in English and Dutch and probably French will soon follow. The RAL Colour terminology is available in four languages. An English-German vocabulary was obtained from the University of Stuttgart.

IKPIR published all this terminology data on the Internet as searchable hypertext databases under a common user interface with a CONNET-compliant application interface. The URLs are http://words.forAEC.com/ and http://thesaurus.forAEC.com/. Both servers are hosted by the IKPIR Institute.

The formal co-operation with the BAS Association will also result in a copy of the standard mechanical components library and most probably also the electrotechnical library, both in Dutch.

The search engine in the CONNET site uses the WEB Thesaurus database to expand any search word into the other available languages.

The Web Thesaurus - background
Building on the IKPIR previous experience with on-line thesauri, being aware of the above described concepts and technologies, and the work to be done in the eConstruct project, the concept of the WEB Thesaurus was further defined in I-SEEC.

The combination of thesaurus, dictionary and encyclopedic methods have resulted in a tool to support searching, indexing, and knowledge transfer. This is an integrated multi-lingual, multi-classification database and knowledge management tool. The traditional thesaurus was enhanced with the possibility to define relations specific to the industry that makes objects. It is necessary to be able to define an object as part of another higher integrated whole.

It is a flexible tool primarily oriented to person-computer communication. It is catering for terminology originating from different countries and construction cultures, it can absorb new technologies, it helps to translate concepts but does not exclude the need for a certain degree of professional knowledge to use it and experts to maintain it. It shows where possible the context of terms and phrases, following the adage that information is knowledge only when placed in context.

When consulted it shows words and phrases in different languages and it treats classifications as a specific language. That makes it possible to search on a classification code and return the terminology that is part of that particular classcode.

Organisation
The Web Thesaurus toolset includes organisational instruments to support experts to communicate about terminology, to support editorial decision making, and to organise security and access at various levels. To top off the exercise, a virtual office (the Dutch BouwDesk Extranet application) was tested as an environment supporting the developers of the Web Thesaurus.

The Web Thesaurus needs a tool to manage off-line the terms and relationships. A prototype of this tool has been created, in parallel with the I-SEEC project.

Definitions
A thesaurus is defined as a specific set of terms that is used in a knowledge domain, formally organised so that relations between concepts are made explicit. It handles traditional relationships like broader term, narrower term, preferred term, synonyms, abbreviations, related terms, scope notes, etc. Thesauri and thesaurus technology were mainly developed in the library world to be able to index, store and retrieve physical documents.
A vocabulary, terminology or glossary mostly lists terms and the description or definitions. A bi- or multi-lingual dictionary mostly gives one or more translations without descriptions of the concepts involved. A lexicon can be a dictionary and vocabulary combined. An encyclopedia combines mostly words with descriptions and images illustrating the referenced words or phrases restricted to one language. A phraseology is a list of phrases.

To be able to group concepts or objects, a classification or taxonomy is used. Hierarchical and facet classifications are discerned and used in combination in proprietary classification systems. Well known are the UDC, the Periodic Table of the Elements, the Linnaeus Flora and Uniclass. Depending on the application envisaged, various systems are operational in the European AEC industry:
- the originally Swedish SfB,
- the English and Dutch translations Nl/SfB and CI/SfB, and
- the more recent Finnish Talo 90 and
- EPIC- Electronic Product Information Coordination group.

Apart from these, the Common Procurement Vocabulary (CPV) was developed by the EU in all or many European languages. AEC-related classes are part of it. The main aim of the classification systems in use is to organise information in specific documents like specifications and budgets and the ensuing quality and budget control. The CPV aims at facilitating European trade communication.

With increased use of computers, international trade and cooperation in design and construct processes, and the Internet of course, the demand for an international accepted classification system for construction objects arose and at last resulted in the ISO decisions referred to above. The new object-oriented construction classification is called the Web Lexicon. It will be instrumental in identifying objects unambiguously by their properties.

Specification of Functionality
The data model for the system to be used in CONNET is outlined below.

Structure of dictionary data
- term
- scope notes (within scope notes there can be links to other terms)
- figure

Structure of thesaurus
The structure is standardised according to NISO and ISO standards. The thesaurus is divided into the following parts:
- basic data
- relations between terms
- implicit relations between terms
- multilingual data
- editorial data

Basic data:
- NLT: Dutch Term
- NLSN: Dutch scope notes
- FIG: Scope notes
- URL: Pointer to URL where can we find additional information on current term

Relations between terms:
- BT/NT: Broader/Narrower terms
- WT/PT: Wider/Part terms
- RT: Related Terms
- UF/USE: Used for/Use instead (synonyms)

Implicit relations:
On the basis of data entered a computer program can calculate implicit relations:
- automatically generates inverse term relationship
- automatically searches the hierarchical tree of the term, that are not directly connected to the term
- adds implicit relations from scope notes
- add links to thesaurus where other language is used as primary language

Multilingual data:
- UKT: UK term
- DET: German Term
- FRT: French Term
- FIT: Finnish Term
- EST: Spanish Term
- SIT: Slovenian Term

In each translation there can be multiple translation into each language. Relations between languages are not one to one. Organisation of the web thesaurus enables easy upgrading, with quite a high level of automation, to any other primary language.

Editorial data:
- CD: Creation date
- CB: Created by
- AP: Approval date
- AB: Approved by

Anybody can add a term; this is added to the list of proposals and an Administrator (Editor) reviews the terms, make evaluations, relations, translations and adds to the list of approved terms.

Description of Thesaurus Tool
A demo is running at http://thesaurus.foraec.com/data/nl. It displays synonyms and allows the user to broaden and choose a narrower term.

Figure 5 Dutch Thesaurus Tool

The current system works in Dutch. Any other language can be set as the primary language, but translations need to be provided for the language engine as well as for displays, listings, forms, etc. The full data model is not yet functional. Some batch processing and integration tools also need to be developed.

FUTURE COLLABORATION

CONNET currently consists of some 13 organisations from 9 countries. The goal is to grow to 15-20 organisations from 10-15 European countries over the first year.

The most suitable instrument to achieve the above objectives seems to be a European Economic Interest Group (EEIG). This has organisational characteristics that enable it to take on functions typical of permanent forms of co-operation (e.g. international product marketing network). It is similar to a partnership with the important consequence of unlimited joint and several liability for its debts. It has legal neutrality so all members on placed on an equal footing. The formalities for formation are stated to be very simple – draw up a written contract, register at an appropriate registry of State in which the EEIG has its official address. The official address can be transferred between States. An EEIG can be formed for a limited or unlimited period. No start up capital is required and all forms of contributions are possible. It is a very flexible instrument in that members can decide all financial, voting, issues. It has full and autonomous legal capacity. It is governed by Regulation EEC 2137/85, which takes precedence over national law. It can enter into e.g. EU contracts. It is important to note that its purpose is to develop/facilitate economic activities of its members (not just the EEIG).

Levels of Possible Collaboration for new entrants
Potential partners could:
1. Link an existing/developing information system into a national gateway
2. Establish an information system in a country where one does not currently exist, either by
   - Developing independently but using the CONNET API (Application Programming Interface)
   - Re-using existing systems that have already been developed
3. Establish a national gateway where it doesn't currently exist
4. Register links and data with existing CONNET services, such as
   - URLs for relevant services
   - URLs for indexing by a crawler to provide coverage of national 'news' to CONNET sites
   - Provide data for TIC, Who's Who, or other existing services to enable these data to be incorporated in or hosted by existing CONNET/I-SEEC services

Note that to achieve the full functionality of the CONNET systems it is necessary to implement the API calls to allow compatibility with CONNET/I-SEEC systems and to allow queries to be passed between systems and results returned to originating system. The data models and APIs for all services have been documented and are freely available.

The advantages of building on the CONNET infrastructure and work to date are, depending on the level of adoption, to:

- Bring a wider audience to existing systems, enabling users of European and national systems to find their sites.
- Allow information on their publications, national organisations, specialist equipment etc. to be brought to the notice of the users of CONNET systems in the 9 countries currently participating in I-SEEC. For those countries without their own system this is a fast way of getting one set up.
- Help I-SEEC systems to be more compatible with their own systems and enable the latter to be improved for existing users by using the API of existing CONNET services.

- Link existing systems into a wider set of systems allowing increased functionality and generality (through the ability to widen searches to a European level) for users of both the newly linked system and also for those of the other linked systems.
- Provide an enhanced free text index on the CONNET News Service (currently indexing 19,000 sites) that includes the information on your existing systems and will also bring a wider audience to existing sites.

REFERENCES

Aitchison, J. and Gilchrist, A. (1987) Thesaurus construction; Aslib; London.

Hutchison, A. 2001 "CONNET Data Model and API report". www.connet.org/datastructures for CONNET

ETTN (1999) "European Technology Transfer Network". http://ettn.jrc.it/default/

TNO Rapport 2001-BKR-R021: Promotie I-SEEC/ WEB Thesaurus; M.Groosman.

URL1 "WordNet" - a Lexical Database for English http://www.cogsci.princeton.edu/~wn/

URL2 "E-Construct". http://www.econstruct.org/