CCLRC

C L R C

RUTHERFORD APPLETON LABORATORY AND DARESBURY LABORATORY
BUSINESS AND INFORMATION TECHNOLOGY DEPARTMENT

EUROCRIS

CRISs, Architectures and CERIF Keith G Jeffery

18 January 2004

1. INTRODUCTION

CRISs exist; some of them have existed for a long time. They have different purposes / requirements, data models and implementation platforms. However, there is a general recognition that it would be useful to provide a homogeneous view over these heterogeneous CRISs, especially in the ERA (European Research Area) context.   There are several ways to implement a solution to the problem of homogeneous access to heterogeneous CRISs. This short paper attempts to present and compare them. Some background information is attached as an Annex; this provides explanations and definitions of terms and also some previous experiences.

Apart from harvesting, which ignores the syntax (structure) and semantics (meaning) of the data and just does text string searching, all architectures rely on a predicate query over a known schema (available or derived by schema reconciliation) thus allowing search terms or values to be related to an entity/attribute and thus to a domain. Example: the string 'green' could occur under attribute 'family name' in entity 'person' or within attribute 'abstract' or 'title' in entity 'project'. The use of query under a schema ensures that the query is meaningful and should have adequate recall (coverage) and relevance (precision).

Most techniques rely on navigational metadata to access hosts of CRISs. The catalog techniques use in addition associative descriptive metadata to perform the first pass search – rather like the harvesting technique, but using structured and meaningful data under entity/attribute sanction. Those techniques with server-side or client-side wrappers require schema metadata to perform schema reconciliation. Although CRISs of these architectures usually do not use CERIF, CERIF can be used with advantage to define the database schemas.

One technique (Full CERIF) has a uniform assumed schema and so has no need for metadata nor schema reconciliation. However, it relies on each host either being full CERIF compliant or providing (and maintaining) a full CERIF version of the host database.

CERIF provides the optimal solution in the full implementation. All other techniques benefit from using CERIF to define schemas or export formats.

This paper describes the different architectures. Annexes provide references to (1) Metadata, (2) CERIF, (3) the EXIRPTS Protocol and in (4) the complete text (since the report is not easy to obtain) of the ERGO Group Report to the EC recommending the ERGO 2++ architecture.

2. REMOTE WRAPPER

2.1 Architecture

```
                        ┌─────────┐
                        │  user   │
                        └─────────┘

        ┌──────────────┐      ┌──────────────┐
        │ Query form   │      │ Presentation │
        │              │      │ form         │
        └──────────────┘      └──────────────┘
                                     ▲
                              ┌──────────────┐
                              │ Presentation │
                              │ convertor    │
                              └──────────────┘
                                     ▲                    ┌──────────┐
                              ┌──────────────┐            │ schemas  │
                              │ integration  │◄──────────►│          │
                              └──────────────┘            └──────────┘
   ┌──────────┐   ┌──────────────┐   ┌──────────────┐
   │addresses │◄─►│ dispatcher   │   │ receiver     │◄────────────
   └──────────┘   └──────────────┘   └──────────────┘         network
```



```
   ┌──────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────┐   ┌──────────────┐   ┌──────────────┐
   │addresses │◄─►│ receiver     │   │ dispatcher   │   │addresses │◄─►│ receiver     │   │ dispatcher   │
   └──────────┘   └──────────────┘   └──────────────┘   └──────────┘   └──────────────┘   └──────────────┘

   ┌──────────┐   ┌──────────────┐   ┌──────────────┐   ┌──────────┐   ┌──────────────┐   ┌──────────────┐
   │ schema   │◄─►│ Query        │   │ Xml          │   │ schema   │◄─►│ Query        │   │ Xml          │
   └──────────┘   │ convertor    │   │ convertor    │   └──────────┘   │ convertor    │   │ convertor    │
                  ├──────────────┤   └──────────────┘                  ├──────────────┤   └──────────────┘
                  │ Query        │                                     │ Query        │
                  └──────────────┘                                     └──────────────┘
```

<<<<Non-CERIF CRISs>>>>
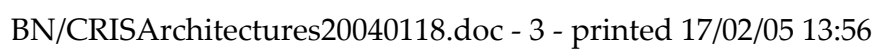
2.2 Description

This architecture provides a simple user query interface to multiple host CRISs. Each host CRIS has to

(a) provide address to the client dispatcher database

(b) provide software for query conversion to local host DML (using host schema)

The use of XML to encode answers (an addition to the basic architecture but indicated in the diagram) dispatched provides some syntactical uniformity but no uniformity of character set, language, semantics. Uniformity in these other aspects can only be achieved through a canonical data model (CERIF). Unfortunately XML cannot represent the full semantics of CERIF, because it represents hierarchies and CERIF represents a directed graph.

2.3 Metadata

This architecture uses schema metadata for query conversion and answer integration. It uses navigational metadata for access to hosts. This architecture does not use CERIF metadata.

2.4 Process

The user inputs a query through a supplied web browser form of the kind 'find the string "widget" anywhere in any host database'. The dispatcher sends this query in some protocol [email | ftp | message] to all hosts with address entries in the address database. Each host converts to its own DML (using the host schema) and executes the query. The results at each host are converted to XML (an addition to the basic architecture but indicated in the diagram) and dispatched back to the user who receives one XML file per host (each with differing syntax and semantics). The integrator takes the result sets and using the host schemas (or preferably XML DTD equivalents of the schemas) reconciles them to a uniform result set which is converted for end-user viewing by the presentation converter.
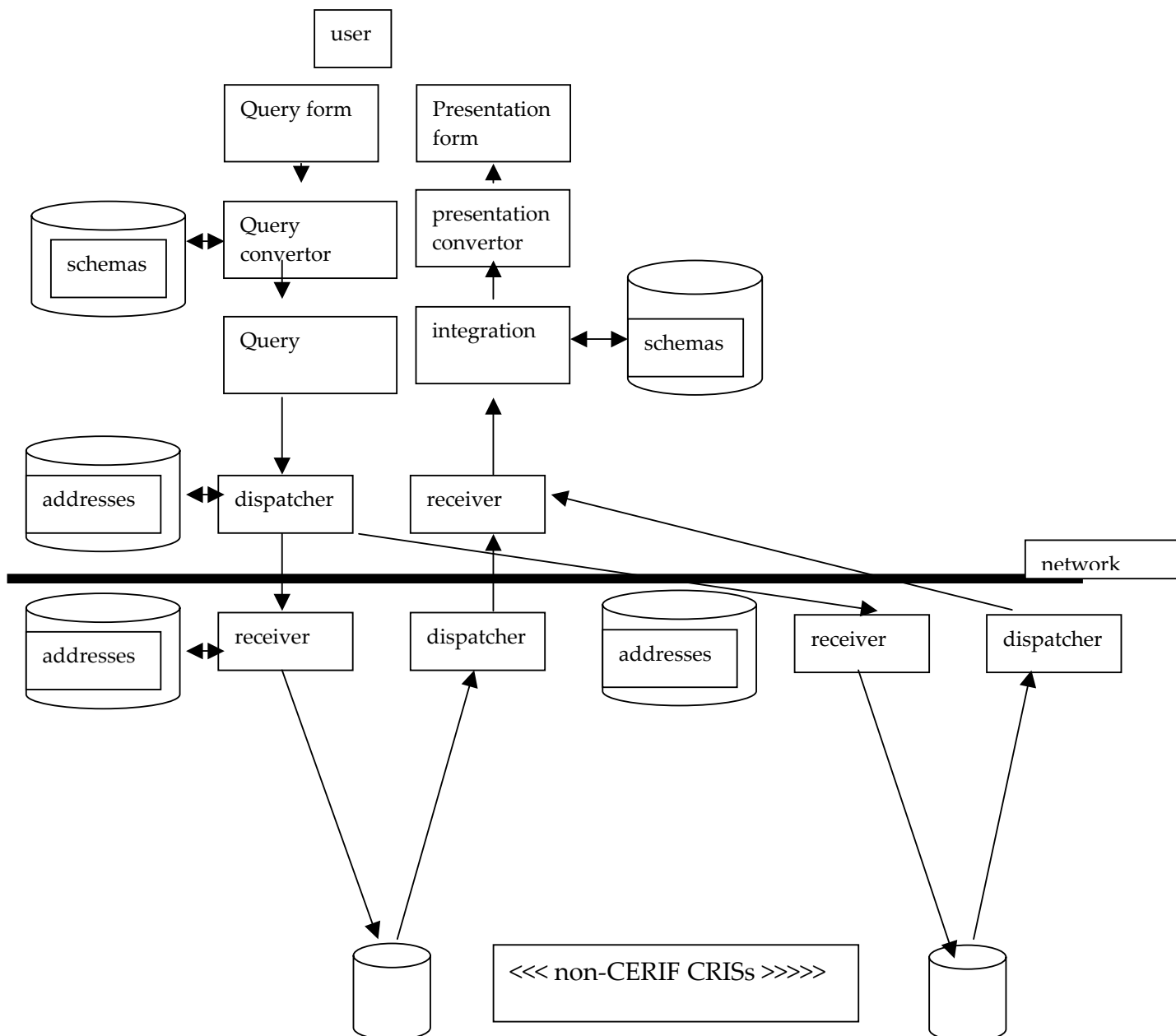
Advantages and Disadvantages

(a) user needs only web browser and simple query form

(b) host has to write query converter

(c) host has to write XML converter (to a specific XML DTD?)

(d) query expressivity very limited

(e) user client has to write an integrator for the answers

3. LOCAL WRAPPER

3.1 Architecture

3.2 Description

In this architecture the hosts have only to provide a receiver and dispatcher; they receive queries in their own DML and dispatch results in their own data model. All conversion responsibility is on the client. The client provides queries for each host from the user query, mediated by the host schemas and integrates the results from each host, using their schemas, to an answer for the end-user presented through a user-defined presentation converter (e.g. XML, HTML….).

3.3 Metadata

This architecture uses schema metadata for query conversion and answer integration. It uses navigational metadata for access to hosts. This architecture does not use CERIF metadata.

3.4 Process

The end-user generates a query in some arbitrary language, using a query refinement interface and web form. The client software converts the query to the target DML for each host using the host schemas stored (and updated by the hosts) at the client and dispatches them using the addresses database. Each host receives a query in its own DML, executes it and returns the result in its own form via the dispatcher to the client receiver. The integrator takes the result sets and using the host schemas reconciles them to a uniform result set which is converted for end-user viewing by the presentation converter. CERIF could, with advantage, be used as the uniform schema for result integration.
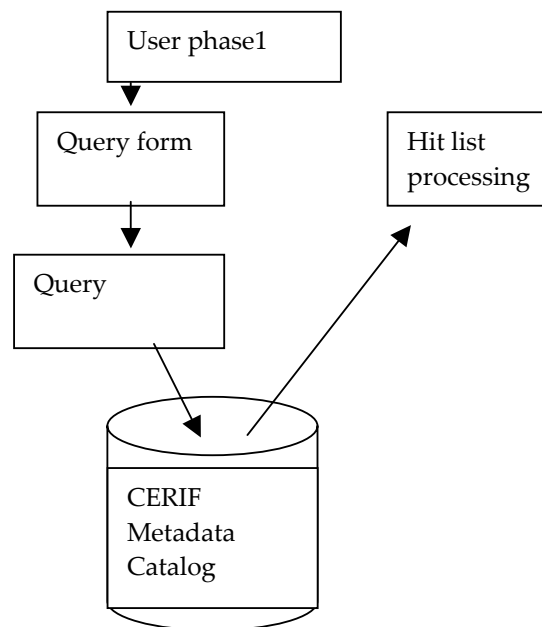
3.5 Advantages and Disadvantages

(a) each host has only to supply and update its schema to the client (all clients if there is not a central query server)

(b) each host has no software to provide except receiver and dispatcher

(c) the client (if it is a central service) has a very large workload

(d) if there is no central service then each client has to have all schemas supplied and updated

(e) the client software has to include a complex query refiner

(f) the client software has to include multiple complex query converters

(g) the client software has to include a complex answer integrator

(h) the client software has to include a presentation converter (complexity depends on specification of presentation required and complexity of the answer structure)

4. CATALOG

4.1 Catalog Only (ERGO Pilot)

4.1.1 Architecture

```
           User phase1
               │
               ▼
          Query form          Hit list
               │              processing
               ▼                 ▲
           Query               │
               │               │
               ▼               │
           ┌─────────────┐
           │ CERIF       │
           │ Metadata    │
           │ Catalog     │
           └─────────────┘
```

4.1.2 Description

This architecture provides a canonical subset data model – CERIF metadata model – with one character set, one language, one syntax (structure) and one semantics. This provides the homogeneity.

4.1.3 Metadata

This architecture uses associative descriptive metadata (CERIF metadata catalog)

4.1.4 Process

The CERIF metadata catalogue is populated from all hosts which provide a local converter from their data model to CERIF metadata (one character set, one language, one syntax (structure) and semantics (meaning)). The end-user has a query form which queries the catalog and obtains a 'hit list' of results. Experience indicates such results satisfy ~ 80% of queries; however if more detail is required the architecture provides the capability for accessing the hosts (see next section).

4.1.5 Advantages and Disadvantages

(a) simple query on union catalog (which may be centralised or replicated)

(b) possibly not all required entities and attributes in catalog

(c) effort to populate catalog; requires converter at each host to supply CERIF metadata

4.2 Catalog plus pull (ERGO 2++)
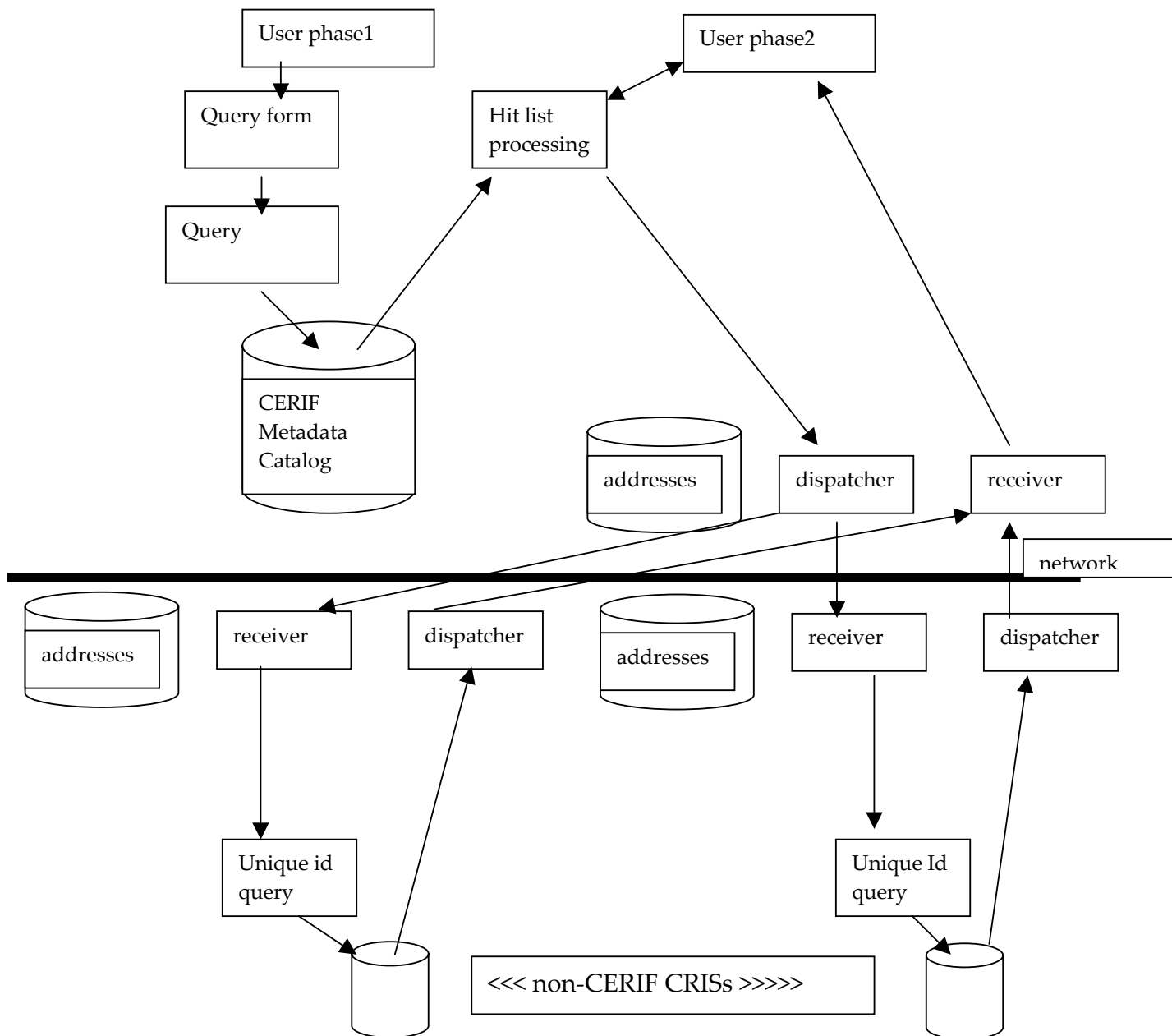
4.2.1 Architecture

4.2.2 Description

In addition to the Catalog-only model, this architecture allows a subsequent access to all hosts with hits in the CERIF metadata catalog to collect the detailed information the hosts are willing to supply. There is no further selection by attribute value nor projection of attributes, everything is 'pulled' if it relates to a hit record in the catalog. Ideally, the hosts convert to CERIF export model to provide uniformity but this is not mandatory.

4.2.3 Metadata

This architecture uses associative descriptive metadata (CERIF metadata catalog) and navigational metadata for host addresses.

4.2.4 Process

The hit list is edited by the end-user and then sent to the dispatcher which sends to each host the Unique Ids (primary key values) of the selected records for which further information is demanded. The host sends the answers back via dispatcher to receiver and thence to the user. No attempt is made to make homogeneous this detailed heterogeneous information which may have different character sets, language, syntax and semantics. Conversion to CERIF export model could be advantageous for integration by the end-user.

4.2.5 Advantages and Disadvantages

(a) advantage of simplicity as for catalog-only architecture

(b) advantage of additional information provision

(c) disadvantage that additional information is heterogeneous (unless converted to CERIF export data model)

(d) disadvantage of hosts having to maintain entries representing their database content in the CERIF metadata catalog

5. FULL CERIF

5.1 Architecture

user

Query form

Presentation
form

addresses

dispatcher

receiver

network

addresses

receiver

dispatcher

addresses

receiver

dispatcher

Query

Query

<<<<< CERIF CRISs >>>>>
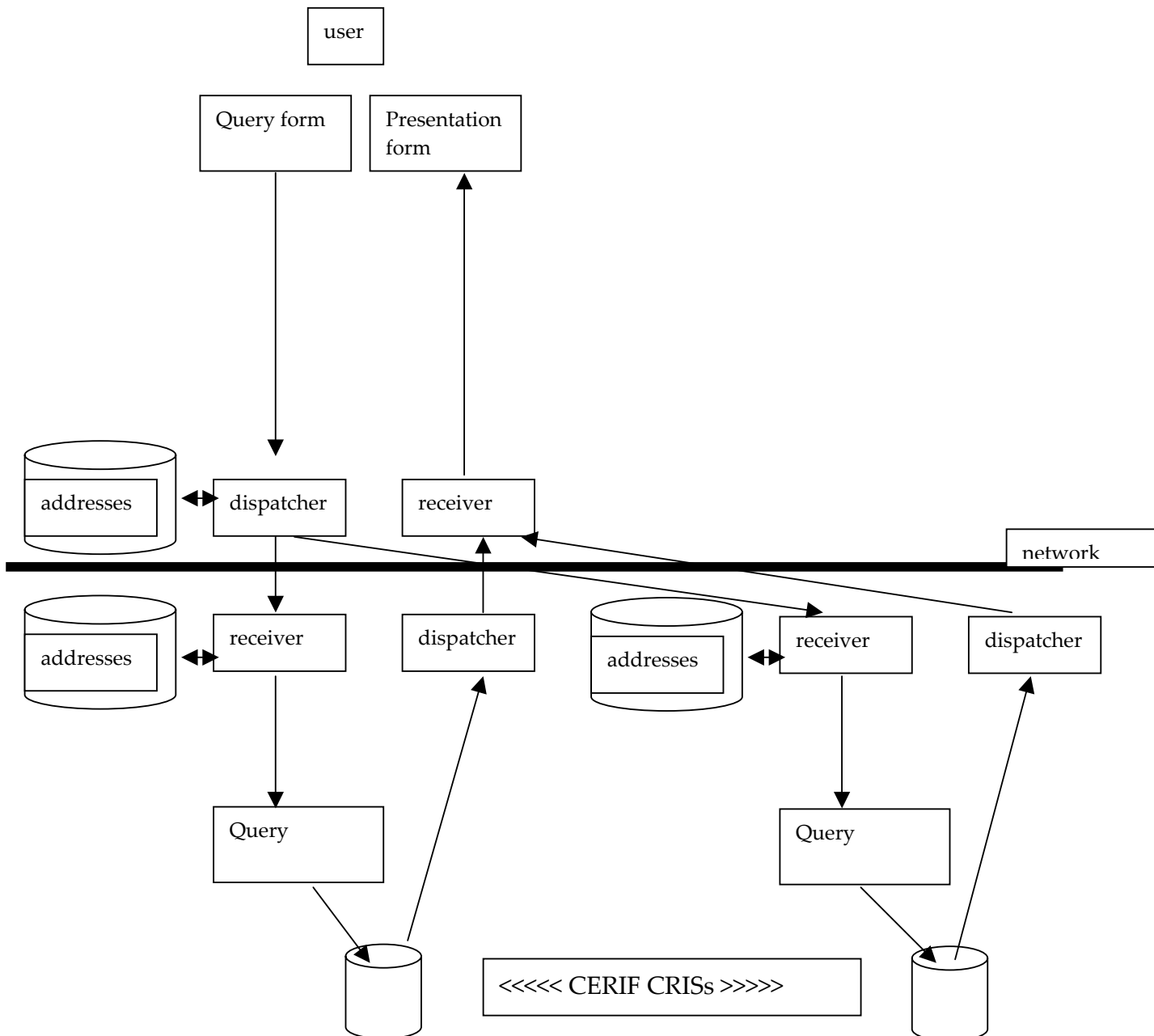
5.2 Description

This architecture relies on the existence at each host of a full CERIF model database, either as the host database itself or a version of the host database converted to full CERIF model. This provides a completely homogeneous solution which is very simple.

5.3 Metadata

This architecture uses navigational metadata for host addresses. No other metadata is required as homogeneity is achieved through the full CERIF model.

5.4 Process

The process is straightforward; through a webform the end-user queries (knowing the CERIF schema) and using normal distributed database technology the query is passed to all hosts; the answers are all in CERIF form so integration is automatic.

5.5 Advantages and Disadvantages

(a) very simple and easy to use for the end-user

(b) each host has to either run a full CERIF model database or provide a full CERIF model version of the host database

6. HARVESTING

The concept of harvesting information from the whole WWW has been introduced. The power of modern search robots (to construct the catalog) and search engines (to search the catalog) is quite remarkable. However, much information is unavailable to harvesting being hidden in databases which may have a webform for query but which do not expose their information on webpages. Furthermore, the search robots usually take around 2 weeks to search the web and so the catalog is not up-to-date. A CRIS harvesting system should be more specific than, say, Google; this implies it searches only URLs known to be entrances to CRISs. The architecture (with catalog and reference to more detailed information) is not unlike the Catalog architecture of ERGO, but ERGO is based on structured data searchable under entity and attribute.

6.1 Architecture

6.1.1 Construction Phase

Catalog of documents
with associative
descriptive metadata

Crawling robot

Html pages

converter   converter   converter   converter

CRIS   << non-CRIF CRISs >>

6.1.2 Search Phase

User phase1

Query form

Query

User phase2

Hit list
processing

Harvester
Associative
descriptive
metadata
catalog

addresses

dispatcher

receiver

network

addresses

receiver

dispatcher

addresses

receiver

dispatcher

URL
query

URL
query

Html pages
from CRIS

Html pages
from CRIS

6.2 Description

This architecture relies on fast indexing of visible web pages by a search robot generating an associative descriptive metadata catalog which is then searched by the user; hits are followed up with a click on the URL to make available the detailed original web page.

6.3 Metadata

This architecture uses associative descriptive metadata in a catalog derived from the search robot and used by the search engine and navigational metadata for host addresses.

6.4 Process

First a search robot traverses the web; it may be instructed to search for certain terms but more likely is general. It constructs an associative descriptive metadata catalog as it goes, usually one entry per web page visited; the catalog also includes navigational metadata: the URL of the webpage indexed. This implies that any CRIS has to provide a set of web pages replicating the data in the CRIS to make it available to the search robot. Techniques are emerging to make structured or semi-structured databases visible to robots but there is no generally accepted technique yet.

The user then queries the catalog, and for every hit (meeting the search term(s)) receives a metadata record; clicking on the navigational metadata (URL in the metadata record) provides the original webpage.

6.5 Advantages and Disadvantages

(a) the host has to provide a copy of the database as webpages to be available to the search robot and subsequent accesses based on clicks from URL of metadata

(b) the query is based on existence of term(s); constraining by entity or attribute is not possible (without sophisticated xml form processing)

(c) the results are unstructured and one page at a time (click on URL in metadata catalog to see page); this inhibits statistical processing or report generation

(d) easy to implement and maintain (although may be ~2 weeks out of date)

(e) familiar interface for many WWW users

7. CONCLUSION

Clearly a full CERIF architecture provides maximum homogeneity and ease of use. However, it requires all hosts either to have their CRIS in CERIF or to provide a CERIF compatible version of their CRIS and make that version available to the federation system.

CERIF can, with advantage, be used as the canonical model for conversion from other CRISs when integrating using either remote or local wrapper techniques. It reduces the (n*(n-1)) interconversion problem to (n), where n is the number of participating CRISs.

CERIF metadata provides structured query capability in the catalog model(s), distinguishing this technique from harvesting.

Under any efficient architecture, CERIF remains the core technology for homogeneous access to heterogeneous CRISs.

ANNEX1 Metadata

Jeffery, K G: 'Metadata': in Brinkkemper,J; Lindencrona,E; Solvberg,A: 'Information Systems Engineering' Springer Verlag, London 2000. ISBN 1-85233-317-0.

ANNEX 2: CERIF

http://www.eurocris.org/   'about' ==> 'taskgroups' ==> 'cerif'

ANNEX3: EXIRPTS Protocol

Naldi F, Jeffery K G, Bordogna G, Lay J O, Vannini-Parenti I
A Distributed Architecture to Provide Uniform Access to Pre-Existing
Independent, Heterogeneous Information Systems
RAL Report 92-003

Remarks to the ERGO working group :

1.) Please check this draft, it is not yet short enough.
   Mark things for deletion/shortening.

2.) Please check also, if this draft is clear enough,
   and do suggest modifications where appropriate.

Kind regards, Bernd Niessen

DRAFT

European Research Gateways Organisation (ERGO)

Final Report to the
INNOVATION PROGRAMME COMMITTEE

(Primary Part)

Version 1.0

17/2/05 13:56

ERGO (15) Final.doc

Table of Contents

8. 1. EXECUTIVE SUMMARY

A clear need for assisted uniform access to heterogeneous research databases is confirmed as predicted in the Terms of Reference.

On the basis of in-depth analysis of

- the perceived user requirements
- existing partial solutions both on organisational and technical levels
- existing information technology
- existing considerations on legal and political aspects

the working group considered

- the spectrum of possible models
- within the spectrum, variants of the optimal (catalogue) model.

The proposed solution combines elegant use of modern information technology (without special research components required) responding to the user needs and information providers' interest, with particular attention being paid to political, security, privacy aspects, as well as de iure or de facto standards. Other aspects, such as copyright and legal, have been considered in the light of EC directives.

The working group recommends an ERGO solution according to ERGO Proposal 2 , details see in chap. 6, which costs about 3.4 MECU and should be available to users in Member States about two years after the favourable project decision by the IPC, i.e. 13 months after project start, hence by 1999-03-30. This project will cover ERGO Phase 1, lasting for three years.

In addition to the above costs, 5 staff*days will have to be invested by the Database Providers for each database accessed via ERGO.

9. 2. INTRODUCTION

As directed by the Innovation Programme's work programme, to :

- facilitate access to national scientific and technical information services;
- identify the main existing information sources and to evaluate access possibilities and the potential for the utilisation of these sources at European level;
- develop the network bridges and instruments permitting access, on a harmonised basis, to these sources from other countries in the Community;

it is intended to set up a project on how to launch and operate gateways to Member States R&D databases. This is called ERGO, **E**uropean **R**esearch **G**ateways **O**rganisation.

10. 3. ERGO TARGET MARKET

10.1 3.0 Basic actors and their roles

The Target User (TU)

Role: The consumer of the information and, of course, the main target of the ERGO project.

The Data Provider (DP)

Role: Any Organisation (e.g. University, Public/Governmental Institution, Private Industry) with managerial/decisional autonomy, which funds research and collects and makes available information on research activities.

The Database Producer (DBP)

Role: The Organisation which makes available the information coming from the Data Provider.

The Data Nodes (DN)

Role: Each site which hosts the available full information. It is usually the same site where the original database. is hosted. It is possible that several Data Nodes are linked to the same Catalogue Node.

The Catalogue Nodes (CN)

Role: The sites which host one of the replicated copies of the ERGO catalogue, and makes the service available to the end user (Target User).

The Reference Node (RN)

Role: Technical point of reference for the ERGO system. Provides monitoring of the services, to ensure that updates sent from Catalogue Nodes (CN) are propagated to the other Catalogue Nodes.

The ERGO Management Board (EMB)

Role: Management and monitoring the strategic evolution of ERGO

The ERGO Task Force (TF)

Role: The technical management of the ERGO service

10.2 3.1   Target Users categories and user requirements.

Since 1995 the ERGO activities form part of the INNOVATION-Programme activities, and therefore the following users groups should be considered as first priority users [1]:

- Innovators in industry
- Researchers in research institutes
- Service and advisory organisations active in the field of research and technological development. (as Innovation Relay Centres)
- Information intermediaries
- Policy Makers and R&D Planners
- and to a certain extent, the Media

For these user groups, ERGO must meet the most relevant requirements within the limits of affordable effort required from database providers.

The ERGO **user interface** must be easy to use for both occasional and regular professional users. Therefore, standard Web browsers will be supported. It is desirable that the power needs of expert users are not overlooked.

During the initial Phase 1 of ERGO, the effort will be restricted to information on "research projects". The Council Decision refers to "national scientific and technical information services" as a whole.

10.3 3.2   Simple but viable charging

At present, many of the existing research databases in various countries are freely accessible. However, there are others in certain countries for which a fee is required from the users. ERGO will include a technical mechanism for charging the users (including monitoring). For full services, users will have to pay a flat subscription rate.

10.4 3.3   Reflections on the issues of data language and query language

ERGO will contain data in both the original database language and the commonly used reference language of the research community. Provided that funding will be available, ERGO will offer, as an extension, machine translation for query and research results in other languages of the European Union.

---

[1] The Council decision mentions as main target groups : enterprises, innovation supporting structures, technology providers and programme managers.

11. 4. SITUATION TODAY

11.1 4.1 Inventory

The working group analysed a number of existing inventories of relevant information sources. The following summary gives an outline on available candidates for an immediate integration into ERGO, which have been confirmed by the members of the working group.

Summary (### countries, ### DBs per country, types of DBs, etc.)

TO BE INSERTED

11.2 4.2 Relevant Projects

The working group identified many projects which had some component of relevance to ERGO, *including the complementary project on EuroCRIS Code of Best Practice*. These components were analysed and have been included where appropriate.

11.3 4.3 Relevant Existing Models: comparison

The working group considered both centralised single and distributed multiple models and recommended a distributed multiple approach. This allows for a high degree of European harmonisation without infringing the optimal solutions within individual Member States.

Five different architectural models have been analysed, reaching from a simple to implement Web Hyperlink approach leaving all the complexity to the user, over a catalogue based model solving the majority of the problems in an elegant way, till a full knowledge based model with high system investment and great user benefit.

The following diagram illustrates the models to building gateways evaluated by the working group.

WebLinks	SF	EXIRPTS	Q2	MIPS

user effort   system investment
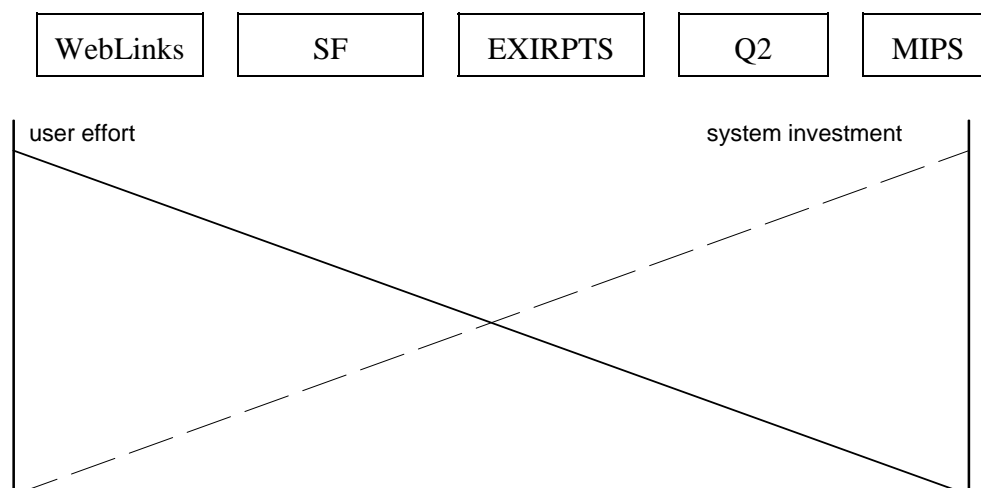
The "**WebLinks**" model is based on a World Wide Web oriented solution characterised by hypertext links which help users to connect to other services. It leaves all the complexity of user

interfaces of these different services to the users.   Complexity means different in user interfaces, database structures, coding schema, retrieval language, data language, display formats, etc.   The different services are located in different countries. The only pre-condition for information access lies in the knowledge of the web location.

'SF' refers to the Search Fields Gateway solution, which has some advantages over the "WebLinks" Model but it does not provide true integration of information.

'Exirpts' refers to the catalogue based solution originally pioneered in the G7 Exirpts project. This prototype allows to assess current research projects funded by the participating institutions. Grace to its common catalogue, users benefit from a common structure, format and contents including one common user interface. Having been initialised by the top management level, acceptance by the participants was trivial. Technical implementation started immediately. The lesson we can learn from this project tells us that the most critical success factor is: Availability of data, which comes with regular updates from each participating host.

Only when all relevant updates become an essential part of the administrative procedures of each database producer, viability and success of such a service can be assured.

Q2 and MIPS are both Knowledge-Based Assisted Solutions (KBS).

Q2 model relies on the total and up-to-date emulation of terminal sessions to each and every database host being connected to this service.   This concerns all levels of interfaces, such as telecommunication protocols, user interfaces, query languages, screen layout, structure of databases, coding, languages, etc.

The critical success factor lies in the complexity of the maintenance for continuous adaptation of all these interfaces to the database hosts. For these reasons, implementation of this model is limited to connecting only a small number of database hosts.
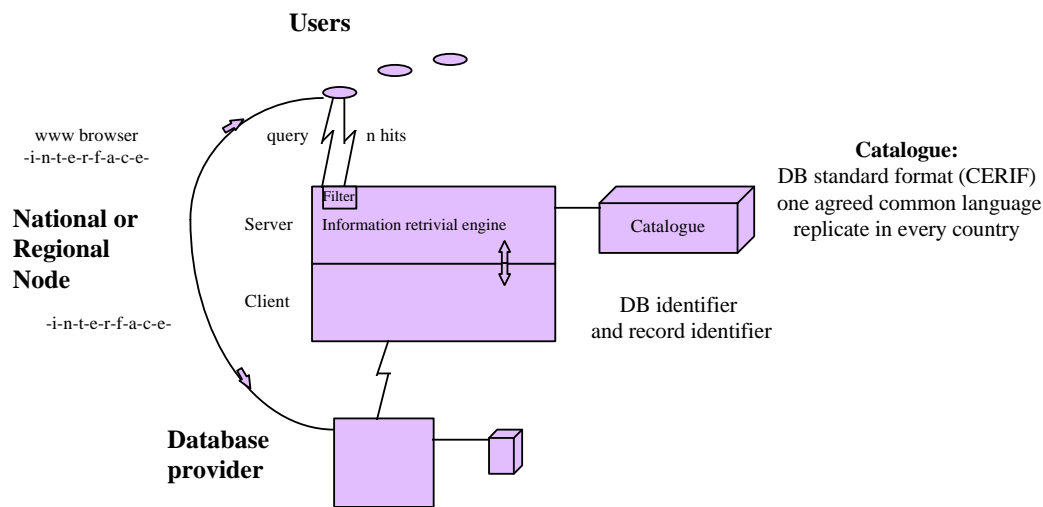
Q2 refers to the EC JRC ISPRA directed project. It should be noted that Q2 has problems when integrating more than a few target host databases.

'**MIPS**' refers to the 'Knowledge-Base Assisted Solution" demonstrated as a pilot system in the application domain of tourism by the ESPRIT MIPS project (ESPRIT project number 6542).

11.4 4.4   Recommendation of Catalogue Model

On the basis of cost / benefit, and taking into account non-functional requirements, the working group recommended the adoption of the catalogue model as being the basis for detailed proposed solutions. The advantages of this model are flexibility, performance, viability, adaptability. It combines the advantages of negotiated logical centralisation with physical distribution management and control.

11.4.1.1 The Catalogue-Based Model

Users

www browser
-i-n-t-e-r-f-a-c-e-

**National or
Regional
Node**

-i-n-t-e-r-f-a-c-e-

query   n hits

Filter
Information retrivial engine

Server

Client

Catalogue

Catalogue:
DB standard format (CERIF)
one agreed common language
replicate in every country

DB identifier
and record identifier

**Database
provider**

Filter: The filter allows each Database Provider to decide what information is provided.

A system with the architecture represented in the diagram will handle requests according to the process outlined below:

a) The user interface is a WWW browser with forms capability (and possibly Java for enhanced end-user dialogue control and help)

b) The user enters the query on the form

c) The query is handled by the regional node

d) The query engine expects the query in a commonly agreed reference language and the query parameters presented in a standard form against the catalogue schema, such as CERIF

e) The number of matches in the catalogue is the number of hits

f) The query use the CERIF information in the catalogue as provided by the data provider; it is the decision of the data provider how much information to make available in the catalogue and how much to retain for (paying) access

g) The list of hits is presented to the end-user together with the rules and costs of requesting full information from the data providers; the user may abandon the query or agree to proceed (and pay)

h) For each hit there is a Unique Identifier (UID): this consists of the catalogue node id, the data provider id, the database id and the unique record number within that database.

i) The hits for a particular region are sent to that regional node as a message consisting of records of originating regional node id, query id, UID. It accepts UIDs, converts them into queries for the target data provider databases, wraps them in standard protocol for Email and sends them to the data provider databases

j) The data provider databases return (Email) to the regional node the documents (one per UID) as answers.

k) The regional node assembles the answers from its region into a folder of documents and send (Email) to originating node

l) The originating regional node accepts all the folders and assemble them into a complete portfolio for ending to the original querying user.

Clearly the regional node will maintain transaction logs. This will allow recovery actions to be taken on time-out, lack of delivery etc. The regional node can also implement access and security policies consistent with local legal requirement and the commercial requirements of their database providers. Regional nodes may implement value-added facilities including translation, additional query assistance, facilities for hyper linked multimedia display....

The regional nodes will require:

a) adequate computing power and data storage

b) excellent networking connections

c) skilled technical staff

d) agreements with regional data providers

Benefits

1) the architecture insulates the end-user from changes in data provider systems

2) the architecture provides for a free, open market system where data providers can choose what they offer at what price and users can choose for what they will pay at what price

3) the architecture insulates the data provider from many changes to the system; only provision to the catalogue of CERIF compatible records and Email access for query and answer to the database are required

4) the architecture allows regional nodes to negotiate with their data providers and allows local policies on payment and security

5) the architecture allows for development with more advanced techniques towards the "Knowledge-Base Assisted Model".

12. 5. ALTERNATIVES WITHIN THE CATALOGUE MODEL

The Working Group decided ERGO to be based upon a catalogue model. Within this proposed model, the Working Group elaborated a so-called "staged approach", in which several alternatives are proposed, where each alternative always is built on the inferior one. The different alternatives, explained in detail in the next session, therefore are called "Steps". 6 steps are provided, of which 3 are selected to be reasonable solutions. For these three alternatives, there is still the choice between different optional extensions.

There are two technical solutions, centralised and distributed. Each has separate advantages and disadvantages. There is clearly political advantage in a distributed solution, which, combined with the technical advantages, led the working group to the solution recommended.

One of the nodes will be managed by the Commission (e.g. CORDIS) and serve ERGO as the reference node.

12.1 5.0   Staged Approach ERGO Steps I to VI

The working group deliberately chose a series of evolutive steps providing progressively increased levels of information delivery and enduser comfort.

Each step builds upon the previous one but requires more resource for development.

12.2 5.1   Description of alternative ERGO solutions, following a staged approach

12.2.1 5.1.1. *Basic principles of the proposed staged approach*

The basic principles followed by the ERGO Working Group in designing a range of alternative concepts to facilitate access to R&D-databases in the Member States are:

- In Phase 1, ERGO will handle only R&D-projects databases.

- Start with a useful and affordable step in, which is already in line with the ERGO mandate.

- Each additional step brings additional value to end users.

- Keep the service / system flexible for enhancements, such as
  - multilingual versions
  - variable number of nodes
  - further types of databases

- Make best use of existing market products : i.e. do not invest in the development of tools, such as:
  - machine translation
  - electronic invoicing
  - electronic indexing.

As a matter of fact, ERGO is only concerned with publicly accessible information in the Member States' databases. Member States can also keep on serving their national users, while ERGO permits common access to EU-wide R&D-information.

12.2.2 5.1.2. *Schematic overview of staged alternatives*

The description of alternative solutions below is based on the gradual meeting of the needs of the ERGO users.

The NBOI-directory of R&D databases can be considered as a zero step.

Schematically, the solutions can be pictured as follows :

Table 1 : Staged ERGO Alternatives

Step	Aimed result	Proposed solution
I	To know which hosts have relevant information on reply to certain query	Catalogue is used for basic retrieval, without delivering information
II	To get freely available information from catalogue	Catalogue is also used for information delivery (as much as authorised by the host)
III	To get charged information from each host of interest	ERGO initiates delivery of up-to-date relevant information from host (user pays directly to each host)
IV	To get charged information from each host, centrally delivered by ERGO	Deliver up-to-date relevant information from host directly via ERGO (single point of sale)
V	To get all the information interactively, during a single user session	Like step IV, but includes also interactive document delivery service
VI	To get other than textual information, if available	full multimedia document delivery (when available at source)

At any of these steps, following the state of art and the evolution of related technologies, further extensions might be considered such as shown in the diagram below, that positions the different gateway system models in the context of the ERGO stages approach.

Table 2 : Possible Extensions based on available technologies

Extension	Aimed result	Proposed solution
A	Delivery of information in language preferred by user	Introduction of a viable machine translation system
B	Electronic charging, in a transparent way for the user, and to minimise the administration efforts	Introduce a system for :
B.1	- automatic invoicing	1) electronic invoicing
B.2	- automatic payment	2) electronic payment
C	User assistance for query formulation (whether or not multilingual)	Intelligent user interface with knowledge based query formulation
D	Indexing ERGO information to avoid the need for classification codes, and to assure a good recall	Introducing an automatic indexing tool

12.2.3 5.1.3. Description of the proposed staged ERGO alternatives : 6 Steps

ERGO Step I : Catalogue of European R&D-projects databases in a common format, as an index for retrieval
"Which hosts have an answer to my question ?"

Description

The user wants to know which of the relevant databases do contain records that are of interest : records responding to his query.

An ERGO catalogue will be built and maintained that contains associated databases and their entries. This catalogue can be used as an index for retrieval.

A possible data model for the catalogue might be:

Minimum data
(those information fields being available for each record and every ERGO user):

- Record Identifier
  (Database Provider, Database, Record Control Number within the latter)

- Update Date

- Title, in the commonly used research working language

- Title, in the original language (without indexing)

- Abstract (which might be kept invisible for fee based databases, but should be available for indexing and query purposes) and/or keywords, in the commonly used reference language

- Contact address and person
  either the project director or principal investigator, or an intermediate contact (e.g. the database provider, in case of fee-based databases)

- Starting-date

Optional data elements:

- Keywords or Abstract - Classification codes

- Ending date

- Project funding

- Other additional CERIF data elements

In Step I, this catalogue will be used only to identify those databases responding to the user query, together with the number of hits in each of those databases.

If no further tools are provided, the user will already be able to make a selection of the hosts with information responding to his query. To receive the records of his interest, the user must then take contact with the selected database providers or hosts.

Efforts and costs

For the user, the catalogue is a free tool to help to identify relevant database providers and their information.

In case of a simple centrally accessible catalogue as described in Step I, no arrangements are required between database providers and the central ERGO services on registration and charging. All contacts between users and providers are bilateral and directly between user and data provider. The catalogue is only a tool to help the user to select the providers of his interest.

So, the user will have to handle the distinct access agreements per database provider / host. Also, the user will have to get successive access to each selected database, and to handle the separate registration and invoicing.

To build up and maintain such a catalogue, regular uploading of the common dataset from the selected R&D-projects databases is to be organised by the host organisations. This will be mainly a one-off effort to set up the uploading procedures, which can be launched at regular interval.

In this case, the required user effort is quite high, and also the provider has to invest much labour and effort to satisfy the ERGO clients.

On the other hand, this step might be of importance for hosts/providers to market their databases via ERGO. The ERGO Working Group wants to avoid databases to join the ERGO system only until this step, since the real ERGO objective is to facilitate user access to the databases.

The central cost for the basic catalogue solution, that allows for simple retrieval only, is relatively low.

However, to have a well working retrieval system, the knowledge based query formulation (Extension C, see below) might be included, by which the user would be assisted in the retrieval process by a knowledge based system. Additionally, a multilingual retrieval system is one of the options.

The introduction of these tools will seriously influence the cost for this alternative.

ERGO Step II : Catalogue with selective data delivery from the catalogue
"Which data can be delivered - for free - by the catalogue ?"

Description

- All data of Step I are in the catalogue.

- The index files of the catalogue are used for retrieval of all relevant items, independently from the question if these fields can be viewed.

- The catalogue allows for interactive viewing of all the catalogue information, that is authorised by the relevant host.

Efforts and costs

ERGO should make arrangements with each of the Database Providers, which of their information will be given to the users for free.

The user will still have to make the effort to contact each of the hosts from which he wants to get full information, if not available for free. For these commercial databases, the user will still be confronted with the distinct access agreements and will have to handle the separate invoices.

From the user's point of view, the major limitation of this stage is that the user has to contact individually each database provider to get full information.

However, at central ERGO level not much extra effort and costs are involved, in comparison to Step I.

ERGO Step III : Initiate delivery of relevant information from Database Providers
ERGO as a contact point to the database providers

Description

Additional possibilities to those in Step II :

- ERGO sends Email to the user-confirmed Database Provider(s) for document delivery.

- Database Provider delivers information to user.

- User pays directly to each host, i.e. host sells to user.

Efforts and costs

At central level, a feature should be integrated to forward the user query to the selected hosts (database providers). This will bring a considerable extra cost to ERGO.

From the users' point of view this step is much more convenient than the previous one, since ERGO contains a central information ordering system, initiated by the request of the user. A weak point for the user is the charging administration.

For many of the Database Providers this step will require a big effort, mainly in building a system which allows to deliver automatically the information asked by the user based on his query. This effort is mainly a one-off cost.

Database Providers will still have to handle the charging administration with each of the user requests.

ERGO Step IV :
Direct (not interactive) delivery of information from Database Providers to Users

Description

Additional possibilities to those in Step III :

- ERGO fetches up-to-date relevant information from hosts and brings it directly to end user via Email. This does not work interactively, but full information is emailed after a time delay.

- User pays to ERGO : single point of sale.

Efforts and costs

From the users' point of view, this step is much more comfortable than the previous one: only one point of sale is foreseen.

Also for the providers, this looks much more attractive, since they do not have to deal with the charging administration for each information request coming through ERGO.

The introduction of an electronic invoicing and payment system is recommended (see extensions B.1 and B.2 below) and will be applied as soon as technology becomes available.

The extra cost at ERGO level will be considerable.

Remark

A possible disadvantage recognised by the ERGO Working Group is that some of the hosts would prefer to keep a client file, while this step does not provide any direct contact between the user and the database provider. For example, in the case, where a certain host has also other databases or interesting products available, such a client file is important for promotional actions about their other products.

A solution to this problem might be a flexible concept, by which providers may make the choice to join at Step III or Step IV.

An alternative could come with a central registration system integrating all users of each database provider being associated with ERGO.

ERGO Step V :
Interactive delivery of information

Description

Additional possibilities to those in Step IV :

- As an additional user option, ERGO fetches up-to-date relevant full information from user confirmed database providers and brings it directly to end user during the same interactive user session.

Efforts and costs

This step will provide a comfortable system to get direct access to each of the databases, with a single user friendly interface. For the users, this system will be the easiest to use.

The integration of an electronic charging system will be necessary in this step (cf. extensions B.1 and B.2 described below).

This solution will require high investments at the central ERGO level, but only once at the database producers' level.   As soon as the ERGO-to-Host interface will be established, these host organisations will be insulated from direct ERGO user access.

ERGO Step VI :
Full multimedia document delivery service, when available at source

Description

In addition to bibliographical descriptions, ERGO fetches - when available at source - relevant multimedia objects from hosts and brings them directly to end user.

ERGO follows those hosts offering hypertext links from their bibliographical information to full multimedia documents, and offers them to ERGO users.

Efforts and costs

Will depend on the technological evolution in this field.

12.2.4 5.1.4 Description of the ERGO extensions to consider

The ERGO extensions as described below, depend on both the state of the art of the technologies and availability of market products for machine translation, multilingual retrieval and electronic payment.

ERGO extension A :
Machine translation

Description

Use of machine translation systems should be considered for two different purposes:

- user assistance during query formulation

- rough translation of information delivered to the user

If suitable machine translation systems will be available, this would give large benefit to the user.

ERGO extension B :
Electronic charging (invoicing & payment)

Description

Since the charging aspects in the above described steps are most time consuming for both users and ERGO, the introduction of an electronic invoicing and an electronic payment system would be most welcome.

Two distinct tools are required to facilitate charging :

B.1 : an electronic invoicing tool

B.2 : an electronic payment tool

The working group expects such tools to come on the market at low cost, soon.

ERGO extension C :
Knowledge based query formulation

Description

ERGO helps user to formulate the query. For this purpose, an intelligent user interface is offered which provides expert knowledge in the specific research area for best quality service response in terms of:

- **Recall** (within the scope of the EU-databases considered) :
  All existing and pertinent information will be hit, even when it is only indirectly related to the subject

- **Relevance** :
  Only relevant information will be brought to the end user.

Description

Additional features for automatic indexing of the catalogue are investigated. These systems (an evolving market) help avoiding complex classification schemes, manual keywording, and a thesaurus with controlled terms.

There is development underway, based on an advanced **semantic language**[2] system and statistical analysis of the records, by which in the ERGO catalogue might be indexed automatically.

Such a tool might be related to the Knowledge based query system, based on the same semantic language.

12.3 5.2   Cost, Effort and Benefit analysis

Tables and figures

12.3.1 5.2.1 *Overview of relevant elements related to benefit, efforts and costs for the distinct actors in ERGO*

The following aspects of

1. user benefit

2. provider benefit

3. provider effort

4. ERGO effort/cost

have been considered in detail.

Four groups of possible elements of benefit, efforts and costs are grouped following categories:

a. General aspects (concerning promotion, transparency, ease of access)

b. Query formulation and handling

c. Information delivery

d. Contents of delivered information

---

[2] Semantic language might be compared with a thesaurus, but different terms might be mentioned in different contexts and also synonyms and related are embodied. This tool should be multi-lingual, at least covering the main EU languages used in the field of R&D.

e. Administration

These elements are used in the following comparison of possible alternative solutions to construct ERGO.

12.4 5.3 Proposal of alternatives for feasible ERGO solutions

12.4.1 5.3.1 *Comparison of the proposals based on the staged approach : Scaling the benefit and efforts in the proposed solutions*

The proposed solutions are compared by "scaling" the users' and providers' against the ERGO and providers' benefits and efforts. The purpose of this scaling exercise is to visualise the advantages, efforts and costs associated with each of the recommended options. The parameters used for scaling reflect the following aspects:

- general (promotion, transparency, ease of access)

- query formulation and handling

- information delivery

- content of delivered information

- administration.

<<< Short summary of tables and a figure, by LVW >>>

12.4.2 5.3.2 *Topology of the network and interconnection of the ERGO actors*

Fig. 1 shows the general topology of the network and how the actors are interconnected.   The   arrows   show   the   functions   carried-on   by   the Organisation/Institution.
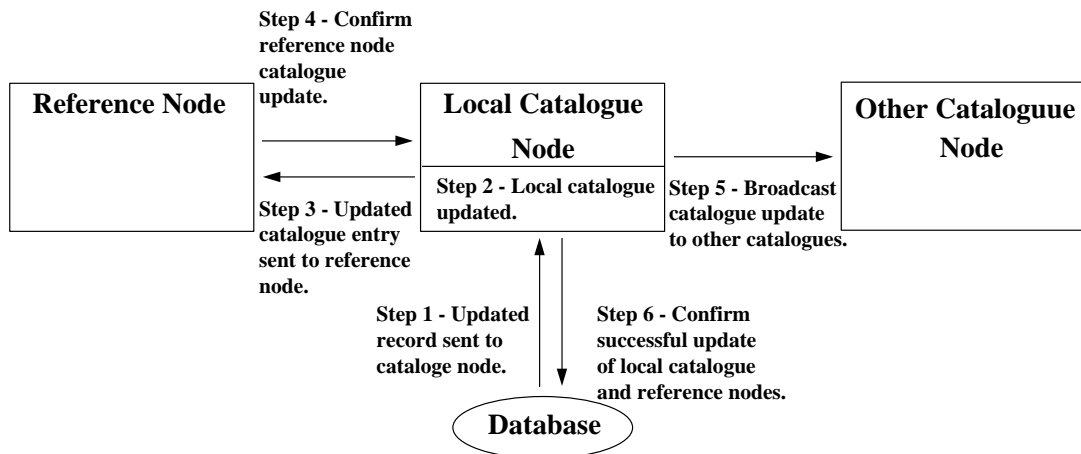
Figure 1 - Catalogue Update Process

13. 6. ANALYSIS OF 3 ALTERNATIVES

The working group considered in detail three alternative proposals, for which design, benefits and costs have been elaborated. According to a staged approach, the following steps are proposed as appropriate alternative solutions.

	User perception	ERGO perception
Step III	Get charged information from each host of interest	Initiate delivery of up-to-date relevant information from host (user pays directly to each host)
Step IV	Get charged information from each host centrally delivered by ERGO	Deliver up-to-date relevant information from host directly via ERGO (single point of sale)
Step V	Get all the information interactively, during a single retrieval session	Like step IV, but also with interactive document delivery service

Due to the characteristics of the opted staged approach, the realisation of a higher step includes implicitly the implementation of all the lower steps facilities.

Critical success factors have also been regarded.

13.1 6.1   Three ERGO proposals

13.1.1 6.1.1   ERGO proposal N°1 :

<div style="border:1px solid">

Initiate delivery of relevant information from host(s),
ERGO as a contact point to the hosts

</div>

13.1.1.1 6.1.1.1   *Technical proposal and options*

The main components of proposal 1 are shown on the diagram on the opposite page. The features of the proposal are :-

User Interface is based on a standard WWW browser. User queries the catalogue using a form which supports Boolean and text searches, on-line help is provided to assist query formulation. Catalogue records which meet the users search criteria are displayed on the user's screen, the user can then refine the search criteria or select a record for delivery from the database producer. Selected records are delivered by Email directly from the database producer to the user; payment issues are resolved between the database producer and user.
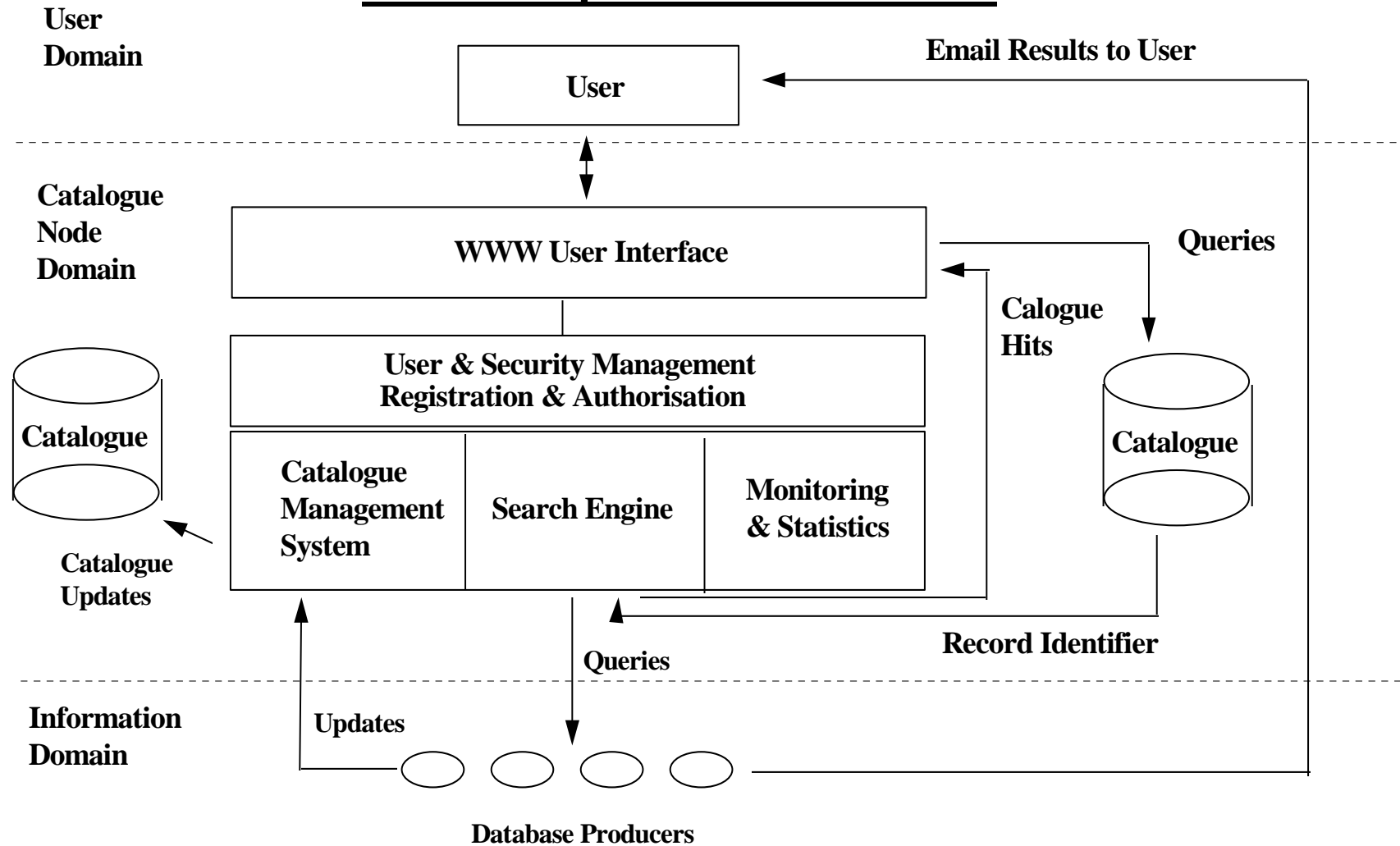
User management and statistics collection facilities are provided to allow monitoring of system usage and also to control of access to data. Facilities are provided to allow users to register on-line.

Catalogue management system provides facilities to update the catalogue and to distribute the updated catalogue records to the reference node and other catalogue nodes. The diagram below shows the sequence of steps followed when an update is received from the host database.

Search Engine is responsible for searching the catalogue for entries which meet the users search criteria. If a user decides to select a record from one of the data node databases the search engine is responsible for generating the query to select the record in the database. It should be noted that in this case the selected record is sent by Email to the user.

Extensions to this proposal include the implementation of intelligent query formulation and context sensitive help in the user interface.

ERGO Proposal 1 - Architecture

User Domain

Catalogue Node Domain

Information Domain

User

Email Results to User

WWW User Interface

Queries

User & Security Management
Registration & Authorisation

Calogue Hits

Catalogue

Catalogue

Catalogue
Management
System

Search Engine

Monitoring
& Statistics

Catalogue Updates

Record Identifier

Queries

Updates

Database Producers

13.1.1.2 6.1.1.2   *Resources for Phase 1 / Proposal 1*

Item	Human Resources	Budgetary Requirements	
Server & local workstation hardware & software for each catalogue node.		15 kEcu set-up costs = 2 kEcu per year maintenance =	30k 12k
Software development for all nodes.	5 staff*years all ERGO nodes. *(Note:Each database provider 5 staff days per database.)*	150 kECU per staff*year =	750k
Operation of catalogue node	1 staff*year per year	150 kECU per staff*year =	900k

Total ERGO budgetary requirements for Phase 1 =   **1 692 kECU**

In addition to these costs, 5 staff*days will be invested by the Database Providers for each database accessed via ERGO.

13.1.2 6.1.2   ERGO proposal N°2

> **Delivering up-to-date relevant information from host directly via ERGO, ERGO as a Single Point of Sale**

13.1.2.1 6.1.2.1   Technical Proposal and Options

The main components of proposal 2 are shown on the diagram on the opposite page. The features of the proposal are :

User Interface is based on a WWW browser User query the catalogue using a form which supports Boolean and text searches, intelligent query formulation and on-line context sensitive help is provided to assist query formulation. Catalogue records which meet the users search criteria are displayed on the user's screen, the user can then refine the search criteria or select a record for delivery from the database producer. The user has the three options for the delivery of selected records :

Option a :   Each data node providing records sends a separate Email to the user, and also a copy to the ERGO catalogue node.

Option b :   Each relevant data node sends the information via Email to the ERGO catalogue node, which collates these deliveries and sends the full package to the user, on completion.

Option c :   Like option b, but with additional feature allowing the user to preview the status of the collation process.

In proposal 2 ERGO takes responsibility for charging. The charging approach has two levels of complexity, the first level where the user pays on registration and the second where support for on-line electronic payment is provided. The second approach will give the flexibility to support ad-hoc ERGO users.

User management and statistics collection facilities are provided to allow monitoring of system usage and also to control of access to data. Facilities are provided to allow users to register and pay on-line.

Catalogue management system provides facilities to update the catalogue and to distribute the updated catalogue records to the reference node and other catalogue nodes. The update process for the catalogues is identical to the process described for proposal 1.

Search Engine is responsible for searching the catalogue for entries which meet the users search criteria. If a user decides to select a record from one of the data node databases the search engine is responsible for generating the query to select the record in the database. It should be noted that in this proposal, the user has the option to specify that ERGO collects the results and emails them to the user. The diagram below describes the process of retrieving records and then making delivering them to the user..

The processing depicted in the diagram is as follows :-

1) The user enters a query and the search engine identifies those records in the catalogue which meet the criteria entered by the user. The results of the query to the catalogue are displayed to the user.

2) When the user decides that a full copy of a record is required he marks the record on the results list. The marked records are sent by the search engine to a distributor process.

3) The distributor process identifies if the selected record is in database to which the process has direct access or if the record is in a database which is connected to another catalogue node. If the database is directly connected the distributor formulates a query and sends it to the database. If the database is connected to another catalogue the distributor delegates responsibility to the distributor process of the catalogue to which the database is connected.

4) The receiver process is responsible for collecting retrieved records from directly connected databases and from databases connected to other catalogues. If the user preference is to be emailed directly by the database then the receiver only records that a record has been received. If the user preference is for records to be collated by ERGO then the receiver sends the retrieved record to the results collator.

5) The results collator sends one Email to the user when all results have been received. It also provides facilities to allow the user to identify the number of records of the query that have been retrieved.

Querying of Multiple Catalogues

Figure 2 - Selection, Collation & Delivery of Records
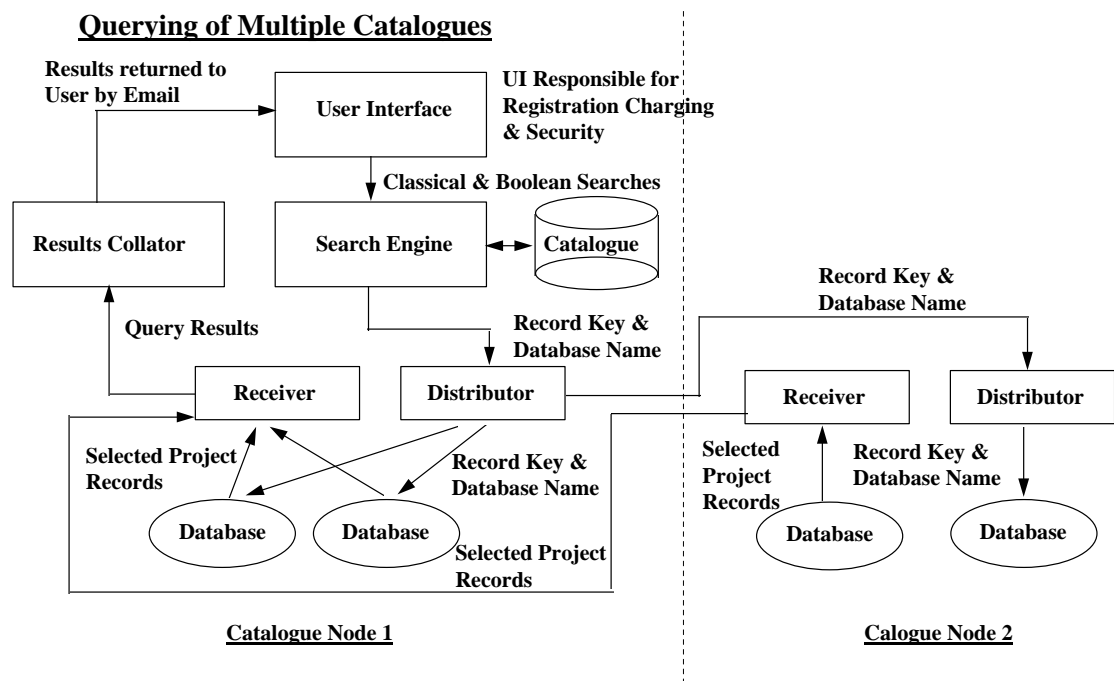
Extensions to this proposal include support of a multi-lingual interface and implementation of machine translation facilities to allow support for multi-lingual queries and catalogue.

ERGO Proposal 2 - Architecture

13.1.2.2 6.1.2.2   Resources for Phase 1 / Proposal 2

Item	Human Resources	Budgetary Requirements
Server & local workstation hardware & Software for each catalogue node.		30 kEcu set-up costs for each node. 60k 4 kEcu per year maintenance for each node 24k
Software development for all catalogue nodes.	10 staff *years all nodes. *(Note:Each database provider: 5 staff*days per database.)*	150 kECU per staff*year 1 500k
Operation of catalogue node.	1.5 staff*years per year	150 kECU per staff*year 1 350k
Machine translation (MT) extension.		400 kECU to adapt software for all nodes 400k

Total ERGO budgetary requirements for Phase 1 without MT =   **2 934 kECU**

Total ERGO budgetary requirements for Phase 1 incl. MT =   3 334 kECU

In addition to these costs, 5 staff*days will be invested by the Database Providers for each database accessed via ERGO.

13.1.3 6.1.3 ERGO proposal N°3

> **Controlled Information Delivery from Database Producers via ERGO to User, allowing for either *interactive delivery within session* or *interactive monitoring* of the ERGO query and collation background process.**

It is understood that all user features of proposal 2 are also available for proposal 3, as alternatives.

13.1.3.1 6.1.3.1 Technical proposal and options

The main components of proposal 3 are shown on the diagram on the opposite page. The features of the proposal are :

User Interface is based on a WWW browser User query the catalogue using a form which supports Boolean and text searches, intelligent query formulation and on-line context sensitive help is provided to assist query formulation. Catalogue records which meet the users search criteria are displayed on the user's screen, the user can then refine the search criteria or select a record for delivery from the database producer. The user has the four options for the delivery of selected records :-

Option a) : Each data node providing records sends a separate Email to the user, and also a copy to the ERGO catalogue node.

Option b) : Each relevant data node sends the information via Email to the ERGO catalogue node, which collates these deliveries and sends the full package to the user, on completion.

Option c) : like option b), but with additional feature allowing the user to preview the status of the collation process.

Option d) : Records retrieved from the database producer are delivered interactively to the user during the current session. The user can also see the delivery status on-line.

The charging approach has two levels of complexity, the first level where the user pays on registration and the second where support for on-line electronic payment is provided. The second approach will give the flexibility to support ad-hoc ERGO users.

Proposal 3 provides multi-lingual support at two levels : firstly, the user interface will be supported in the major European languages, and secondly, the catalogue will be multi-lingual. The multi-lingual aspects of the catalogue will be implemented using machine translation technology.

User management and statistics collection facilities are provided to allow monitoring of system usage and also to control data access. Facilities are provided to allow users to register and pay on-line.

ERGO Proposal 3 - Architecture
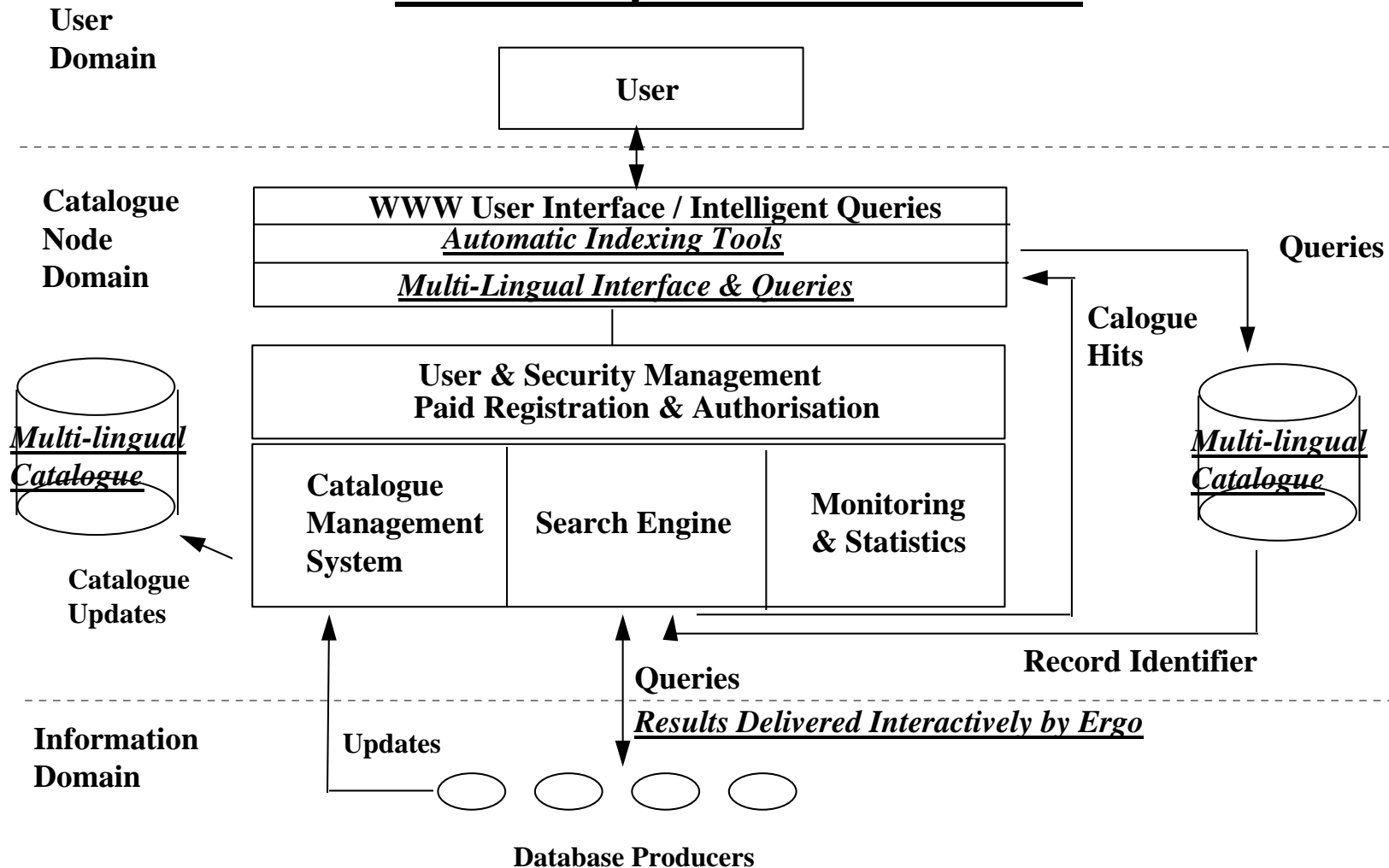
The Catalogue management system provides facilities to update the catalogue and to distribute the updated catalogue records to the reference node and other catalogue nodes. The update process for the catalogues is identical to the process described for proposal 1.

The Search Engine operates in identical way to the approach described in proposal 2 but the results collator can either send the retrieved records by Email or return them interactively to the user's active session.

Extensions to this proposal include the implementation of an automatic indexing tools and a knowledge based user interface..

13.1.3.2 6.1.3.2   Resources for Phase 1 / Proposal 3

Item	Human Resources	Budgetary Requirements
Server & local workstation hardware & Software for each catalogue node.		150 kEcu set-up costs for each node = 300k 30 kEcu per year maintenance for each node = 180k
Software development for all catalogue nodes.	30 staff *years all nodes. *(Note:Each database provider 15 staff days per database.)*	150 kECU per staff*year = 4 500k
Operation of catalogue node	3 staff*year per year	150 kECU per staff*year = 2 700k
Automatic indexing and knowledge based user interface extension.		400 kECU to develop software for all nodes 400k
Machine translation.		*400 kECU to adapt software for all nodes 400k*

Total ERGO budgetary requirements for Phase 1 without MT =   **8 080 kECU**

Total ERGO budgetary requirements for Phase 1 incl. MT =   8 480 kECU

In addition to these costs, 15 staff*days will be invested by the Database Providers for each database accessed via ERGO.

For each of the three proposals, the suggested extended features will incur costs greater than those given above.

13.1.4 6.1.4 *Selection of the most recommendable solutions*

For each of the Steps, some optional choices of extra features are proposed. Briefly the proposed alternatives are:

1. **Step III**, including the adoption of an :

   - automatic indexing tool, optionally combined with :
     - a machine translation tool (tool A)
     - an intelligent user interface.

2. **Step IV**, including tools for :

   - automatic indexing,
   - electronic invoicing and electronic payment,

   optionally combined with :

   - a machine translation tool (tool A)
   - an intelligent user interface.

3. **Step V**, including tools for :

   - automatic indexing,
   - electronic invoicing,
   - electronic payment

   and also an intelligent user interface, optionally combined with :

   - a machine translation tool.

The following table gives an overview on these most recommendable ERGO solutions.

Recommendable ERGO solutions

o = optional	m = mandatory	- = not relevant	proposed solution	extra recommended

extension ➜ ⬇ staged Step	A machine translation	B.1 electronic invoicing	B.2 electronic payment	C intelligent user interface	D automatic indexing tool
Step I : catalogue	o	-	-	o	o
Step II : catalogue + delivery from catatlogue	o	-	-	o	o
Step III : delivery from host	o	o	-	o	o
Step IV : delivery from host, ERGO single point of sale	o	m	m	o	o
Step V : interactively	o	m	m	o	o
Step VI : multimedia	o	m	m	o	o

13.1.5 6.1.5 *Strategy schemes for phases and schedules for each of the proposed alternatives for the gateway*

To be completed after finalising previous chapters

*   *CHART NEEDED WITH PHASED WORKING STRUCTURE, INCL. MONITORING, incl. technology aspects (e.g. use of proven technologies)*

*   *SHOULD INCLUDE WORKING STRATEGY/PROCEDURES*

*** *MAURITZ VANDEGRAAF ?*

INCLUDING OPTIONS TABLE WITH ALSO THE OPTION TO INCLUDE THE ASSOCIATED MEMBER STATES

*   *This chapter will contain a matrix to guide the Committee in the decision taking : an overview of the principal facts for the most suitable solutions on a.o. benefit for the information providers, benefit for the users, risks and costs, and supplementary options (or substeps) which might be chosen.*

TO BE REVISED

============

In terms of user services, the ERGO service is divided into 3 major phases:

Phase 1

The timing of this falls within the time span of the current Innovation Programme.

The objectives of Phase 1 are outlined below :

a) About 30 Current Research Databases will be covered.

b) Data providers are needed from a majority of Member States.

c) Include databases of different languages and both free and fee paying databases.

d) At least two World Wide Web nodes to monitor potential performance problems.

Phase 2

The objectives of Phase 2 are outlined below :

a) Extend number of nodes.

b) Extend information domain to include further types of current research information, such as organizations, scientists, research programs, and research results.

c) Offer access to other classes of users other than WWW users.

d) Extend the functionality of the service in line with user needs, perhaps in the direction of knowledge base assistance.

Phase 3

The objectives of Phase 3 are to extend the information domain to other research data collections of interest to the European user.

This ERGO Report covers mostly aspects relating to Phase 1.

Administrative Schedule

(May 1996:	*Submission of Interim Report)*
(June 1996:	*Positive opinion of Innovation Programme Committee (IPC) on Interim Report)*
Febr. 1997:	*Submission of Final Report to IPC*
March 1997:	*Positive opinion of IPC on Final Report / Decision to proceed with ERGO Phase 1*
Aug. 1997:	*Open Call for Tenders for ERGO Phase 1 contract*
Oct. 1997:	*Closing date for tenders*
Feb. 1998:	*ERGO Phase 1 Development contract awarded and project development commences*
March 1999:	*ERGO Phase 1 service starts operation* *with 2 Catalogue Nodes and about 30 Databases becoming fully accessible via ERGO.*

13.2 6.2   Funding considerations

The working group assumes that the Innovation Programme will fund the first three years of the ERGO project. After evaluation, a different funding regime will be proposed, which will rely on Subsidiarity and local financing of regional / national catalogue nodes.

13.3 6.3   Time scale, costs and planning

The IPC will authorize Phase 1 of the ERGO project in March 1997.

On this basis, an open call for tenders will be launched by the Commission in August 1997.

ERGO Phase 1 implementation starts in February 1998.

Phase 1 of the ERGO service starts operation in March 1999 with 2 Catalogue Nodes and about 30 Databases becoming fully accessible via ERGO.

After evaluation of the results of Phase 1, which is at 24 months after project start, the Phase 2 will be launched to incorporate further national and regional nodes and their associated database providers.

13.4 6.4   Future developments beyond "projects"

Other types of CRIS (Current Research Information Systems), such as organisations, expertise, programmes, results, publications, etc., can be added to the ERGO information environment. These additional and required facilities will be costed during year 3 of the proposed ERGO project.

14. 7. RECOMMENDATIONS

It is therefore recommended that ERGO Proposal 2 is funded by the Innovation Programme by 3 334 kECU and implemented through standard Commission's procedures.