EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

EUROHORCs
EUROPEAN HEADS OF RESEARCH COUNCILS

Window to Science
Information Systems of European Research Organisations

Report of the EUROHORCs – ESF Working Group on a Joint Research Information System

www.esf.org

ESF
The European Science Foundation (ESF) is an independent, non-governmental organisation of national research organisations.
Our strength lies in the membership and in our ability to bring together the different domains of European science in order to meet the scientific challenges of the future. ESF's membership currently includes 77 influential national funding agencies, research-performing agencies and academies from 30 nations as its contributing members.
Since its establishment in 1974, ESF, which has its headquarters in Strasbourg with offices in Brussels and Ostend, has assembled a host of research organisations that span all disciplines of science in Europe, to create a common platform for cross-border cooperation.
We are dedicated to supporting our members in promoting science, scientific research and science policy across Europe. Through its activities and instruments, ESF has made major contributions to science in a global context.

EUROHORCs
EUROHORCs is a European association of the heads of research funding organisations and research performing organisations. Since its establishment in 1992, EUROHORCs has become an active player in the field of European research policy by promoting and enhancing inter-council cooperation and serving, amongst others, as an advisory body to the European Commission. By creating an informal platform for discussion, producing policy statements and initiating joint activities, EUROHORCs seeks to strengthen its influence on European research policy. EUROHORCs uses the European Science Foundation (ESF) as an implementation agency.

Report published by the European Science Foundation, September 2008

Suggested citation:
ESF (2008): Window to Science. Information Systems of European Research Organisations. Report of the EUROHORCs – ESF Working Group on Joint Research Information System. Strasbourg: European Science Foundation.

Report Editing: Alexis-Michel Mugabushaka

ISBN: 2-912049-86-5

Preface and Acknowledgements

In December 2006, ESF was requested by EUROHORCs to convene a Working Group to explore the benefits of linking up the research information systems of EUROHORCs Member Organisations and to identify viable technical options to achieve this.

This report records the activities of the Working Group from the time it was convened in January 2007. It provides an overview of research information systems of EUROHORCs Member Organisations, discusses the added value of a joint research information system and various technical models. It also reproduces the recommendations that the Working Group submitted to EUROHORCs in January 2008.

The Working Group hopes that this report and the data it gathered will be a useful resource for anyone interested in the further development of research information systems. It hopes that the discussions about a joint European Research Information System, recorded in the report, will contribute to a greater understanding of the key technical and organisational challenges and will guide future efforts to realise such a system.

The Working Group would not have been able to accomplish its tasks without the support and contributions of many individuals.

We wish to thank in particular Neil Jacobs, Executive from the Joint Information Systems Committee (JISC), for his substantial input into the discussions on the appropriate model of a joint research system.

The Working Group is also deeply grateful to those officials from EUROHORCs Member Organisations who provided information on the research information systems of their respective organisations.

In addition, the Working Group wishes to thank the participants to the workshop it organised on 27 June 2007 in Stockholm (Appendix 2/3).

Neil Williams,
Chair of the Working Group

Members of the EUROHORCs-ESF Working Group:
Neil Williams (Chair of the WG), ESF
Gerry Lawson, RCUK
Jesper Aven, VR (until August 2007)
Jürgen Güdler, DFG
Ruud Strijp, NWO
Alexis Michel Mugabushaka, ESF

All views, conclusions and recommendations expressed in this report are those of the Working Group and do not necessarily reflect the views of EUROHORCs, ESF or their Member Organisations.

Contents

1. Introduction

Traditionally, research-funding organisations and public research institutions reported to their stakeholders through annual reports and other dedicated publications recording their activities.

Although those publications remain an important – and perhaps irreplaceable – source of information on the activities of the research institutions, in recent years they have turned to the Internet to meet the growing demand for accountability and to serve the information needs of the various stakeholders.

Going beyond the listing of their activities on their Internet sites, an increasing number of organisations have developed complex information systems that provide, in real time, the information on funded or performed projects, on researchers involved and the results of the research.

In the context of increasing European cooperation in research and cross-border research funding, those systems have a potential to facilitate growing interactions between national research organisations.

It is against this background, that the European Heads of Research Councils (EUROHORCs) – the association of major national funding agencies and research performing organisations in Europe – started to explore the benefits and feasibility of a system that would federate the research information systems of its members.

The EUROHORCs General Assembly in its meeting on 10 June 2006 agreed to look at practical aspects of how the systems of its Member Organisations could be linked.

On 11 October 2006 a workshop was organised by the Netherlands Organisation for Scientific Research (NWO) (which at the time hosted the EUROHORCs secretariat) to discuss the linking of national systems in terms of needs and technical solutions. The workshop recommended that EUROHORCs pursue this undertaking and, in December 2006, EUROHORCs asked the ESF Office to "*convene and run a Working Group to draft a business plan on how to link up the databases of EUROHORCs members*".

WORKING GROUP ON A JOINT RESEARCH INFORMATION SYSTEM: MANDATE AND MODUS OPERANDI

The ESF Office invited the organisations that participated in the Hague workshop to nominate representatives to the Working Group and other EUROHORCs Member Organisations with substantive experience in Research Information Systems were approached.

The Working Group began its activities in January 2007 and set out to define the scope of the project. An initial draft of the project outline was submitted to EUROHORCs Steering Committee meeting on 12 February 2007 for guidance. This consultation shows that the case for the need for and benefit of a joint system had not been sufficiently established and needed to be put on the agenda of the Working Group.

The tasks of the Working Group were therefore defined as follow:

(1) To make an overview of existing Research Information Systems;
(2) To assess the added value of a joint system;
(3) To identify appropriate models of a joint system; and
(4) To make recommendation to EUROHORCs on how to proceed further.

The Working Group operated mainly through e-mail exchanges and teleconferences. It issued a questionnaire to all EUROHORCs members (supplemented by a desk search) and organised a workshop to which all EUROHORCs Member Organisations were invited. On 25 and 26 June 2007, the Stockholm workshop, hosted by the Swedish Research Council, was attended by 22 participants from 16 organisations (see Appendix 2 and 3 for the workshop agenda and list of participants respectively).

This report of the Working Group presents an overview of research information systems in selected organisations (Task 1). It also discusses the added value (Task 2) and appropriate models for a joint information system (Task 3). The recommendations (Task 4) on how to build to achieve the goals of a joint research system, which the Working Group submitted to EUROHORCs Steering Committee meeting on 29 January 2008 are reproduced in the Section 4 of this report.

2. Research Information Systems of EUROHORCs Member Organisations

The initial request to ESF to convene a Working Group referred to "*linking up the databases of EUROHORCs' Member Organisations*".

The Working Group interpreted "databases of EUROHORCs members" to mean Research Information Systems (RIS). These are defined as tools that '*provide access to and disseminate research information'. This includes People, Projects, Organisations, Results (publications, patents and products), Facilities and Equipments* (Jeffery & Asserson, 2006)[1]. This definition excludes systems used to record or execute financial transactions, the databases used to manage the peer review of grant proposals, grant application tools and other "internal systems" (whose access is restricted).

Box 1: Data Collection

The data on which this report is based were mainly collected through a questionnaire sent to all EUROHORCs Member Organisations. The questionnaires were sent out on 30 April 2007 and a reminder on 23 May. In addition an Internet search was done. In total, information for 26 organisations could be collected.

Of those 26 organisations, six indicated that they had no research information system at the time but were in a preparatory phase of establishing such a system. Three organisations did not have systems fulfilling the definition criteria. In the activities of the Working Group, only 17 research information systems were considered.

Research Information System (RIS) in this report thus refers to systems:
- Which provide online access to funded projects (by research-funding agencies) or on-going research activities (by research-performing organisations);
- Have as informational entities: projects, programmes, organisations, and (additionally) outputs;
- In which information is organised in searchable databases (simple listings of projects on the homepage do not fulfil this criterion).

The first task of the Working Group was to collect basic information on Research Information Systems from EUROHORCs Member Organisations (see Box 1). Their brief description is provided on the following pages.

[1] Jeffery, Keith and Asserson, Anne (2006). CRIS : central relating information System. In: Asserson, Anne and Simons, Eduard (eds). Enabling Interaction and Quality: Beyong the Hanseatic League. 8th International Conference on Current Research Information Systems (May 2006, Bergen, Norway). Leuven: Leuven University Press, 109-119.

Organisational and national research information systems

Table 1 gives brief descriptions of systems used by 17 EUROHORCs Member Organisations. However, it should be noted that those systems differ in one major way.

Most of the systems described in Table 1 contain information on research projects funded or performed by the institution which operates them.

There are however four systems on the list, which acts as national portals to the research activities of a larger research system: (1) The Flanders Research Information Systems, FRIS; (2) The Estonian Research Information System, ETIS; (3) The Slovenian Research Information System, SICRIS; and (4) The Dutch National Academic Research and Collaboration Information Systems (NARCIS).

The research funded or performed by EUROHORCs Member Organisations are, in that case, only a subset of those national research information systems.

Although the activities of the Working Group focused primarily on the research information systems operated by EUROHORCs Member Organisations (institutional RIS), in this report, reference is occasionally made to the national RIS, as they play a key role in the debates about a joint European RIS.

Examples of other national research information systems not included in Table 1 are:

Bulgaria: BulCRIS, Bulgarian Current Research Information System
www.cris.government.bg

Croatia: Who is Who in Croatian Science
http://tkojetko.irb.hr/en/

Czech Republic: R&D Information System of the Czech Republic (IS VaV)
http://aplikace.isvav.cvut.cz/index.jsp

Hungary: HunCRIS Hungarian Research Information System (http://nkr.info.omikk.bme.hu/)

Ireland: Expertiseireland
www.expertiseireland.com/

Switzerland: ARAMIS, information system for projects of the Swiss Federal Administration
http://www.aramis.admin.ch/

2. Research Information Systems of EUROHORCs Member Organisations

AUSTRIA	**AUSTRIAN SCIENCE FUND (FWF)**
	The Austrian Science Fund (FWF) operates a **"Project Database".** Covering the period 1992 to the present, it lists all projects approved by the FWF. Updated twice a week, the database can be searched using different criteria such as principal investigators, research institutions, key words characterising the research topic and funding scheme.
	www.fwf.ac.at/en/projects/projekt_datenbank.asp
BELGIUM	**RESEARCH FOUNDATION FLANDERS (FWO)**
	The Flanders Research Information Space **(FRIS)** is a portal that supplies information about research activities in Flanders. There are several options to search/browse the system: by research projects, organisations or persons.
	www.researchportal.be/
CZECH REPUBLIC	**CZECH SCIENCE FOUNDATION (GAČR)**
	The Czech Science Foundation (GAČR) has a publicly accessible system called **"Project Database"** that lists research projects supported by the GAČR. It contains a selection of information from the operative database. The database can be searched by the following criteria: supported/completed project, scientific fields and key words. The updating is done regularly. The Czech version of the research information system contains more detailed information about Czech scientists.
	http://pala.gacr.cas.cz/web/Seznam_php-en.htm
ESTONIA	**ESTONIAN SCIENCE FOUNDATION**
	ETIS is the Estonian Research Information System, which includes information concerning research and development organisations, researchers, research projects, and the results of several research activities. At the same time, ETIS is also a channel for submitting and examining various applications and approving the reports on applications and projects. Numerous research and development institutions use ETIS as an internal research information system.
	www.etis.ee
GERMANY	**GERMAN RESEARCH FOUNDATION (DFG)**
	The German Deutsche Forschungsgemeinschaft (DFG) has a system called **GEPRIS** with information on research programmes and projects. The database, which covers the period 1999-2004, can be searched by free texts entry through the title of the projects and the abstract as well as via the record of the principal investigator and his institution address. The data are taken from a database used to process the applications and validated by the researchers themselves.
	www.dfg.de/gepris/
NETHERLANDS	**NETHERLANDS ORGANISATION FOR SCIENTIFIC RESEARCH (NWO)**
	The Netherlands Organisation for Scientific Research (NWO) has a database of research projects. Updated every week, it contains information on 9 000 predominantly ongoing projects in all scientific fields. The database can be searched by free texts as well as criteria such as principal investigator, research institution, funding scheme, NWO division and NWO theme. NWO projects are also included in the NARCIS (National Academic Research and Collaborations Information System) portal. It is planned that NARCIS will be integrated into one system with DAREnet, a search service which gives free access to academic research output in the Netherlands, covering a broad collection which guarantees digital accessibility to the full text without any restrictions.
	www.nwo.nl/projecten.nsf/pages/losearch_eng/ www.narcis.info/narcis/index.php
NORWAY	**RESEARCH COUNCIL OF NORWAY**
	The "Project Database" of the Norwegian Research Council provides an overview of all funded projects since 1997. Principal investigator, research field, institution, funding scheme/programme and year of funding, can search the database. A free text search is also provided. Many of the projects have titles and summaries in English, but most of the entries are in Norwegian.
	http://www.forskningsradet.no (and follow the link to «Project Database»)
PORTUGAL	**FOUNDATION FOR SCIENCE AND TECHNOLOGY (FCT)**
	The Portuguese Foundation for Science and Technology (FCT) has made available data on supported research projects since 1998. There is currently no consistent interface.
	www.fct.mctes.pt/pt/apoios/projectos/bd/

2. Research Information Systems of EUROHORCs Member Organisations

SLOVENIA	**SLOVENIAN RESEARCH AGENCY (SRA)**
	The Slovenian Research Agency has the **SICRIS** information system, which is being developed and maintained by the Institute of Information Science (IZUM) in Maribor and the Slovenian Research Agency (SRA). SICRIS also allows viewing of presentation pages of more than 500 European projects of the EU Framework Programmes directly from the Project Database within the CORDIS system.
	www.sicris.izum.si
SWEDEN	**SWEDISH RESEARCH COUNCIL (VR)**
	VR-Proj is the research project database of the Swedish Research Council (VR). It contains information about all research projects, various grants, positions and fellowships that have been funded by the Council since 2001. The database is continuously updated, and approved proposals are published with the project description and funds. The database has both a full text search area and an advanced search area.
	http://vproj.vr.se
SWITZERLAND	**SWISS NATIONAL SCIENCE FOUNDATION (SNF)**
	The **"Project Database"** of the Swiss National Science Foundation (SNF) contains research projects and other subsidies it has supported since 1957. Updated daily, the database allows a search by name of researchers involved and their institutions, the discipline of the project etc …. A full text search by key words is also possible.
	www.snf.ch/prodb/webforms/frameset.aspx
TURKEY	**SCIENTIFIC AND TECHNOLOGICAL RESEARCH COUNCIL OF TURKEY (TÜBITAK)**
	The Scientific and Technological Research Council of Turkey (TÜBITAK) operates a **"Project Database"**. Covering the period 1997 to the present, it lists all projects funded by TÜBITAK. Updated continuously, the database can be searched using different criteria such as principal investigators, research institutions, key words characterising the research topic and funding schemes. The system has an interface currently in Turkish but an interface in English is planned.
	http://mistug.tubitak.gov.tr/proje/index.php
UK	**ARTS AND HUMANITIES RESEARCH COUNCIL (AHRC)**
	The **AHRC Award Listing** contains information on post-doctoral (research) grants and information on museums, galleries and collections, which AHRC has funded. The database can be browsed via subject/ panel, scheme, institution and searched via project title, award holder, institution and year and holds around 5000 awards.
	www.ahrc.ac.uk/awards
	ENGINEERING AND PHYSICAL SCIENCES RESEARCH COUNCIL (EPSRC)
	The UK Engineering and Physical Sciences Research Council (EPSRC) provides information on the research it funds through its **"Grants on the Web"** System. EPSRC programme, research subject (topic), industrial relevance (sector), socio-economic theme, scheme, UK region and organisation can search the database. There is also a free text search and grant progress check facility. The grant progress checker data is updated daily and all other data is updated weekly.
	http://gow.epsrc.ac.uk
	ECONOMIC AND SOCIAL SCIENCES RESEARCH COUNCIL (ESRC)
	ESRC's **Awards and Outputs Database (AOD)** is the systematic backbone of ESRC Society Today, the internet site of the Council. It contains records from ESRC's managed information systems on individual awards going back as far as 1975. Much of the data was migrated from a predecessor online database called "Regard". At an aggregate level there are over 9,000 awards and around 100,000 award outputs contained on the site – and crucially, they are linked directly to ESRC managed information systems.
	www.esrcsocietytoday.ac.uk/
	NATURAL ENVIRONMENT RESEARCH COUNCIL (NERC)
	Grants on The Web (GOTW) provides information on research grants (and fellowships). It is updated daily and can be searched by the following criteria: Researcher (or the institution), region, scientific classification fields or free text search (in the abstract). The user also has the ability to filter data based on a number of criteria (such as grant type, status, classification and date). Users can go to the core details of the award and can export search results into Excel. In addition to GOTW NERC also hold details of funding of studentships on SOTW (Studentships On The Web) and of research projects undertaken by NERC's research centres.
	http://gotw.nerc.ac.uk/

3. Joint Research Information System: which added value?

The idea of linking the research information systems of various European institutions has been discussed in various contexts for many years.

Previous efforts to establish international platforms to exchange research information include:

- **European Research Gateways Online (ERGO),** a project undertaken by the EC to provide a one-stop shop for European R&D project information. The pilot phase was successful in collecting data on over 90 000 nationally funded R&D projects. The system was discontinued mainly because of lack of funding to maintain the system and update the information.

- **EXIRPTS (A Window on to Worldwide Research),** a project initiated by the national research organisations of the G8 countries in 1987. The project was discontinued after the conceptualisation phase in 1989.

- In 2002, the Netherlands Organisation for Scientific Research (NWO) launched an initiative to provide research management tools, which could be used, for effective expert or referee finding, proposal comparison, project management and evaluation, etc. The project was based on I-Research, a system using Collexis® "fingerprinting" technology. More than 20 organisations from different countries expressed an interest in participating in a pilot phase. The initiative was discontinued in 2005.

The early debates on the technical implementation of a European research information system culminated in the development of a Common European Research Information Format (CERIF) in 1991 (see Box 3).

More recently, the intensification of European collaboration in research and the idea of the "European Research Area (ERA)" have given a new impetus to the debate.

The Working Group asked the EUROHORCs Member Organisations what benefits did they see in linking up their systems with those of other organisations?

Summaries of the replies show a variety of perceived benefits of a joint system:

- Giving access to information on national research activities to an international audience and thus making the research of their own country more visible (especially mentioned by agencies in smaller countries).
- Getting information on national or foreign research activities:
  - benchmarking own funding activities to those of other organisations;
  - helps to identify and compare research profiles of researchers, institutions and even countries in specific fields;
  - helps to identify "hot fields of science";
  - helps in gap finding (e.g. to direct resources to research activities that have not yet been done elsewhere in Europe).

- Help managing national funding and research activities:
  - finding experts;
  - development of an international accessible peer review system.

- International and interdisciplinary cooperation:
  - supporting international networks of scientists;
  - by presenting research activities in all scientific fields, a common CRIS can help to overcome even disciplinary frontiers.

- Evaluation: data can be used for evaluative purposes (benchmarking of performance, describing collaboration networks, etc.).

Box 2: How are the RIS of EUROHORCs Member Organisations used?

The questionnaire included a question on the usage of the Research Information Systems.
Most organisations do not systematically collect information on the users (their profiles, their interests). From the few replies based on anecdotal evidence, the following five user groups can be deduced:

- The internal administration
- Academic community, including researchers
- Administration of universities, research organisations etc.
- Government bodies
- Journalists and the general public

In considering the above-listed potential benefits of a joint research information system, the Working Group had to critically assess which benefits could be realistically delivered by a joint system, built on the existing research information systems as *they are currently.*

3. Joint Research Information System: which added value?

In the view of the Working Group, any "added value" of a joint system will depend on two main factors: (1) How the "joint system" is built (its architecture and the services it provides); and (2) Whether the underlying Research Information Systems of individual organisations are broadly similar.

The consultation made it clear that the research information systems serve mainly to provide information on research that a given organisation funds. It is only recently that some organisations have started to track usage records and learn about what information users are looking for. Anecdotal evidence shows that users are likely to be administrators of funding agencies and funded institutions; researchers and journalists, as well as the general public (see Box 2).

In the view of the Working Group, the introduction of a joint system should not involve radical changes in the existing systems of Member Organisations. The challenge is to make existing information more widely available.

3.1 Added value of a joint research information system

The **main benefit** of a joint system is to make maximal use of information currently made available on the Web, but scattered in a range of systems and formats. The joint system would allow multinational searchers of different systems. It can be used alongside other well-established systems such as bibliometric databases, Google and Google scholar, etc.

Its **unique feature** is that it could provide a single point of entry to information about projects (and related researchers and organisations) that have successfully undergone a competitive selection through peer review.

In the consultation phase, most research-funding agencies indicated that the main advantage of a joint system would be to supply information about the research and researchers that they fund to an international audience and thus make their activities more visible internationally.

- A joint system can be used by researchers who would like to find others working in similar fields, or institutions with a strong research portfolio in a specific field or topic (In addition to the other means that they use such as publications, scientific conferences and EC databases).
- A joint system can be useful to administrators from research-funding agencies who wish to compare their research portfolios with those of other organisations. This can help to identify and compare research profiles of researchers, institutions and even countries in specific fields; to identify "hot topics" and gaps; and provide input in discussions on potential European collaborative programmes. With a common classification system, it could also be used to map resources (funding, number of researchers, etc.) devoted to different research areas in different countries.
- Administrators from research funding agencies could also use this system to find experts on a given topic. This information could be used (in combination with other sources of information) to identify referees or any potential conflict of interest (collaboration in a project, similar research topics likely to be a source of competition).
- A joint system could be used by journalists and the public to identify experts on a given topic across Europe.

Box 3: The CERIF Standard

The Common European Research Information Format (CERIF) was developed with the support of the EC (European Commission) in two major phases: 1987-1990 and 1997-1999. It is a standard; recommended by the European Union to its member states.

Since 2002 care and custody of CERIF has been handed by the EC to EuroCRIS, a not-for-profit organisation dedicated to the promotion of Current Research Information Systems (CRIS).

CERIF is neutral as to architecture; the data model can be implemented as a relational, object-oriented, RDF/OWL XML database, or as an information retrieval (including Web) system.

Core CERIF Entities

CERIF 2006 has four Core Entities: Project, OrganisationUnit, Person, and ResultPublication that all interact with each other (see Figure 1).

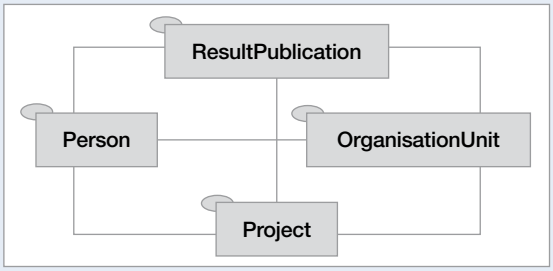


Figure 1. CERIF Core Entities

2nd Level CERIF Entities

Besides the core entities that represent key players (Person, OrganistionUnit) and their activities (Project, ResultPublication). CERIF captures the context of players and their interaction in the wider range of a research environment. Figure 2 demonstrates the core entities and some of the 2nd level entities within the complexity of their interactions.

Source:
www.eurocris.org
Text provided by Neil Jacobs, JISC

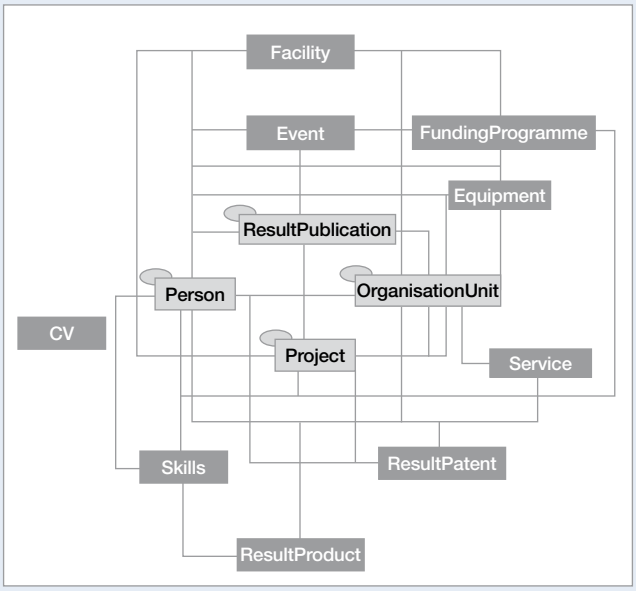


Figure 2. CERIF Core and 2nd Level Entities

CERIF Link Entities

For connecting the core and 2nd level entities, CERIF uses so called Link Entities (see figure 3). A link entity always connects two, either core or 2nd level entities and stores additional information about the established connection.

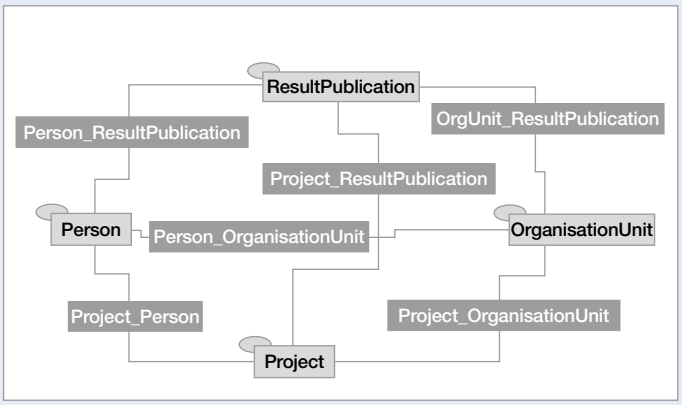


Figure 3. CERIF link Entities connecting Core Entities

Semantic layer

Layered over the above syntactic structure is a set of multilingual classifications that provide for fine-grained specification of the precise object and link properties that are relevant to a particular domain.

4. Joint Research Information System: which models? [2]

A brief survey of the literature, and consultation with domain experts, has identified the following three basic models, with some variants:

1. Central database model
2. Distributed database model
   a. Remote wrapper
   b. Local wrapper
3. Web crawling and search ("Google")

Clearly, this is not an exhaustive list – other variants are possible – but it does summarise the main relevant approaches. All of the models can be implemented with a portal application front-end, so that the functionality can be presented within Members' own web environments.

Before describing briefly each model and variant, the following clarifications/definitions are provided:

In this report, **Service** has a meaning approaching that in the concept of "service oriented architecture", wherein information systems are designed as modular units with defined interfaces ('services') between them.

By ***Canonical data model***, we mean here a common description of the objects and relationships that are widely agreed to be important in a particular domain. CERIF is such a canonical model (see Box 4).

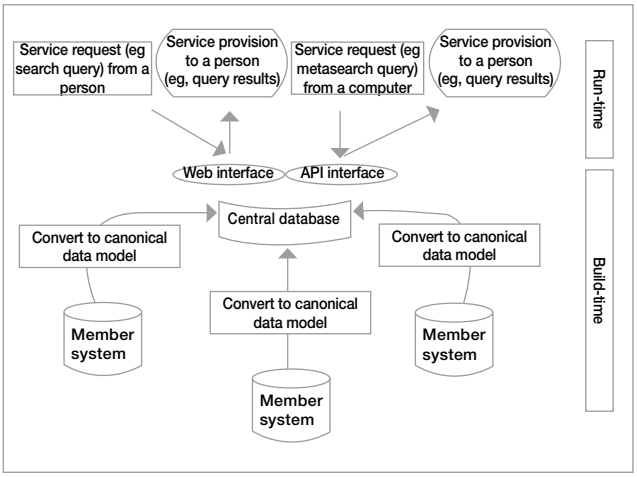
4.1. Central database models

There are two main variants of the central database models:

a. Central database model with regular batch update
b. Central database model with update via machine interfaces

4.1a. Central database with regular batch update

The development of the central database requires the selection of a canonical data model2 (CERIF being the obvious candidate), and the conversion of each dataset to that data model, including agreed syntax and semantics. Updates (also in the canonical data model) need to be supplied at regular intervals to ensure the currency of the central database. The original conversion and all updates should include as much of each Member's dataset as is required to fulfil the services that need to be provided. Effort is required to validate both the original conversation and all updates.

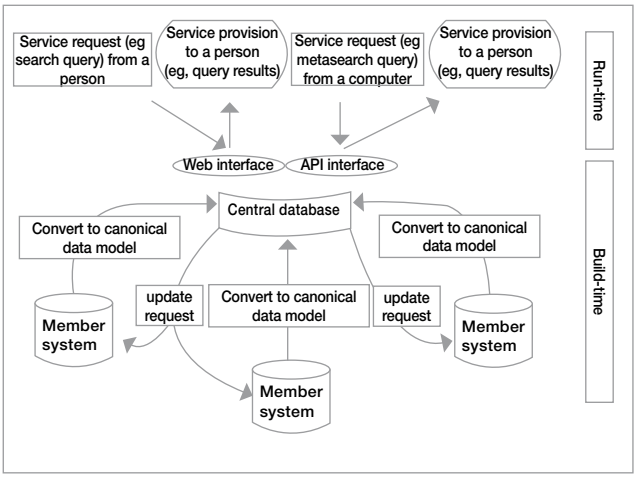


1a. Central Database with regular batch update
Architecture/processes

4.1b. Central database with very frequent update via machine interfaces

As in the previous model, the development of the central database requires the selection of a canonical data model and the conversion of each dataset to that data model. Updates are automatically prompted by the central database and supplied from Member's systems (for example, using harvesting protocols such as Open Archives Initial Protocol for Metadata Harvesting, OAI-PMH3), perhaps daily or even hourly.

The original conversion and all updates should include as much of each Member's dataset as is required to fulfil the services that need to be provided. Effort is still required to validate both the original conversation and all updates.



1b. Central Database with very frequent update via machine interfaces
Architecture / processes

[2] This Section is a based on a briefing prepared by Neil Jacobs, JISC

The central database model in both its variants has a range of advantages as well as disadvantages.

Advantages

1. Because the data model syntax and semantics are agreed, a wide range of services are available.
2. Services offered will be fast because they require interaction with only one database.
3. Services offered will be reliable because, at request time, they rely on only one database.
4. Service speed and reliability will be consistent across all datasets that have been added to the central database.
5. Services will be based on data that is reasonably current.

Disadvantages

1. There may be legal objections to moving some or all of the data across organisational and probably national boundaries.
2. Dedicated and ongoing effort is required to maintain the central database and its update and validation operations.
3. Dedicated effort is required in converting Members' datasets to the canonical form, and this effort will be needed again when the Member's data model changes.
4. The whole development process needs to be completed before any services, and therefore added value, are realised.

Examples of a central database model with regular batch are:

(a) ERGO project stage 1 1999-2000 (which used an earlier version of CERIF): http://cordis.europa.eu/ergo/
(b) EUROHORCs demonstrator developed by uniCRIS using CERIF2004)

Examples of central database models with frequent update via machine interfaces are:

(a) IDEAS (1984-87) and EXIRPTS (1987-89) both had the same architecture, wherein (1) there was a central catalogue (metadata from all the different databases in canonical format) which the user accesses with a query; (2) when the user gets hits from the query on the catalogue, they can request detailed information from original databases (3) information is supplied either in host format or in a canonical CERIF-like format.
(b) The following examples are from the domain of digital repositories, but the principle is the same: (1) BASE, Bielefeld Academic Search Engine, a multidisciplinary search engine for scientifically relevant Web resources: http://www.base-search.net/; (2) OAIster, a union catalogue of digital resources: http://www.oaister.org/

4.2. Distributed database models

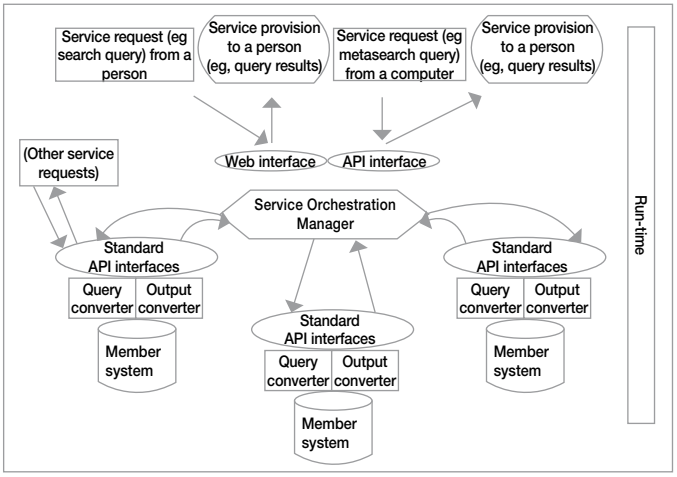
The distributed database model has two variants:
- Central database model with remote wrapper
- Central database with local wrapper

4.2a. Distributed database model with remote wrapper

For this variant, two elements need to be developed but, once running; there is no need for further intervention.

1. Service Orchestration Manager (i.e. an application that handles service requests, directs them appropriately, handles and collates responses and presents them back to the origin of the request). To support the defined services, the application should have a Web and API Interface (API: application programme interface refers to the target through which one computer can interrogate another according to a declared syntax and protocol). The API accepts service requests from both people (via one or more Web interfaces) and from other computers, and directs them as appropriate to relevant targets (API interfaces) offered by Members' systems. It is likely that this will need to work with a canonical data model such as CERIF.`

2. Converters running over each Member's system translate the service request from the Service Orchestration Manager (e.g. query) into something that the system can deal with appropriately, pass the service request to the system, accept its response and convert that back into the data model used by the Service Orchestration Manager. Should the internal operation or data model used by a Member's system change, then the converters at their site would need to be updated.



2a. Distributed database model with remote wrapper Architecture / processes

Examples

This model is common in digital library services, for example using the Z39.50 protocol and, more recently, SRW/U:

(a) TechExtra, a free engineering service cross-searching 30 key sources for articles, books, the best websites: http://www.techxtra.ac.uk/

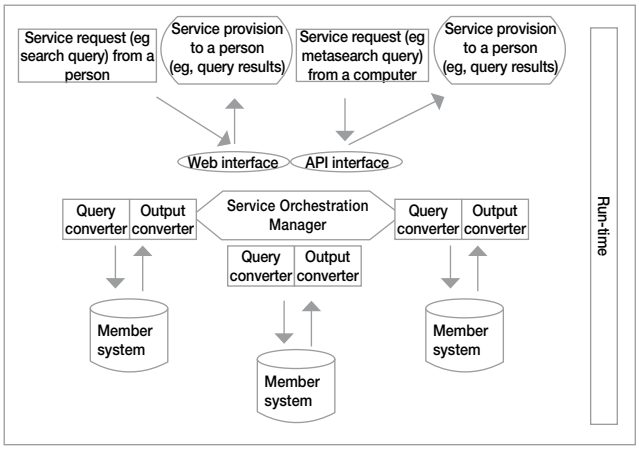
(b) OCLC WorldCat Registry Search, a Web-based directory for libraries and library consortia: http://orweblog.oclc.org/registry/institutions/

4.2b. Distributed database model with local wrapper

Only one element needs to be developed and, once running, there is no need for further intervention. The element needing development is the Service Orchestration Manager, with Web and API interfaces to support defined services and converters for each Member's system involved. This accepts service requests from both people (via one or more Web interfaces) and from other computers. It is likely that service requests will be accepted only if they can be mapped to a canonical data model such as CERIF, which acts as an agreed check for valid requests, and (later) as a structure within which to collate responses from Members' systems. The Service Orchestration Manager then converts these requests into appropriate formats/protocols for each relevant Member's system, and directs them as appropriate to targets offered by Members' systems. It then accepts valid responses from Members' systems, converts them back into the canonical data model, and presents them to the requester as appropriate (a Web page or a defined API). Of course, should the internal operation or data model used by a Member's system change, then the converters would need to be updated.

Examples

None identified.



2b. Distributed database model with local wrapper Architecture / processes

Both models have the following advantages:

1. The architecture can be developed incrementally, once a common data model has been agreed within which further development can take place – so a single interface can be defined to meet Members' highest priority, with further interfaces added as required.
2. Datasets are not moved wholesale across organisational or national boundaries, alleviating any legal concerns that might arise, and local security/privacy conditions can be enforced.
3. Services are always offered on current data.

The main disadvantages are:

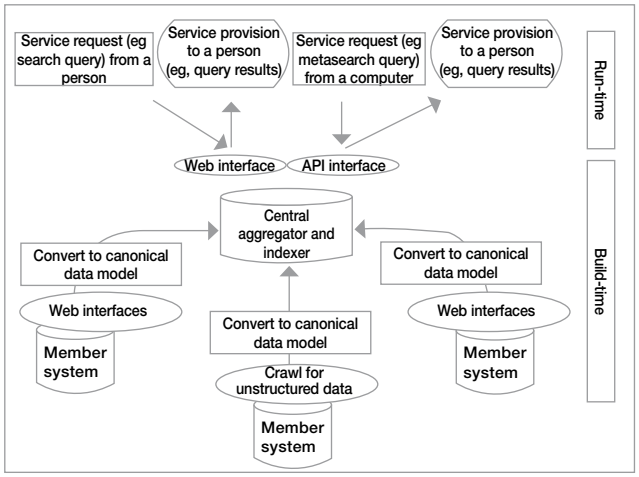
1. Service requests (e.g. queries) are more limited than in central database models in particular to those service requests that are pre-defined, which may not match real requirements. Having said that, there can be flexibility in defining queries within a pre-defined user query form, and agreed syntax and semantics give a large scope for such pre-definition.
2. Unless precautions are taken, the speed and reliability of services are both limited by those of the slowest / least reliable system involved.
3. Maintaining a number of request and response converters at a central point may prove demanding. Members will need to ensure

that any change affecting the operation of their target APIs is discussed with the centre; otherwise, there is a risk that service requests to that Member's system will fail.

Additionally for the distributed database model with local wrapper, it may prove difficult to establish and maintain a common understanding and implementation of the syntax and semantics of service requests and responses.

4.3. Web crawling and search

Web crawlers, aggregators and indexers are readily available, so little development would need to be done. It could just be a matter of installing a local Google search.



Web crawling and search
Architecture / processes

Advantages

Little or no development effort is required.

Disadvantages

1. Service requests are limited to little more than keyword searches, which are readily available from Google and elsewhere.
2. The lack of agreement syntax and semantics render any analytic services (such as counts) too unreliable to be useful.

4.4. Comparing the models

In weighting the advantages and disadvantages of the three models, the Working Group identified the Distributed Model (either variant 2a or 2b) as the model of choice (see Table 2) as discussed in Sections 4.2a and 4.2b.

- The Web crawling/search model has the immense disadvantage that it retrieves unstructured data. This means that none of the benefits of a joint research information system as identified (in Section 3) can be realised.

- Comparing the Central Database Model and the Distributed Model, the former requires far greater efforts to maintain. In addition to technical maintenance, such a system would require continuous efforts to convert the original dataset and to ensure the timely and regular update of the records. Furthermore, even if we are dealing with data which is freely accessible on the Internet, the relative restrictive data privacy regulations in some countries are likely to be a problem when data are moved "wholesale" from one organisation/country to another.

Table 2: Simplified representation of the advantages and disadvantages of the identified models

	Advantages	Disadvantages
Web Crawling Model	Little development effort needed (each information system would just need to expose the information it holds in such a way that the Google (or other) robot can find it and use it.	The retrieved data are unstructured and unsuitable to deliver any of the potential added values identified above.
Central Model	A wide range of services can be offered as the data model syntax and semantics are not a problem.	Considerable effort in the maintenance and update of the central database.There might be legal objections to moving some or all data across organisations/countries.
Distributed Model	Datasets are not moved (just retrieved); services are always offered on current data.	Considerable efforts in maintaining a number of requests and response converters (especially as the contributing systems are likely to change). This effort is much reduced by using a common Format such as CERIF (see paragraph 7.3), since that means that single shared "language" to which all these queries and responses need to translate.

5. Joint Research Information System: which way forward?

The survey of the Working Group showed that a growing number of EUROHORCs Member Organisations have embarked on the development of research information systems. For 17 organisations, information on such systems could be collected and an additional six organisations indicated that they were in the process of planning or implementing such systems.

In line with its terms of the references, the Working Group had to assess whether a pilot project to link the existing systems is recommendable and which model is appropriate for such a joint system. The Working Group had also to critically assess the benefits of such a joint system.

Section 2 of this report discusses a number of benefits and potential usage of such a joint research information system. Section 4 compares three various ways to build a joint system and identify the distributed model as the model of choice.

However, given the tremendous heterogeneity of existing systems (both in content, languages, information entities covered), the identified benefits are not likely to be delivered by just linking up the existing systems in their current status. The Working Group felt that a certain degree of convergence is a sine qua non condition for a useful joint information system.

The Working Group formulated its recommendations in six points reproduced below:

1. Exchange of Information and Experiences

Independently of whether or not a joint research information system is built in the short or long term, the collection of information on existing research information systems and the presentation of selected case studies have proved useful in its own right. This exercise provided for the first time a mechanism for officials from ESF/EUROHORCs Member Organisations who deal with research information systems to **exchange experiences**. It also provided the opportunity for organisations, which are planning to develop online research information systems to take into account the experiences of other organisations.

Recommendation 1

To encourage the exchange of experiences (and independent of any joint information system), EUROHORCs is recommended to maintain a web-page of links to research information systems of their Members and to facilitate the networking of their specialist staff who are involved in the development of research information systems.

The networking of specialist staff members who are responsible for research information systems could be achieved through regular meetings. To minimize time and effort, such gatherings can be organised in the frame of other events/conferences that the potential participants are likely to attend.

2. Pilot Project

The Working Group considered **whether a pilot project of a joint information system should be launched** as suggested in The Hague Workshop in October 2006. The idea discussed was a small-scale project involving research information systems of selected organisations and serving as basis for full-scale research information systems, which would be built incrementally.

The Working Group felt that it would be premature to launch a pilot project of a joint research information system at this stage.

Some organisations were not ready to undertake a joint information system. For example, the UK Research Councils were establishing a Shared Service Centre, which requires significant restructuring of current and historical data, and is making exceptional demands on IT/IS staff time. Thus, the UK Research Councils indicated that a joint information system, although important, could not be a priority for them within the next two years.

Moreover, the Working Group noted considerable divergences in the 'contents' (coverage) of the single research information systems, which would seriously limit the benefits of a joint information system at this stage:

- Many systems do not have abstracts in English and some do not have abstracts at all;
- In some systems the abstracts target the general audience (journalist, public) and are written in 'simple' language while others collect the abstract provided by the researchers in their proposals (and are therefore written in more technical terms);
- Some systems include information on funded PhD theses, others do not;
- Not all systems contain the institutional affiliation of the researchers involved in the project (some provide only the name of the Principal Investigator);
- The systems use different classification schemes (notably of research fields);
- Meaningful comparisons across systems require a comparable nomenclature of funding schemes (a travel grant to attend a conference differs significantly from a grant to establish a centre of excellence).

In the light of the potential benefits of joint systems, the Working Group holds the view that the idea should not be abandoned altogether. However, the Working Group considers that more effort is needed to make the research information systems of single institutions more convergent, and this should precede any attempt to build a 'joint research information system'.

Recommendation 2

The Working Group recommends not launching a pilot phase at this stage.

3. Cooperation with EuroCRIS

To improve convergence there is no need to reinvent the wheel. The CERIF standard (Common European Research Information Format) was specifically developed to provide extensible European data storage and transfer formats for Current Research Information Systems (CRIS). In the view of the Working Group CERIF is the system of choice. It was developed with the support of the European Commission and has been recommended by the European Union to its member states as a standard data model for R&D information in Europe. The custodian of the CERIF standard is the EuroCRIS organisation, and it is available at nominal cost (www.eurocris.org). This is a not-for-profit organisation and corporate membership is available.

Efforts should be made therefore to raise the awareness about CERIF, and national research organisations should explore further, how CERIF can be used as a basis to achieve the convergence of their research information systems. The goal is clearly not that all national research organisations use the same system or the same data model, but rather to take into consideration the convergences of their system with that of their counterparts when developing their systems further.

Recommendation 3

The Working Group recommends EUROHORCs to encourage their Member Organisations to develop their systems in a more convergent way. The issue of a joint system should be discussed again at the first EUROHORCs General Assembly in 2010.

In this process, collaboration with EuroCRIS, a not-for-profit European forum dedicated to maintaining standards and best practice in Current Research Information Systems' (and other key actors in developing research information systems at the European level as appropriate) would be beneficial. Preferably, collaboration with other partners should be based on a Memorandum of Understanding (MoU). The MoU, which shall be made public, should make clear that any purchase of services — especially if the joint system is to be built at a later stage — would be made in a competitive and transparent procurement process.

Recommendation 4

The Working Group recommends close collaboration with EuroCRIS to facilitate the convergence of research information systems of research-funding and research-performing organisations.

4. Assessment of the Commitment of EUROHORCs Member Organisations

Before any actions are initiated, it is essential to assess the commitment of EUROHORCs Member Organisations to the idea of developing their research information systems and work towards the long-term goal of integrating them in a joint distributed European Research Information System. They should be asked to indicate their interest and whether they are willing to devote the necessary resources (mainly in staff time and budget to attend meetings) to this undertaking.

Recommendation 5

The Working Group recommends that the interest and commitment of EUROHORCs Member Organisations (in working towards the convergence of their research information systems) be assessed prior to launching the initiative. The initiative should go ahead only as a EUROHORCs initiative if there is critical mass of organisations, which indicate their interest and commitment. The Working Group recommends at least eight organisations as a threshold.

5. Implementation of the recommendations

The process of working towards greater convergence of the research information systems of EUROHORCs Member Organisations includes a number of activities, which would be initiated and coordinated in the years 2008-2010.

They include, among others:

- Exploring the possibilities for close cooperation with EuroCRIS and prepare the MoU on which such cooperation should be based.

- Organising meetings and other means to exchange information and experiences.
- Updating and maintaining an inventory of research information systems and contact details of responsible staff members.
- Defining a set of minimum requirements that single research information systems must meet in order to be integrated in a joint research information system and assessing how "ready" EUROHORCs members are.
- Organising information and training sessions for officials from EUROHORCs/ESF Member Organisations on CERIF.
- Exploring in a meeting of technical staff the range of services which can be developed on the basis of a joint information system (i.e. which sort of services can be done and perhaps provide additional added value which is as yet not identified).

The Working Group has identified two options to implement the above activities.

Option 1: EUROHORCs could set up a new Working Group or a Task Force modeled on the other existing EUROHORCs Working Groups.

Option 2: EUROHORCs could ask the ESF Office (as its 'Executive Agency') to implement the recommendations. In this case, ESF would be assisted by a 'reference group' drawn from among representatives of EUROHORCs interested Member Organisations.

Recommendation 6 (Implementation of the recommendations)

The Working Group presents two options for the EUROHORCs General Assembly to consider as on how to implement the recommendations made in this report:

(1) By setting up a Working Group or a Task Force

(2) By asking the ESF Office to implement the recommendations, in which case ESF will be assisted by a 'reference group' drawn from among representatives of EUROHORCs interested Member Organisations.

In either case, EUROHORCs General Assembly is recommended to specify its objectives and formulate a clear mandate.

If the EUROHORCs General Assembly wishes to continue to explore the possibilities of linking its Member Organisations' research information systems to each other, it should to express its preference for one of both options.

6. Another step necessary in encouraging the EUROHORCs Member Organisations to develop their systems in convergent ways is to create an inventory of their research-funding schemes and the schemata they use to classify research fields. The goal of such a study (focusing on taxonomy of funding schemes and classification of scientific fields) should not be to replace the existing classification schemes but to develop, potential cross-mappings against a standard (e.g. based on the OECD standard – FRASCATI Field Classification). An ESF Member Organisation forum on Research Evaluation is already discussing this idea (relevant in comparing the evaluation approaches in different research funding agencies) and a EuroCRIS Working Party on research classification exists, which can take this forward.

The report of the Working Group and the above listed recommendations were submitted to EUROHORCs in January 2008. It was decided not to launch a pilot phase as recommended and not to pursue further the idea of a joint research information system for the time being.

Appendices

Appendix 1: Executive Summary of the recommendations of the Working Group to EUROHORCs

The Working Group was convened in January 2007 by ESF at the request of EUROHORCs, to "*draft a business plan on how to link up the databases of EUROHORCs members*".

In discussions with EUROHORCs Steering Committee and General Assembly, the tasks of the Working Group were defined as follows:

(1) To make an overview of existing Research Information Systems
(2) To assess the added value of a joint system
(3) To identify appropriate models of a joint system
(4) To make recommendations to EUROHORCs on how to proceed further.

The ***overview*** shows that it is mainly research-funding agencies that have been particularly active in developing research information systems in the recent years. Those systems complement annual reports and serve mainly for accountability purposes (providing information on funding activities and sometimes recording the output of the funded projects). Researchers, administrators from other organisations, journalists, also use them. In all information on 17 research information systems was collected and six organisations indicated that they were planning or further developing their systems towards fully fledged research information systems.

The Working Group believes that there are clear ***potential benefits*** in establishing enabling mechanisms to combine the information contained in those systems. A unique feature of such a system would be to provide a single point of entry to information about projects (and related researchers and organisations) which have successfully undergone a competitive selection through peer review. Such a system would give the research information systems of participating organisations an international audience and help make maximal use of information currently available on the Web but scattered in a range of systems and formats.

A ***distributed model*** in which a central node sends structured queries to the systems of different organisations and formats, the output being a pre-defined format, was identified as the model of choice.

This model would minimise efforts and resources needed in developing and maintaining such a system.

Recommendations

Given the various scopes (both in content, information entities and languages), the disparate data models and architectures of existing systems, the Working Group believes that a joint system at this stage would not realise any potential and that it would be premature to launch a pilot joint system as suggested at a EUROHORCs workshop held at NWO Headquarters in October 2006 (and indeed implied in the EUROHORCs request to convene this Working Group). In the view of the Working Group, a fair degree of convergence of the systems is the prerequisite to a useful joint system.

The Working Group recommends not launching a pilot phase at this stage, but instead recommends facilitating a process in which EUROHORCs Member Organisations can continue sharing information and experiences on their respective systems and in which organisations interested in a joint system can work together to achieve a convergence of their systems.

The Working Group formulated the following six recommendations to EUROHORCs General Assembly for its considerations:

Recommendation 1:

To encourage the exchange of experiences (and independent of any joint information system), EUROHORCs is recommended to maintain a webpage of links to research information systems of their members and to facilitate the networking of their specialist staff who are involved in the development of research information systems.

Appendix 1

Recommendation 2:

The Working Group recommends not launching a pilot phase at this stage.

Recommendation 3:

The Working Group recommends EUROHORCs to encourage Member Organisations to develop their systems in a more convergent way. The issue of a joint system should be discussed again at the first EUROHORCs General Assembly in 2010.

Recommendation 4:

The Working Group recommends close collaboration with EuroCRIS to facilitate the convergence of research information systems of research-funding and research-performing organisations.

EuroCRIS is a not-for-profit European forum dedicated to maintaining standards and best practice in Current Research Information Systems. The cooperation should be based on a Memorandum of Understanding which identifies concrete actions to be taken, the results to be achieved and set a clear timetable.

Recommendation 5:

The Working Group recommends that the interest and commitment of EUROHORCs Member Organisations (in working towards the convergence of their research information systems) be assessed prior to launching the initiative. The initiative should go ahead only as a EUROHORCs initiative if there is critical mass of organisations, which indicate their interest and commitment.

The Working Group recommends at least eight organisations as a threshold.

Recommendation 6 (Implementation of the recommendations)

The Working Group presents two options for the EUROHORCs General Assembly to consider on how to implement the recommendations made in this report:

(1) By setting up a Working Group or a Task Force
(2) By asking the ESF Office to implement the recommendations, in which case ESF will be assisted by a 'reference group' drawn from among representatives of EUROHORCs interested Member Organisations.

In either case, EUROHORCs General Assembly is recommended to specify its objectives and formulate a clear mandate.

Appendix 2: Agenda of the Workshop organised by the Working Group 25-26 June 2007

Windows to Science:
Information Systems of European
Research Organisations
An ESF – EUROHORCs workshop
Stockholm, 25- 26 June 2007

The aim of the Workshop is
- to exchange experiences and practices on existing RIS
- to discuss the added value of a joint Research Information System
- to explore and document the advantages and disadvantages of various models of joint research information systems (distributed/central database models)

Programme

Monday 25 June 2007 : Session 1: Research Information Systems of EUROHORCs Member Organisations Particularities and common ground (*Chair: Neil Williams*)	
12:00 – 13:30	An informal get together – sandwich lunch
13:30 – 14:00	Welcome Address *by Annette Moth Wiklund (VR, EUROHORCs)*
14:00 – 14.30	Research Information Systems of EUROHORCs Member Organisations – An Overview *Jesper Aven (VR) and Alexis-Michel Mugabushaka (ESF)*
15:30 – 16:00	Coffee break
16:00 – 16:30	GEPRIS/DFG – *Holger Hahnen*
16:30 – 17:00	Swedish Research Council (VR) – *Jesper Aven*
17:00 – 18:00	The Scientific and Technological Research Council of Turkey (TÜBITAK) – *Ramazan Acum*
18:15 – 18:30	Coffee break
18:30 – 19.00	Netherlands Organisation for Scientific Research (NWO) – *Ruud Strijp*
19:00 – 19.30	Swiss National Science Foundation – *Marcel Kullin*
20.30	Dinner

Tuesday 26 June 2007 Session 2: Towards a Joint Research Information System (*Chair: Neil Williams*)	
09:00 – 10:30	Joint Research Information system: where is the added value? *Jürgen Güdler (DFG) & Ruud Strijp (NWO)*
10:30 – 11:00	Coffee break
11:00 – 12:00	Models of joint Information systems: distributed/central database models *Neil Jacobs* (JISC)
12.00 - 12.30	Closing of the Workshop

The Workshop is kindly hosted by the Swedish Research Council
The presentations are available under:
www.esf.org/activities/science-policy/corporate-science-policy-initiatives/research-information-systems-2007.html

Appendix 3: Workshop Participants

Czech Republic
Marek Simacek; Czech Science Foundation

Denmark
Per Kolbeck Nielsen; Danish Agency for Science, Technology and Innovation

Finland
Maj-Lis Tanner; Academy of Finland

Germany
Andreas Trepte; Max Planck Society (MPG)
Holger Hahnen; German Research Foundation (DFG)
Juergen Güdler; German Research Foundation (DFG)

Hungary
Bernadett Kovacs; Hungarian Scientific Research Fund (OTKA)

Ireland
Kevin Flynn; Enterprise Ireland

Italy
Giovanni De Simone; National Research Council (CNR)
Massimiliano Di Bitetto; National Research Council (CNR)

Netherlands
Ruud Strijp; Netherlands Organisation for Scientific Research (NWO)

Spain
José Luis Cereceda; Spanish National Research Council (CSIC)

Sweden
Jesper Avén; Swedish Research Council (VR)

Switzerland
Marcel Kullin; Swiss National Science Foundation (SNF)

Turkey
Ramazan Acun; Scientific and Technological Research Council of Turkey (TÜBITAK)

UK
Neil Jacobs, JISC

COST
Christer Halen

ESF
Neil Williams
Monique van Donzel
Alexis-Michel Mugabushaka

Print run: 300 - Latitude :

EUROPEAN SCIENCE FOUNDATION
SETTING SCIENCE AGENDAS FOR EUROPE

1 quai Lezay-Marnésia | BP 90015
67080 Strasbourg Cedex | France
Tel: +33 (0)3 88 76 71 00
Fax: +33 (0)3 88 37 05 32
www.esf.org