The Need for a Standardized Current Research Information System (CRIS): A "Call to Arms"

ERIC H. ZIMMERMAN, KEITH G. JEFFERY
Bar Ilan University, Israel and CCLRC Rutherford Appleton Laboratory

Summary

In today's world of growing complexity in science, there is a need for systems to assist researchers, administrators, strategists, opinion-formers, entrepreneurs and innovators and also the general public. Systems are needed to provide both information for decision-making and support to the process of knowledge-creation. Here we investigate the former, proposing a call to action in Europe.

1 The Requirement

Access to information on current research activities throughout Europe is an essential requirement for the success of EU innovation policy. The key asset in European R&D consists of ideas, technical reports, publications, patents, prototypes, products and know-how – leading to technology transfer and wealth creation, and to the generation of new R&D ideas. The key added value to be achieved is the pan-European approach to the generation of and exploitation of R&D. There is a need for information on currently relevant R&D information to be made available widely to encourage both innovation and new, improved R&D.

The innovators in industry and services, the academics pushing the frontiers of R&D, the decision makers in governments and R&D funding agencies all require easy-to-use access to R&D information. The raw data sources are the R&D Information held by funding agencies and other information providers in the EU states. These are held for the particular purposes of the agencies and the particular clients of the information providers. They are heterogeneous and unconnected. The potential for European wide exchange is not being exploited.

There is a need for the end-user of research information to be able to access this data through a uniform survey-level interface and to be able to integrate and compare the information between data sources. This "common interface" must not only address the content (what must be exchanged) but also the format of such information (how it should be presented). This information must be presented in a uniform way, at least at summary level. The Web and database technology are the way to provide this.

The definition of this uniform information description platform requires:

- The definition a full CRIS data model which will cover the database structures of the majority of existing CRIS;
- The definition a set of data models which could provide examples for data exchange (since there are an infinity of possible exchange data models between CRIS);
- The definition of a metadata data model to provide a uniform summary-level view over heterogeneous information sources.

Easy access to information must address not only the availability of information with a common definition and format but also how that information could be retrieved by the end-user. The end-users need to be able to search, European wide, for information on a particular research topic or theme. Subject indexing of the information is the obvious key in this respect. Classification should be consistent for all the research information sources; otherwise users will not get consistent results when they retrieve information. Since end-users also use different languages, the controlled indexing terminology proposed should have the same meaning in all languages.

2 The Role of CRIS

2.1 Introduction

CRIS can be a facilitator of cross-discipline research. Increased specialization in modern science has added urgency to the need for communication. CRIS can be used to power the communication networks of 21st century science. An intelligently designed CRIS may provide context to science. With ever-increasing cross-disciplinary science being performed across borders, the importance of preserving context cannot be overstated. As Gibbons has stated, Mode 2 knowledge is produced within a context of application. The involvement of different perspectives, languages (and discipline-oriented jargon) and cultural settings, by teams of people with heterogeneous skills, necessitates that a system to preserve context exists. CRIS can be used as the engine of Collaboratories and collaborative white-boards, and also to build repositories of published and "unpublished" literature, as we have discussed earlier.

2.2 Information to Knowledge

Knowledge Management has, it may be argued, two main functions: "community of practice" building and adding structure to a rich body of knowledge so that it might be used and reused. CRIS can serve both functions effectively. The management of "communities of practice" necessitates the development of expertise data-

bases or "competency dictionaries". Mode 2 Knowledge Production is performed in transient group settings. Networks are developed on an ad-hoc basis. When a given problem is solved, these networks are oftentimes dissolved. It is vital that the experiences of these groups be captured. This knowledge management is crucial if new knowledge is to be transferred to solving a new problem. Hence, the need to capture gray literature is apparent. Gray Literature is increasingly recognized as an important area of research information study, because, as Keith Jeffrey has written, "in a R&D environment [it] represents the cutting edge of this knowledge and so its management is of utmost importance". As Jeffery writes, with gray literature an organization documents and stores its knowledge assets.

2.3 Information to Innovation

CRIS can be the natural bridge between scientific information and innovation information. These are two different yet complementary types of information, necessary for the presentation of the "complete picture" of science. Whereas the general aim of science is to understand nature and society, innovation seeks to create new products, processes and services. In recent years we are witness to the growing dependency of innovation on science. This rapidly expanding body of knowledge should be captured by CRIS.

2.4 Information Tsunami

With ever increasing amounts of information available it is becoming more important to be able to find the relevant information needed. Productivity cannot be jeopardized by the time needed to sift through the vast amount of data and information, of varying quality. Efficient retrieval can be accomplished by using CERIF to control the "language of science". The key here is to develop and maintain a global taxonomy of science. This will help all actors drive science forward, by filtering out the "noise" associated with high recall and low precision. Though some stages of the information coding can be automated, the intervention of human information specialists would be crucial in order to ensure the proper monitoring and control of language, concepts and rules.

The European Union's Innovation policy aims at improving and strengthening the generation and exploitation of current and new Research and Development (R&D) projects as well as technology transfer. To this end, access to information on current research activities throughout Europe is an essential requirement. New R&D ideas can emerge thanks to a pan-European approach for information sharing and exchange. There is thus a need for a convenient tool to spread relevant R&D information widely to encourage innovation and new, improved R&D as well as wealth creation.

3 The Solution

3.1 CERIF: Introduction

CERIF - developed by a pan-European team of experts: CERIF91 (1988-1991) and CERIF2000 (1997-1999) (www.cordis.lu/cerif) - provides a practical common standard for information contents and for subject indexing. Further, controlled value lists ease the collection and exchange of data. To provide easily searchable information implies adapting common rules. To this end, it is important to use standard controlled vocabularies. The use of standard controlled vocabularies and standardised data structures, as described further in the CERIF guidelines, should be as wide spread as possible, in order to make both data providers and end-users familiar with common CRIS characteristics.

CERIF is this common language that fosters the diffusion of information across Europe. The EU Commission's recent Green Paper on Public Sector Information [1] emphasizes the importance of access for European citizens to publicly funded information and equally the opportunities for economic growth and employment that it provides. All Member States are taking initiatives with regards to public R&D information – some quicker then others. European Union policy should therefore aim to have all Member States arrive at the same point, and as quickly as possible, by organizing seminars, studying case studies, adopting standards and establishing networks of user groups. CERIF in addressing Public Sector R&D information is dealing with an area of economic activity with a high growth potential, and is crucial to the ongoing competitiveness of European industry.

3.2 Taming Heterogeneity

Funding agencies hold the raw R&D Information and other information providers throughout the EU amongst which are the so-called Current Research Information Systems (CRIS). However, data is heterogeneous and unconnected. The potential for European-wide exchange is being under-exploited, although there is a need for accessibility to and comparability of data through a uniform query interface from various sources. The end-users need to be able to search, European-wide, for information on a particular research topic or theme. The ERGO pilot project (www.cordis.lu/ergo) demonstrated feasibility. Responsibility for the development of CERIF has now been handed by the EU to euroCRIS (www.eurocris.org) a not-for-profit organization registered in the Netherlands.

---

[1] "Public Sector Information: a key resource for Europe", Green Paper on Public Sector Information in the Information Society, European Commission, COM (1998) 585 http://www.echo.lu/info2000/en/publicsector/gp-index.html

3.3 Best Practice

The Code of Good Practice (CGP) was developed as a guide for both new and existing producers of CRISs. The intention was to focus clearly on the reasons for having a CRIS and on the main components of the system. The CGP is not a comprehensive guide to building and developing information systems. In this respect, CERIF is more of a hands-on construction guide.

As a stand-alone system each CRIS plays an important role for the host institution or organization. Together, a collection of CRISs is potentially a very powerful information tool, the true value of which can only be harnessed if interoperability can be achieved. Universal adoption of the CGP by CRIS producers for both new and existing CRISs will be a significant step towards realizing this goal and will provide CRIS users and data providers alike with a framework for knowledge transfer. Likewise, there will be greater scope for the CRIS institutions to exchange data for mutual benefit or commercial gain. The increasing accessibility of information through developments in technology further emphasizes the need for consistency (with regard to information exchange) to ensure that the wealth (and potential diversity) of information available locally is accessible globally.

CRIS powered by CERIF, administered by research administrators and information specialists, is the answer to many problems. Such a make-up provides up-to-date information on R&D for innovators to drive wealth-creation. It provides for scientists seeking partners to cooperate – especially in cross-disciplinary or cross-national projects where the researchers with the specialized skills required are not in their usual circle of peers. It informs funding agencies and governments such that they can make strategic decisions about R&D based on facts. It provides for access to R&D information for intermediaries (brokers) and for the media. It also provides information on R&D for citizens. The scope is local, national, regional and international. The coverage can spread from public-funded R&D to encompass jointly funded R&D (with commercial organisations) and onwards to all R&D, which is deemed to be neither commercially, nor security sensitive.

4 Next Steps

4.1 CRIS at the Centre

CRIS can and must take center stage as the mechanism of information provision in the world of science. The need for quality scientific and technological information, available to all, is apparent. In our knowledge-based society, where so much is dependent on scientific discovery, we must be positioned to intelligently make use of information and apply it to address the problems and issues of the day. CRISs,

naturally so, have developed independently, based on the needs of individual organizations. They have been useful for the organization but rarely more widely. We are now able to characterize them and we are finding ways to assist them to exchange data or to intercommunicate. Early evidence suggests that this ability is useful to a wide user group. We can even see how to communicate with the CRISs all at once, and get the best of all of them into a global view. This would be ideal for a large user community.

4.2 CRIS and ERA

Clearly, much has been done. Yet, there is much more to do. We know now the road we must travel to make linked CRISs useful for a wide group of users. Interconnected CRISs are the lifeblood of the ERA, indeed of global scientific communication. The rapid growth of the knowledge economy means that the prosperity of Europe is now inextricably linked to its capacity to innovate. Only through innovation leading to the creation of new products and services can present standards of living be maintained and enhanced. But a fragmented research base, which the evolution of the ERA must overcome, handicaps Europe. Realization of the ERA will require ready access to high quality current research information on the people, projects and processes that together constitute the ERA.

The timely exploitation of scientific and innovation information is a prerequisite for the success of the European Research Area. Europe must be positioned to leverage the vast quantity of research information generated by its scientific endeavors for the good of society. At the core of this new science system is the need for a robust research information system.

There is therefore a need for systems to meet the diverse needs of researchers, administrators, strategists, policy-makers, opinion-formers (especially in the media), entrepreneurs and innovators; and also the general public, whose negative perception of innovation is a major problem. CRIS *(Current Research Information Systems)* exist to collect and disseminate research information. This information includes People (researchers and inventors), Projects, Organizations, Results (publications, patents and products), Facilities, and Equipment. CRIS therefore have the potential to provide the information resource from which the needs of European society can be met. Europe's history has led to a fragmentation of research effort. We must now put the pieces together.

4.3 Progress to Date

Within the CRIS community, experts on national CRISs have worked together to produce a technological infrastructure to provide convenient integrated access to national RTD information. Products to date include the CERIF recommendation

(http://www.cordis.lu/cerif) and the ERGO pilot demonstration project (http://www.cordis.lu/ergo). The technology era we live in mandates that we extend the current technologies with emerging Information and Communication Technologies (ICT), including homogeneous access to heterogeneous distributed databases using techniques from GRIDs and Web services, with domain ontologies, for expert assistance in query and answer elucidation, and for schema reconciliation and multilingual capabilities.

As this technological effort advances, there must be constant dialogue with the owners and providers of information in universities, research institutes and government agencies, as well as with the end users – the researchers, policy-makers, administrators, entrepreneurs and others involved in the knowledge economy. We need to create a space where they can meet and exchange opinions. Only in this way can Europe develop the information infrastructure for a true European Research Area.

4.4 Stakeholder Association

Future large-scale R&D information systems can succeed in Europe because there is much experience and there exists the necessary financial means. With regards to ERGO, INDARD and other failed attempts at building intelligent CRISs, it is to no fault of available technology, but rather to a lack of will, that so many skilful researchers are still working in isolation, oftentimes on similar problems, many of which have already been solved elsewhere. For lack of a suitable communication system many scientists just do not know of these findings until such time as the results have appeared in scholarly journals or pre-print servers, oftentimes a year or more later. Instead of sharing such knowledge broadly for the benefit of mankind, the right hand does not know what the left hand does and therefore the majority of researchers are still performing below their potential. In a nutshell: synergy by definition means giving and taking. As long as this is not really understood, even as we proclaim to be in the age of information sharing, the existing technology can't help.

An association of R&D information stakeholders is needed. This group would serve to nurture the CRIS-community by keeping eye on the added value. It would foster initiatives to counter the negative aspects of the fast growing information "wealth". It would keep abreast of new and emerging technologies and provide for material that can be easily implemented. It would also be positioned to lobby and tender to keep our needs on the radar screen of decision makers in Brussels and elsewhere. CRIS-expertise is essential (knowledge is power). The association can be a fountain of creativity by detecting the needs and responding to them with the best system by looking for alternate ways of bringing "the mountain to Mohammed".

5 Contact Information

Eric Zimmerman
The Department of Information Science, Bar Ilan University
Ramat Gan, Israel

e-mail: zimmee@mail.biu.ac.il
Website: http://www.biu.ac.il/RA/www/rserch/eric/index.html

Keith G. Jeffery
Director, Business and Information Technology Department
CCLRC Rutherford Appleton Laboratory

e-mail: k.g.jeffery@rl.ac.uk
Website: http://www.bitd.clrc.ac.uk/Person/K.G.Jeffery