Anne Asserson and Keith G. Jeffery

SYSTEMVERNETZUNG

Current Research Information Systems (CRIS): Past, Present and Future

Civilised society has always carried out research. Such research is documented from early Chinese civilisations, and may well have been documented in Neolithic times. Indeed, the paintings in caves such as Lascaux appear to be both an observational record (of local fauna) and a modelling or simulation of intended action (hunting). The key point is that research leads to wealth creation and improvement in the quality of life. The problem is that the process to create wealth or life improvement from research outputs is little-understood and apparently non-deterministic. However, there is a general belief that if one documents the research activity, and the research output, then the opportunities for wealth creation or improvement in the quality of life are increased.

During the last 50 years, there has been e-documentation of research information. Systems were usually restricted to research grants awarded by research funders or received by universities – i.e. systems essentially for financial management and accounting. Only much later were systems implemented for research outputs such as publications. The grants databases in each country were used not only for record keeping and financial management, but also extended to check that stages in the awarding and management of research projects funded had been completed. They were developed with national objectives in mind. They could provide information such as how many universities receive grants in biochemistry? or what is the total amount of research funding by year spent on research in chemistry?' Such management information informed decision-making on research priorities. These systems generally were in batch processing mode; one of the earliest to 'go online' (for use by research funding staff) was the UK system in the late seventies. Online access from university applicants or recipients of grant funding was not trialled until the mid-eighties and only went live in the nineties. The story is similar in other countries.

Linking National CRIS

The potential strategic value of CRIS for decision-making could be enhanced if research managers in one country could also have access to the information on research funding in other countries. The first breakthrough was in the early eighties. The heads of research funding organisations in France, Italy and UK set up a pilot project to interoperate their research documentation systems. The IDEAS project (1984-1987) was successful demonstrating a query in English accessing Italian and French (as well as UK) databases and returning the answers in English where they could be integrated. This project used a special protocol running over email systems (this was more than 10 years before WWW) and at that time networking speeds were rather slow.

However, the demonstration to Heads of research funding of the G7 countries led to a successor project, EXIRPTS (1987-1988), where the concept was extended to all represented countries. The technology involved each national database sending an agreed subset of information in a common syntax and semantics to form a common catalogue replicated at each site and updated there; queries accessed the catalogue, found the hits matching the predicate and returned the hit

list. The user then selected the records of interest and automatically from his local host system requests in the language, syntax and semantics of the target database in each country were sent and answers received and integrated. These two projects demonstrated clearly the need for management and harmonisation of syntax and semantics if the results were to be useful.

European CRIS

In the late eighties, informed by IDEAS and EXIRPTS, and stimulated by the conference of European rectors' conferences, the CREST working group of the European Commission (EC) set up a group of experts to agree a common data format for exchange of research information. CERIF (Common European Research Information Format) was released in 1991 as CERIF-91. The format consisted of one record per research project, and had other information as attributes. Subsequent experience with this format proved its deficiencies: the format was too restrictive in that it did not allow multiple occurrences of information in attributes and important relationship information was either lost or not recordable. It was restricted to English and a single (limited) classification system which itself rapidly became out of date.

The deficiencies of CERIF-91 led to the re-forming of the expert group in 1997 and the publication of CERIF-2000. This format is a fully relational structure allowing n:m relationships between all entities. Thus a single record describing a person can be related via relationships with role and time interval to publications (e.g. as author, co-author, reviewer), to research projects (e.g. as grantholder, project leader, researcher), to organisational unit (e.g. as leader, member). Furthermore recursive relationships can be represented such that the relationship between organisational units (e.g. a research group is part of two different departments within one university) can be recorded. CERIF 2000 allows any language for recording and allows multiple classification systems for any major entity/attribute combination. CERIF is an EU recommendation to member states (i.e. effectively an EU standard).

In parallel the ERGO pilot project, based on the protocols and techniques used in EXIRPTS, was implemented with approximately 100,000 records from 20 countries. Unfortunately the EC decided not to fund a full system.

euroCRIS

euroCRIS has existed informally since the first international CRIS conference in 1991. During much of the nineties euroCRIS existed as the euroCRIS platform for promotion of CRIS best practice. In 2002 euroCRIS was re-founded formally and registered its charter as a not-for-profit organisation in the Netherlands in 2004. Currently with more than 100 members, euroCRIS runs the biennial CRIS conference, organises annual strategic seminars and biannual members' meetings – usually associated with a workshop on a hot topic. euroCRIS was given by the EC the task of maintaining and developing the CERIF standard.

CERIF2006

CERIF2006 develops the CERIF concept to its logical conclusion by separating the semantic layer from the syntactic layer. This means that the semantics associated with e.g. the relationship between a person and a research publication can be expressed in any language and with any meaning – but the meaning has to be declared such that it can be understood by others, using the CERIF classification system.

CRIS at the Core of Research Information

During the late nineties and early years of this century, repositories of research publications were constructed, usually utilising the Dublin Core (DC) metadata standard. euroCRIS members realised immediately the potential of CRIS alongside such repositories to give contextual information and also to provide richer metadata which is structured formally (unlike DC) so that it is computer-understandable as well as computer-readable. Furthermore, in some research institutions it was realised that – using a CRIS – one could cross-link publications to research datasets and software, thus providing a very rich environment for the researcher or the reviewer.

During the early years of this century these ideas were extended with proposals that CRIS should be cross-linked to systems in any organisation for finance, human resources, project management, workflow, customer relationship management etc such that management information on research activity (i.e. research information) could be presented – drawn from all these various systems – via a CERIF-CRIS in a homogeneous form for use in decision-making.

euroCRIS now has a vision of CRIS at the core of other systems acting as the cross-linking database system providing homogeneous views over multiple heterogeneous systems in one organisation or across organisations. The vision extends to CRIS in an ambient, mobile computing environment and utilising GRIDs as the e-infrastructure such that CRIS is part of the workbench of a researcher, research administrator, research manager, innovator/entrepreneur, or media specialist.

CRIS: The future of Europe?

The aim of the ERA (European Research Area) is to achieve a free movement of research and knowledge (researchers, ideas, products, patents, publications, access to facilities and equipment) to parallel that of capital, goods and people (later with services added) in the spirit of the original European Economic Community. This implies information interoperation between heterogeneous national systems recording research information and avoidance of wasteful replication. It also implies interworking of national and EC-funded research initiatives.

The thesis is that research provides raw material for wealth creation and improvement in the quality of life. To this end among the 'Lisbon targets' in 2000, three percent of GDP spent on research by 2010 was included. A recent green paper (http://ec.europa.eu/research/era/pdf/era_gp_final_en.pdf) highlights what has been achieved and how much more needs to be done.

In the modern IST (Information Society Technologies) world, an e-infrastructure is assumed. It provides networking anywhere with superposed services. For research these services include access to research facilities, support of the research project proposal process, access to the research outputs (products, patents, publications) and support (information and processes) for the researcher, research manager, policymaker, research funder, innovator/entrepreneur and media specialist in their day-to-day work. CERIF-CRIS can provide the basis for these services.

CRIS Benefits

The benefit to Europe of interoperating CERIF-CRIS is great. Strategists can analyse the research funding of other countries and adjust their plans accordingly. Research Funding organisations can more effectively plan cooperating or competing programmes of research. They can adjust research priorities to coincide with the Framework programmes of the EU. They can also find

appropriate reviewers for research proposals and compare statistically research performance (measured by outputs against expenditure). Research managers at research institutions similarly can track performance and plan their strategies accordingly. Innovators and entrepreneurs can find ideas for commercialisation. Researchers can more easily find others working in the same or adjacent fields of research and track down their projects, methods, outputs (products, patents, publications) so stimulating competition or cooperation. This is particularly important in emerging multi- or cross-disciplinary research activities where the researcher is not working with their familiar community. The media can access a wealth of research information to provide substance for 'research stories' of interest to the general public – in so doing encouraging discussion of the moral, social and economic aspects of the research as well as the technological and philosophical.

CERIF-CRIS provide even more than this. Linked with publication repositories and repositories of research datasets and software, the CRIS provides context within which the outputs can be understood. The CRIS provides the ability to an enquirer to link the datasets and software to the publication and to 'repeat the experiment' in silico with their own data or software.

The CRIS provides the basic reference information to drive workflows – for research itself and also for associated administrative and management processes such as procurement, accounting, recruiting, staff management, project proposal generation, project management, publication, patent registering etc. Cross-linked with other systems within an organisation (such as finance, human resources, project management etc) the CRIS becomes the unifying core with homogeneous information presentation from heterogeneous sources.

In this way a CERIF-CRIS can provide the consistent, research-focused interface to the various actors in the research community, shielding them from the complexity and heterogeneity of other systems and providing an interoperating environment with the CRIS of other organisations.

Contact:

Anne Asserson
Universität Bergen, research department
e-mail: anne.asserson@fa.uib.no

Professor Keith G Jeffery
Science and Technology Facilities Council (STFC)
e-mail: keith.g.jeffery@rl.ac.uk