CERIF BEST PRACTICE

Anne Asserson

Department of Research Management, University of Bergen, Nygardsgaten 5, Bergen, Norway
Email: anne.asserson@fa.uib.no

ABSTRACT

CRISs (Current Research Information Systems) are becoming increasingly important for organizations that are related to research, such as funding organisations, universities, and ministries. A CRIS holds information on research activities, results of research, and competence. A CRIS is useful for assessing a person or department, to show the institution's activity, to monitor scholarly activities, and as a base for the development of research strategy. This could be from a local CRIS, national CRIS, or from interoperable CRISs. A CRIS will be really useful if it is structured and can interoperate with other CRISs. The CERIF model (Current European Research Information Model) is a structured model and is able to give statistics for planning, evaluation, and assessment within an institution or benchmarking among institutions. The CERIF CRISs are able to give multiple views, such as a researcher's CV and an overview of an institution's projects (ongoing or ended) with project partners on an organizational or personal level. The output publications of a project are given for an individual researcher or institution, with linkage to the full text (in the local repository) and a list of journals where researchers or organizations are publishing, events, and an annual report on an individual researcher. A CERIF CRIS is recommended by the EU for interoperability among CRISs. A CERIF provides a one stop shop for users and gives uniform access to full text publications and scientific data. A partial model for people, organisation, and results, not projects, can be used. It is recommended, however, to implement the full model. To secure consistent information, it is also recommended to establish authority lists for people (unique ID, name, organization, position, age, sex, etc.) organsations (name, acronym, address, etc.), journals (title, acronym, publisher, URL, etc), and books (publisher, acronym, address, county, etc.) in the CERIF CRIS.

Keywords: Code of good practice, Assess, Repository, CRIS, CERIF

1 INTRODUCTION

A CRIS (Current Research Information System) is a tool to manage research and research output and handle the day-to-day information management for research managers, administrators, and researchers. A CRIS can also function as an interchange format among networks of CRISs.

The purpose of best practice is to assist and advise builders and users of CRIS to obtain optimum results. It may be that for some purpose a stand alone CRIS is appropriate - perhaps for evaluation of research or for recording the research of an institution. The report "Code of Good Practice for Current Research Information Systems" from January 1998 resulted from an initiative of the European Commission DG XIII (CORDIS) and euroCRIS. It is also important to notice that a survey for a "Code of Good Practice for CRIS" concluded that CRIS providers have a need for a recommendation which:

- Covers a wider scope of research information, including relationships between networks;
- Is adapted to today's information society environment;
- And which allows implementation of emerging technologies.

Traditionally a CRIS has one major focus for entry. It could be focussed on projects, persons, organisational units or results to mention some of the possible main entities. A CRIS can have various implementations, such as IR systems, hierarchic systems of any kind, a relational DBMS, or hypermedia systems.

2 CODE OF GOOD PRACTICE (CGP)

The CGP has been developed as a guide for both new and existing producers of CurreCRISs, large and small. The intention is to focus clearly on the reasons for having a CRIS and on the main components of the system.

There are many organisations, large and small, participating in research projects throughout both Europe and the rest of the world. The continuing evolution of the supporting technology (in particular the Internet and web based technology) is making communication available to even the smallest organisation. The CGP has been created to establish a framework for encouraging interoperation and harmonisation among European (and other) research institutions. It should be regarded as good practice for existing research institutions to review how they operate from time to time as well as to make available their experience to new and emerging institutions.

The CGP should not be used in isolation and must be supported by expertise in the available technology, methodologies, and content in order to be successful. As the CGP has been developed to meet the needs of a wide range of organisations, there will always be some aspects of the CGP that are more relevant than others. The CGP should be applied according to the needs of the individual institution and adapted accordingly.

The CGP for a CRIS should cover change control, good practice, and configuration management. This paper will focus on Good Practice. The proposal of a CRIS will need to cover a Definition of Purpose, Identification of Users, and Definition of Content. The design plan needs a database specification, structure and presentation, classification and indexing, and search and navigation. The information processing plan needs to address a collection plan, data capture guidelines, and quality control.

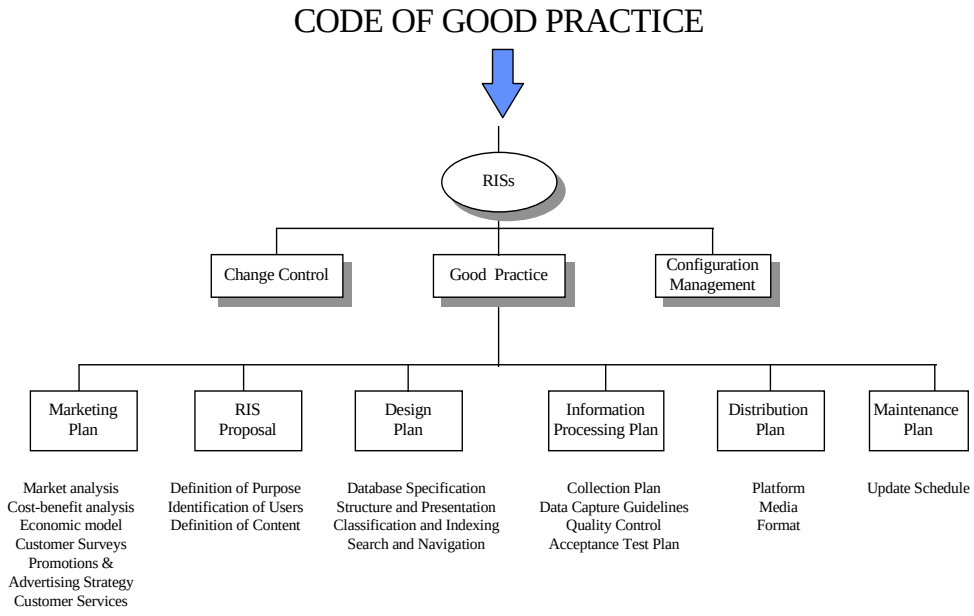

Figure 1: Code of Good Practice (CGP)

The CGP is an initiative of the European Commission DG XIII (CORDIS) and the European Platform for Current Research Database Producers (euroCRIS). The euroCRIS platform was established in 1995 to provide an informal forum for the development of research documentation practices and to stimulate European co-operation in this field.

The CGP was developed following a series of working group meetings with members of the euroCRIS platform and CORDIS and has been reviewed by a number of the CRISs themselves before being published.

3 PURPOSE OF THE CGP

The purpose of a Code of Good Practice is to record and document cases of practice that have been beneficial and successful, from these various experiences establish some guidelines that capture the accepted best practice, and on the basis of these guidelines assist new CRIS developers and users.

As a stand alone system, each CRIS plays an important role for the host institution or organisation. 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 realising 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 cornerstone of the CGP is consistency. This is the single most important benefit that will result from the adoption of the CGP. Consistency will lead to:

- Increased usability of data and value of CRISs to the users (often with varied demands and requirements);
- Increased interoperability between CRISs;
- Reduced operating cost for CRISs;
- Reduced effort for information exchange.

In order to realise the overall benefits offered by the CGP, it is first necessary for all relevant parties to adopt the CGP and then ensure its implementation within their working environment as a recommended standard or norm to be used. The increasing accessibility of information through developments in technology further emphasise the need for consistency (with regard to information exchange) to ensure that the wealth (and potential diversity) of information available locally is accessible globally.

A CRIS will be really useful if it is structured and can interoperate with other CRISs. The CERIF model (Current European Research Information Format) is a structured model and is able to give statistics for planning, evaluation, and assessment within an institution or benchmarking among institutions. This is achieved by recording cases of practice which have been beneficial and distilling the commonly accepted best practice by propagating the experiences as advice and assistance to CRIS developers and users.

CRISs are becoming increasingly important for organizations that are related to research, such as funding organisations, universities, or ministries. A CRIS holds information on research activities, results of research, and competence. A CRIS can also be useful for assessing a person or department, to show the institution's activity, to monitor scholarly activities, and as a base for the development of research strategy. This could be from a local CRIS, national CRIS, or from interoperable CRISs.

The definition of purpose at the institutional level makes a CRIS available as a tool for policymaking. A CRIS can also be used for the evaluation of an institution or an individual researcher. Evaluation based on research outputs is an increasing trend, and the performance based evaluation is increasingly used. The CRIS is also a tool to document the research activities at an institution. A CRIS can also function as a formal log of the research in process as well as historical projects.

The collaboration or co-authoring of a national or international researcher is valuable and important to document. Documentation of research output is also becoming more and more important, and this way of assessing an institution or a research groups is increasing. It is often a problem that an institution does not have a complete overview of its intellectual production. All this information is useful for assisting in project planning and to avoid repetition and duplication.

The purpose of a CRIS for an individual end user is to evaluate opportunities for research funding, to avoid

duplication of research activities, and analyse trends locally, regionally, and internationally. Also it shows the national and international co-authoring and cooperation in projects. Last but not least, a CRIS gives metadata and reference to the full text that is in an institutional repository. It can locate new contacts and new networks and identify new markets for research products.

4 DESIGN OF A CERIF-CRIS

A CRIS (Current Research Information System) should support the whole research process from funding, project information, and submission of articles, to storage of a peer reviewed article in full text within the institutional repository (IR). A CERIF-CRIS is a current research information system structured on the CERIF model. This well structured model makes it possible to document persons, organisations, projects, funding, and output from research: publications, patents, equipment, events, etc. Documentation of the contextual data is becoming more and more important and is a supplement to the traditional documentation of a publication: the need to know where the researcher was when the publications were produced, the relationship between the publication and a project, and the relationship between a group of researchers and their research output.

The CERIF CRIS are able to give multiple views, such as a researcher's CV and an overview of an institution's projects (ongoing or ended) with project partners on an organizational or personal level. The output publications of a project are given for an individual researcher or institution, with linkage to the full text (in the local repository) and list of journals where researchers or organizations are publishing, events, and an annual report on an individual researcher. A CERIF CRIS is recommended by the EU for interoperability among CRISs. A CERIF provides a one stop shop for users and gives uniform access to full text publications and scientific data. A partial model for people, organisation, and results, not projects, can be used. To secure consistent information, it is also recommended to establish authorized lists for people (unique ID, name, organization, position, age, sex, etc.) organsations (name, acronym, address, etc.), journals (title, acronym, publisher, URL, etc), and books (publisher, acronym, address, county, etc.) in the CERIF CRIS.

5 BEST PRACTICE ON THE IMPLEMENTATION OF THE CERIF-CRIS

When implementing a CERIF-CRIS, the advice is to implement the whole model and all the entities, even though they all may not be in use at the time. It is imperative that not only the entities but also all the links and language fields should be implemented. It may seem a bit much at the start, but the experience is that once they are there and CRIS is used at an institution, there is very soon an interest in the other parts. The binary structure of the CERIF model with the linkage between the entities gives the dynamic structure that is the strength of this model.

The CERIF-CRIS can be implemented using a subset or a superset of a full CERIF model. That means that only part of the model needs to be populated. Here are some examples:

For management purposes, it may be interesting to only populate the project part of the model, perhaps with the persons and institutions. To establish a researcher's profile or an expertise database, one could start with people. For an organisation that would like to present itself to the outside, the organisation entity is a start. To present the research output from an institution, such as publications, patents, and products, the result part can be populated. When populating the person and institution parts, one can establish the contextual data of bibliometrics and statistics. For services and facilities and particular equipment, it is possible to start with these entities to show what an institution can provide.

All relationships among the entities are role based with a time stamp. This makes it possible to document a person's change from one position to another within an institution or between one or more institutions.

The use of a classification scheme is recommended. The model is such that any classification scheme can be used. To facilitate retrieval it is recommended to add key words.

Experience with the CERIF model has shown that it does not cover all that an institution needs. Its binary structure makes it possible to extend the model. This should be done in the same structure, that is, add on new entities and link with attributes.

To be able to use the CERIF metadata for exchange of data, it is important that the core of the model is not changed.

6 THE QUALITY OF A CRIS

The next step is to establish a network of CRIS and link them together in a European or global portal. Also one may envisage the linking of CRISs to the e-infrastructure (networks, grids, etc.) and the physical research infrastructure (spacecraft, particle accelerators, synchrotrons, research ships and aircraft, databases, e.g., of economic or population data, databases of video of performing arts, etc.), which is of critical importance.

However, if the individual CRISs are of insufficient quality, linking them together has no advantage and may have disadvantages by giving a false presentation of the state of the research world of interest.

Therefore, the aim is first to get high quality (or at least adequate quality) CRISs, add associated repositories of publications, datasets, etc., link them together, and relate them to the research infrastructure and finally to research policies.

6.1 Authority lists

One important and advisable way of securing the quality of the individual CRISs is to use authority lists. An "authority list" is a controlled vocabulary of descriptive terms designed to facilitate retrieval of information.

The name and one unique instance of the name related to a unique identification a person are necessary. Names are presented in many different forms in articles, a variety of short forms depending on journalistic style. The recommendation is to have the full name of the person in the CRIS and have filters to produce the different reference formats, such as The Vancouver (International Committee of Medical Journal Editors) style, which is used primarily for publications in medicine, biomedicine, medical technology, and allied health sciences; the MLA (Modern Language Association) style, which is most commonly used to write papers and cite sources within the liberal arts and humanities, the Geological Society of America Bulletin reference style; *The Chicago Manual of Style*; and the science reference style, where journal names are abbreviated to save space.

Addresses that link to people or institutions/organisations should be picked from an authority list and recorded only once. Preferably there should be one authority site with the responsibility to update the addresses at the institutional level, where the changes are known.

The same routines should occur with the names of organisations. A research organisation will usually present three levels: the organisation’s name, the faculties, and the individual departments. There should also be an authority list of acronyms and short forms of the organisation’s name. As the CRISs are presented in a web interface and the information is made available globally, the names should be translated into different languages. Finally, the more obvious authority lists, such as those for country and language should be used along with other ISO standards.

6.2 Pre-classification

To avoid inconsistent names of journals and to avoid inputting an ISSN, short name form, etc., it is advisable to pre-classify the journals. The user should choose the journal in an interface where added information is made available on the pre-classified list.

The authority list should have the following attributes: journal name (full, acronym and short), the ISSN, where it is