



CRIS SEMINAR SEPTEMBER 2004
PALACE OF THE ACADEMIES BRUSSELS

CRIS + Open Access = The Route to Research Knowledge on the GRID

Information exchange through institutional
repositories and the European Research Area
(ERA)

Report September 2004



euroCRIS acknowledges gratefully the partners who co-developed the programme, provided speakers and other forms of support. Brief descriptions of the partner organisations can be found in the report of the 2003 Seminar <http://www.eurocris.org/en/seminars/seminar2003/>

Hospitality was provided by the Flemish Ministry Science Division for which we are extremely grateful.



Ministry of the Flemish Community
Science and Innovation Administration

euroCRIS acknowledges the kind provision of the venue (Palace of the Academies, Brussels) by the Flemish Academy.



ROYAL FLEMISH ACADEMY OF BELGIUM
FOR SCIENCE AND THE ARTS

These partner organisations provided (along with euroCRIS) speakers, other attendees, funding and other assistance which has made the seminar possible. The seminar partners agreed that a report should be produced by euroCRIS, agreed with the partners, published for all the partners and placed on the euroCRIS website together with the presentations.



www.eurocris.org

euroCRIS is the not-for-profit organisation promoting best practice in CRISs, spanning the field from raw experimental and simulated data through research management systems to research publications. Through its biennial CRIS conference and annual strategic seminars euroCRIS disseminates best practice. euroCRIS members' meetings share experience and set policy. EuroCRIS maintains CERIF, the Common European Research Information Format.

EXECUTIVE SUMMARY

OA (Open Access) and GRIDs are two of the hottest topics in IT at the moment. Both are extremely relevant to CRISs (Current Research Information Systems) and so euroCRIS took this opportunity to bring together a group of experts in all aspects of the subject area to share experiences and produce both insight and a way forward.

The seminar was extremely successful with intense but friendly discussions. OA was of particular interest to publishers, librarians and academics. Research quality assessment based on publications, citations and the ISI system engaged especially academics, research managers and publishers. More IT technical people held discussions on extending CERIF to incorporate OA using formalised metadata. Legalistic issues and the utilisation of GRIDs for CRIS + OA enjoyed wide interest.

All participants indicated that they had been informed, enlightened and stimulated by the seminar – not only in the formal sessions but also in the breaks and social events organised to permit free intercommunication in a less formal atmosphere.

The main recommendations are:

1. the partners will continue to evaluate both 'green' and 'gold' open access, but will encourage the use of 'green' institutional repositories anyway;
2. the partners will encourage the community to assure their IPR and copyright in order to permit respectively exploitation of R&D results and widest dissemination of scientific results;
3. the partners will encourage the community to utilise CERIF – and particularly CERIF extended with formalised Dublin Core for OA publications – in order to maximise interoperability of CRISs in Europe;
4. the partners will assist the community in setting up scientific workflows with incremental metadata collection in order to improve the metadata quality associated with scientific products;
5. the partners will encourage the community to utilise GRIDs as the IT surface to assist with CRIS + OA
6. the strategic partners will create opportunities for cooperative working to further their joint interests and improve the pan-European aspects of the joint endeavours;
7. following the success of this euroCRIS 2004 seminar, and recalling that of the 2003 seminar, an annual series should be continued addressing topics of importance to the research community in Europe;

1. PURPOSE

The rationale behind this theme is as follows: open access and archives, now economically and technically possible thanks to advances in technology, provide an opportunity and a set of challenges;

- a) will open access publishing be acceptable to researchers to make their work and themselves known, and to encourage scholarly dialogue;
- b) will open access publishing be acceptable to research evaluators forming 'league tables' of research organisations;
- c) will the open access archives be personal (self-archiving), institutional (knowledge of an organisation) or maintained by a scientific community (a learned society or a publisher acting for the community);
- d) will the open access material be available toll-free or charged;
- e) will reviewing be continued as now using panels of reviewers or will reviewing become spontaneous annotations attached to a publication;
- f) will access be through websearch and harvesting or through controlled metadata with thesauri;

The seminar addressed the relationship between CRISs (Current Research Information Systems) and OA (Open Access) Systems; bringing together systems for managing R&D with systems for providing open access to scholarly publishing – the major visible output of R&D – on the emerging European GRIDs infrastructure. The debate over OA is very active with 'green' (institutional repository self-archiving) and 'gold' (author / institution pays publishing) as competing but also complementary processes. The major publishers are experimenting with 'gold' services while 'green' institutional repositories are growing fast. GRIDs, especially through the NGG (Next Generation GRID) Reports (www.cordis.lu/ist/grids) has emerged as a vision for a European IT 'surface' now being implemented progressively especially under the auspices of DG INFSO F2 to provide easy-to-use access to information and computation. CRISs provide both a context for evaluation of - and understanding the background to – scholarly publication. CRISs also provide a management framework for R&D in institutions from funding agencies through national laboratories to universities, as well as a mechanism for interoperating research and development information.

In this topic there is, in addition to general interest, specialist interest partitioned as follows: ESF: publication quality of scholarly work in Europe; EARMA: evaluation of research, league tables, bibliometrics and scientometrics; ICSU/CODATA: publishing metadata standards and the interfacing to original scientific datasets; ALLEA: publishing through and for learned societies; ERCIM: IT to support the process, intersection with the DELOS network and associated projects. euroCRIS, with its integrating interest in all aspects of research information, pulled together our strategic partners. This synergy surfaced the issues and moved forward to resolutions.

2. TOPICS COVERED IN THE PROGRAMME

The programme covered a wide range of issues: technical, administrative, managerial, legal across academic, government and commercial perspectives. It covered scholarly publications and their evaluation including bibliometrics, through management of the R&D process and on to access to scientific datasets and software linked to traditional CRIS information.

Session1: What is happening – ‘Free Access’ *chair: Harrie Lalieu, euroCRIS*

- The new paradigms: Open Access and Institutional archives: - Gold or green publishing, Stevan Harnad, University of Southampton
- The Tension of Open Access: how not-for-profit publishers are reacting, Sally Morris, The Association of Learned and Professional Society Publishers, ALPSP
- The Position of SPARC, Raf Dekeyser, The Scholarly Publishing and Academic Resources Coalition, SPARC®

Session 2: What is happening ‘Payment Access’ *chair: Frank Heemskerk, EARMA*

- Author pays open access, Jan Velterop, Biomed Central
- Access for all by evolution not revolution, Derk Haank, Springer SPG

Session 3: The situation for the researchers *chair: Hans Karow, ESF*

- Chair’s introductory (Ref.: ‘euroSCIENCE Forum 2004’, Sessions on Open Access / New Publishing Needs) Hans Karow, ESF
- The Value Chain, Hans Roosendaal, University of Twente
- Can e-publishing meet the needs of the researchers, and ensure quality, through peer review? Marc de Mey, University of Gent

Session 4: ‘State of the Art’ Electronic Publishing and Research Documentation

chair: Anne Asserson, EuroCRIS

- What does a CRIS add to Open Access Publications? Keith Jeffery, euroCRIS/CCLRC
- How can CERIF facilitate access to institutional archives? Matthew Mascord, CCLRC
- Institutional evaluation and assessment of researchers, Gunnar Sivertsen, Norwegian Institute for Studies in Research and Higher Education

Session 5: Making it possible *chair: Keith Jeffery, ERCIM*

- Semantic Web and the GRID, Brian Matthews, CCLRC
- Open Access Initiative - Protocol for Metadata Harvesting, OAI-PMH, Andy Powell, UKOLN, University of Bath
- Legal Issues, Judy Beck, CCLRC

Session 6: EU perspective and the GRID *chair: Peter Wintlev-Jensen, EU*

- The EU plan for GRIDs in Europe, Max Lemke, DG INFSO F2

3. KEY DISCUSSION POINTS ARISING FROM THE SESSIONS

Open Access (OA) and the threat to Publishers

The cost of library subscriptions rises inexorably and faster than inflation; libraries are forced to cancel journal and proceedings subscriptions so creating a problem for researchers who find it difficult to gain access to the literature they require. However, researchers have two potentially conflicting requirements: to have their work disseminated as widely as possible (which favours OA) and to have their work published through channels with prestige (which favours conventional or 'OA gold' publishing).

The 'gold' route to OA (author institution pays) was discussed including the potential threat to quality should unscrupulous publishers increase volume for commercial advantage. The 'green' route as defined restrictively by Stevan Harnad (hereafter 'pure green') was discussed involving self-archiving, preferably in institutional repositories (for reasons of IPR, publicity and curation) parallel with peer-reviewed publishing. An alternative 'green' includes only institutional repositories - with 'grey' literature as well as peer-reviewed published literature - and a quality mark mechanism for peer-review based on publishers or learned societies. It became clear that there was a potential threat to publishers from this alternative and that many (including Biomed Central, and Springer) are responding by offering 'gold'; however, there may be no cost advantage - and indeed possibly a cost disadvantage - for institutions. All 'green' routes require harvesting / retrieval technology, such as OAI-PMH (Open Access Initiative Protocol for Metadata Harvesting).

All agreed the metadata currently used was insufficiently detailed and formal; detailed metadata assists in precise retrieval of relevant material; formal metadata permits more sophisticated processing including deduction and induction - and thus permits the system to act intelligently. The cost of metadata creation and maintenance was recognised; if metadata becomes more detailed the load on the researcher in inputting is increased. Incremental metadata collection through the process appeared promising.

Institutional OA repositories were favoured over subject-based OA repositories for several reasons: there is institutional interest in their creation, maintenance and curation for reasons of Intellectual Property (IP), publicity for the institution and management of R&D output (including possibly ensuring exploitation considerations - e.g. patenting - were considered). Furthermore, subject-based repositories can be provided at metadata or catalog level by harvesting (with retrieval criteria) institutional repositories and thus forming a 'subject-based view' over the stored primary publications.

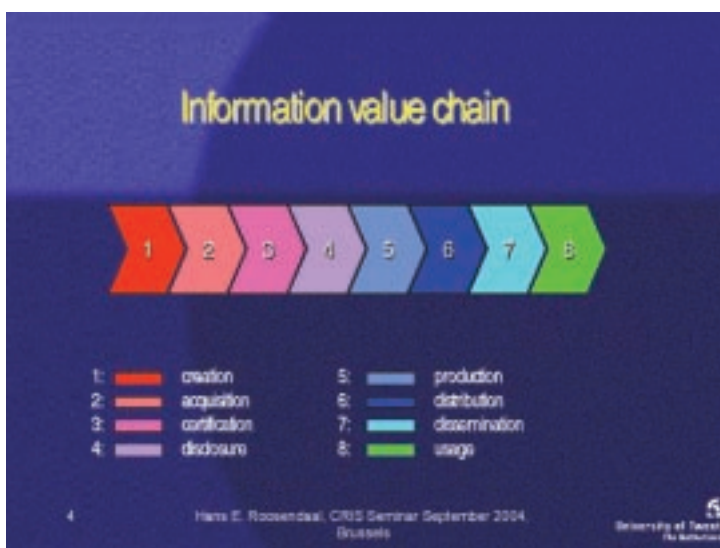
Peer Review and Evaluation

Both 'gold' and 'pure green' maintain existing peer review mechanisms. However, there is evidence that the system is not working optimally; there is the well-known pattern of one good, one bad and one neutral review from three referees, making it difficult to decide on publication. There was discussion of 'free annotation' peer review – possible given widely-used OA institutional repositories - but this found no favour and the present 'peer review college' system was preferred – at least until a better process is found.

On evaluation there was much concern that funding organisations might utilise the ISI system in an unsophisticated way – e.g. to partition funding among institutions - when it is known to have uneven coverage both across and within disciplines. There was interest in a European equivalent of ISI which could correct these imbalances. There was support for automated citation systems over open access publications as pioneered by citebase/citeseer. It was further noted that accesses to OA publications are measured conveniently and provide some evidence of quality. The key role played by CERIF-compliant CRISs in linking publications to persons, organisational units, projects, events, facilities, equipment, patents and products was recognised.

The Scientific Process as a Workflow

There was discussion of the workflow of steps in the publication process (whatever technology is used) and also of workflow in the scientific process from proposal to publication.

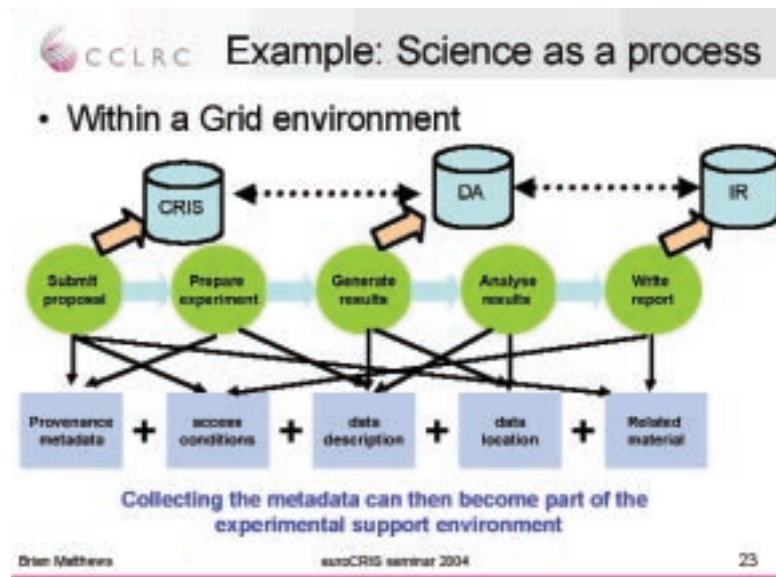


The information value chain (Roosendaal) was discussed representing the different steps of added value by the different stakeholders after the creation of the work by the author. The role of the university in research and education information management (institutional repositories) was discussed. Overall, the information value chain was welcomed as a basis for speculation on

the roles of the stakeholders in the chain, also including costs and benefits.

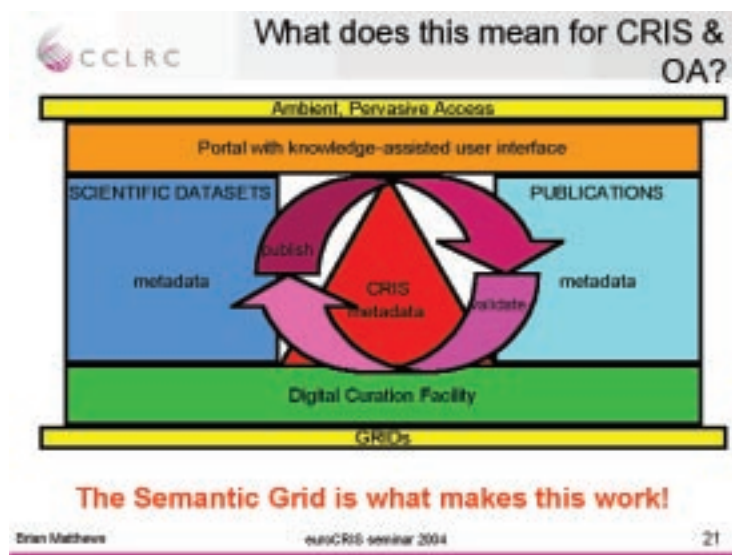
Discussion centred on the editing function of publishers which improves quality in readability, use of terms and layout and is not represented separately in the diagram, but is subsumed in production. Overall, the publication value chain was welcomed as a basis for discussion on costs and benefits.

The publication information value chain is only part (the final step) of a much larger scientific process from proposal through research to publication that may be characterised by this diagram (Matthews). This provides an infrastructure utilising GRIDs technology to support the researcher in the scientific process. This model found favour with the research manager community.



CRIS, OA and GRIDs

It was agreed generally that these three aspects together provided a sound basis for supporting R&D including researchers, researcher institutions, funding organisations, exploiters and research managers. CRISs provide the context and management information for appreciation of publications and with GRIDs technology the



publication via a CRIS can be linked seamlessly to primary scientific datasets and open source software relevant to the application from which the publication was derived. This permits the scientific validation required. The need for digital curation was recognised and endorsed; there is particular concern that changing media (due to improving technologies and reducing costs leaving some media obsolete) and changing formats of material (due to improving technologies and proprietary commercial interests) could render the research record incomplete. The increasing trend towards ambient and mobile computing was also recognised and endorsed: there is a need to provide access to CRIS+OA on the GRID anywhere, anytime, anyhow.

4. CONCLUSIONS

The participants concluded that the seminar topic was critically important to Europe and the ERA and that we should continue to cooperate to push forward the infrastructure to assure the future of Europe led by R&D supported by CRISs including linking them with OA institutional repositories. Specific aspects were as follows:

1. The need to push for 'pure green' institutional OA repositories because they do not impact publishers yet make publications freely available and because they encourage institutions to curate their intellectual property;
2. the need to evaluate further the true costs and benefits of 'gold' publishing; at present not all publisher charges are known and the potential impact in a published 'gold' journal may be greater than in a 'green' institutional repository with or without conventional publishing;
3. the need to extend CERIF as proposed by EuroCRIS members to incorporate formalised (and extended) Dublin Core to embrace OA repositories thus extending scientific management (CERIF) to OA;
4. the need to utilise a similar extension to CERIF to embrace primary scientific data sources and their relationship to the CERIF entities as well as OA publishing;
5. the need to improve quantity (detail) and quality of metadata to assist in retrieving relevant OA scholarly publications;
6. a similar need for quality metadata for primary scientific data and associated OSS (Open source software);
7. the need to find consensus on a scientific workflow and – within it – a publications workflow with incremental metadata input at appropriate stages;
8. the need to utilise GRIDs to improve the CRIS+OA environment with ready access to information, computation and primary scientific data in an easy-to-use environment;

5. ROADMAP FOR THE PARTNERS

Utilising the particular skills and interests of each partner, but working cooperatively, the partners intend to proceed as follows:

1. We shall implement 'green' open access institutional repositories alongside parallel 'gold' publisher open access and conventional publishing;
2. the partners will encourage their communities to assure their IPR and copyright (by retention rather than signing it over to a publisher) in order to permit respectively exploitation of R&D results (through patents, products or licensing) and unrestricted dissemination of scientific results;
3. the partners will encourage their communities to utilise CERIF – and particularly CERIF extended with formalised Dublin Core metadata for OA publications – in order to maximise interoperability of CRISs (including research products i.e. publications) in Europe;
4. the partners will assist their communities in setting up scientific workflows with incremental metadata collection as an integral part of the process in order to improve the metadata quality associated with scientific products;
5. the partners will encourage their communities to utilise GRIDs as the general IT architectural surface to assist with interoperable and economic realisations of CRIS + OA

In addition, following the success of this 2004 seminar, and recalling that of the 2003 seminar, an annual series will be continued addressing topics of importance to the research community in Europe; the strategic partners will create opportunities for cooperative working to further their joint interests and improve the pan-European aspects of their joint endeavours.

Postscript

The presentations, and list of attendees, are available at www.eurocris.org

The contact person for this euroCRIS-organised seminar is:

Ms Anne Asserson anne.asserson@fa.uib.no

euroCRIS IN A NUTSHELL

A Current Research Information System, commonly known as "CRIS", is any information tool dedicated to provide access to and disseminate research information, such as People, Projects, Organizations, Results (publications, patents and products), Facilities, and Equipment.

euroCRIS is established to address issues of current research information systems (CRIS) worldwide, but with emphasis on Europe. CRIS may be organized organisationally, thematically or along the lines of scientific disciplines. Issues are, but not limited to, databases: global, thematical and according to type of information (expertise, projects, institution, facilities and products - including publications); standards and guidelines; best practice; data access and exchange mechanisms; and other data standardization issues within the realm of research, training and development (RTD, R&D).

The primary goals of euroCRIS are to act as the single forum for all interested individuals and organizations to enter into dialogue concluding with resolution of all matters related to the use of information technology in the conduct of all research information system business. euroCRIS supports standardized, streamlined information exchange across all aspects of the CRIS lifecycle as follows:

- Promote and improve communication and interaction between global CRIS
- Maintain and publish the CERIF (Common European Research Information Format) recommendation and any standards endorsed by euroCRIS
- Organize and run the CRIS series of conferences with associated workshops and other events
- Provide a source of expertise in CRIS to members and to others under business arrangements made at the time
- Develop euroCRIS guidelines
- Nurture the CRIS community by all appropriate mechanisms
- Provide a forum for exploring and exploiting new and emerging concepts and technologies (including data quality, standards, etc.)
- Establish a one-stop portal / gateway to international CRIS resources

Membership is open to any organization or individual interested in CRIS - without geographical limitations. Structure, identity and procedures have been laid down in formal statutes (see Short Charter and Full Charter 2002).

For adequate and flexible functioning the completion of major aims and core actions is assigned to Task Groups, as stipulated in the statutes. At present five are established and fully operational.

euroCRIS Contact Address: C/o NIWI, Department of Research Information, P.O. Box 95110, 1090 HC Amsterdam, The Netherlands Phone +31 20 462 8657 Fax +31 20 665 8013

euroCRIS@euroCRIS.org

<http://www.euroCRIS.org>

