From CRIS to CRIS: integration and interoperability

Hornbostel, Stefan

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

From CRIS to CRIS: Integration and Interoperability

Stefan Hornbostel,
IFQ Institut für Forschungsinformation und Qualitätssicherung,
Bonn

Summary

Developments in technology, science policy and within the science sector itself have meant that there is both an ever increasing supply of research information and also an ever increasing demand for information for a variety of purposes. This has led to a blurring of the boundaries between information for the research proc-ess, evaluation for a public reporting and benchmarking for control processes. It is, however, not always possible to carry out new data surveys to satisfy the growing need for information without damaging the science system. Decentralised CRISs (Current Research Information Systems) geared to various different tasks can take on these tasks if they do not remain restricted at locally or within a sector but instead gain greater scope and usability through interoperability.

1 Introduction

Over recent decades, the scientific landscape in Europe has undergone fundamental changes. Far-reaching reforms to the science systems have been carried out in most countries, with the result that the entire system has become more competitive, more internationally and efficiency-oriented, and science policy has increasingly used an output-orientated control system (cp. Zimmermann 2002). The consequences of these changes can be seen in the fact that Europe now makes up a significantly larger percentage of global scientific publications and that citations for articles by European scientists have also increased. Western Europe's article output grew by about two-thirds from 1988 to 2001 and surpassed that of the United States in 1997 (National Science Board 2004, 5-39). The dynamics of this development have coincided with a rapid change in information and communication technology. Rapid exchange of information became important not only for the production of science, but likewise for performance information and quality assessments. Science management must increasingly turn to indicators which are based on quantitative information relating to research activity and quality. This does not mean that the traditional peer-review sys-tem is of no further use, but it is at the limits of its capacity. This consequently begs the question at various levels of the science system as to how it is possible to cope with the growing flood of information supplies and demands. It is scarcely possible to repeatedly carry out new data sur-veys, as they would be far too costly in terms of time and resources. CRISs offer a possibility to gather such information efficiently and to make it available for various different purposes.

2 Driving forces

2.1 Technology-driven acceleration of information exchange

The increase and distribution of scientific knowledge has been a central focus of research in the sociology of science for a long time. The development began with the discovery of book-printing, and has progressed via scientific journals to electronic communication and publication media. At present we are experiencing a fundamental change in scientific communication structures.

This relates not only to the speed with which information can be exchanged, but also to the new possibilities of electronic publication, of saving vast quantities of texts and then making them available all over the world almost instantaneously. This also applies to new possibilities regard-ing the automatic analysis of texts, the exploration of their themes, and their citation and usage. With the increasing amount of information, the strategies of assessment and selection of informa-tion have of course changed as well.

While traditional strategies of coping with the increasing amount of information used peer review on the one hand and review journals on the other hand, today new technologies offer very differ-ent strategies to mark quality, relevance, importance, reputation or usability. This process started with the science citation index. First designed as a retrieval system, it is today of particular impor-tance because it offers the opportunity to evaluate documents, journals, persons, institutions and even countries via citation analysis. Primarily a kind of search engine, SCI has been transformed into a rating machine which has almost monopoly status.

Such technical innovations do not change communication processes alone; they also change the social organisation and the distribution of power in the world of science. SCI is a good example for demonstrating this change. Nowadays, the reputation and sales of journals depend heavily on the journal impact factor which has been achieved, the importance of an article depends on the number of citations, and the negotiating positions of scientists as well as the reputations of Uni-versities depend on their performance in the SCI. This point has been criticised.

Particularly in the humanities and social sciences, SCI, originally developed as a service for North American scientists, has a strong language bias. Publication habits of those disciplines (mono-graphs and collections) are also not taken into account, address information has not been standardised. A hard-to-judge mix of mistakes and bias factors is influencing the specified quality of the SCI. Application-oriented research is hardly surveyed, and last but not least, most informal communication processes (working papers, grey literature) remain invisible.

There have been a variety of reactions to this unsatisfactory situation:

- In Europe, the invention of a European Social Science Citation Index was discussed.
- Recently, competitive products like SCOPUS have been launched.

- Instead of citation analysis, the power of evidence of „usage-metrics" is being examined (WEB-logs, link resolver, download metrics et cetera).
- With open-access publications, the markets of scientific journals are changing, as are the tech-niques of quality assurance (open peer review).
- Self-archiving now supplements publishing in traditional peer-reviewed journals rather than replacing it. Repositories for the self-archiving of scientific publications have been created by many research institutions.[2] Different initiatives are attempting to establish reference systems for these documents, and to develop techniques to harvest metadata and methods to mark the rele-vance of these publications. CRISs that cross-link the print material to other entities provide new options with regard to utilising the collection of publications for the scientists' purpose (cp. Jeffery & Asserson 2004).

In short, we are experiencing a development which will not only lead to a further increase in pub-lications, decentralised collection and saving, and easier access, but also to other techniques for judging the relevance of documents. Unfortunately, the heterogeneity of information, formats and standards is growing at almost the same speed as the amount of actual information. If this large variety of information is not to become a huge data graveyard, we urgently need Current Research Information Systems (CRISs), in order to link this information and make it usable.
Changes in information and communication processes are not induced by technology alone. They have gained momentum through a "push factor" which results from changes to governance struc-tures.

2.2 Policy-driven change of governance in research institutions

Over past decades, the governance structures of scientific systems in many European countries have undergone some fundamental changes. The trend is towards a type of higher education gov-ernance characterised by controlling the output of research institutions instead of the input or individual processes. The catchword is "accountability". That is to say, institutions gain more autonomy on the one hand and are being asked for appropriate output one the other hand, while the system itself becomes highly competitive and the budgets in most countries are more or less stagnating. This is not only a trend at national level but also within the European Research Area.
Research institutions are therefore forced to provide information not only on the input they re-ceived, but also on the research output and the quality of research carried out. Furthermore, these data are not just used to demonstrate achievements; increasingly, they are becoming internal man-agement tools to improve efficiency,

[2] E.g.: Caltech CODA Collection of Open Digital Archives: http://library.caltech.edu/digital/; Australian National University: http://eprints.anu.edu.au/.

to plan and rate performance, to allocate money and to schedule institutional profiles. Finally, the transition to flexible salaries is pushing the develop-ment of local information systems and is leading to institutional data collections – partly for offi-cial use, partly for internal management processes.
As a result, information on research produced on a decentralised basis is growing rapidly. Higher education institutions and other research institutions are developing CRISs of varying quality and performance. They are creating repositories for the publications and products of their scientists, offering information about their staff, publishing information on projects and are reporting on outside funding. More than 4000 universities in Europe are maintaining some kind of research information system.[3]
But it is not just research institutions which are under the pressure of having to report their activi-ties and make them visible to the public. The funding agencies are also developing their own CRIS, which are intended to provide information on projects, persons, institutions and financial resources. The German Research Foundation (DFG) is planning to offer information not only on projects which have already been granted but also on the scientific results of these projects.[4] The area of research funding is a good example for the interrelation between various information and evaluation policies: funds which have been granted on the base of a peer review are seen as an indicator of lively research activities and good quality. In rankings and ratings they are used as indicators for characterizing the research performance of university departments. However, the data with which the funding agencies are working do not quite match the organisational units which are the basis for the rankings.
While rankings are being used as entities organisations which have grown over time, funding agencies are aggregating their information at the level of a classification of research fields, which however is not identical with the field classifications according to ISI. Furthermore, both these systems are only compatible to a limited degree with official scientific statistics. It is only at a very high level of aggregation that the data "match". This level is not, however, suitable for rank-ings or for internal university management.

The following example from a large German university can perhaps show how bad the data situa-tion is and how little the data correspond with each other. The example contains information on the raising of outside funding by an institute for sociology.

[3] Recently the Austrian Agency for Quality Assurance (AQA, www.aqa.ac.at) and the Austrian "Plattform für Forschungs- und Technologieevaluierung" (www.fteval.at) held a conference with the major topic account-able aspects of Research data and -database. Conclusions to appear in: fteval-newsletter: 27, 2006, *Research Information Systems*.

[4] The system of the DFG, which offers information on projects, is called GEPRIS (= Geförderte Projekte Informationssystem). This internet database service provides information on research projects currently funded by the DFG (http://www.dfg.de/gepris/).

Outside funding data of a large German university (in thousand €)								
	1996	1997	1998	1999	2000	2001	2002	2003
Official Statistics	3.1						157.0	159.0
Ranking			513.7	513.7	372.4			
Data of the Institute					261.4	329.2	289.6	

The figures themselves are not of interest in this regard. The important factor is rather that none of the data from the various sources is identical. The first line contains the figures from the official statistics. Data are missing completely for a period of five years; after this, they do not correspond to that of the university. The second line contains the university's data. The data were recorded as part of a regular ranking process at the central university administration. It required considerable effort to ascertain the figures for the corresponding institute at this university. But these figures also do not correspond to figures of the institute itself (third line).

Without doubt, genuine errors were made in this example and there are doubtless universities with information systems which work better than this one. At the same time the example is definitely typical, as the actual problem is that different definitions and delimitations of the investigated units are used at all levels. The figures are consequently not necessarily wrong, but cannot be compared with each other.

Inconsistencies such as this were not noticed for a long time, as corresponding data either did not exist or were only used for purely administrative purposes. This changed with the spreading of rankings and ratings over the last decade. Rankings claim to create comparability. Even if rank-ings are of very different quality and meaningfulness, their effect is not to be underestimated: they have highlighted a high degree of inequality in countries with egalitarian – usually State – scien-tific institutions and initiated a kind of „objectivisation process" in other countries. This means that a reputation may be impaired if in the long run no evidence can be produced to warrant an exceptional place in rankings. However, ratings also have the altered governance structures to thank for their breakthrough. As the competition in the system increases, and as money and repu-tations are increasingly distributed and spread via indicators and media reporting, and as rankings become increasingly important as a means for protagonists to orientate themselves in the science system, so the comparative analysis of institutions, benchmarking, ranking and rating will increas-ingly gain in importance. In Germany, such rankings are now not created by the media alone; the DFG[5] issues its own rankings, and the Science Council[6] will be following its lead in the near future. Outside funding agencies are working

on monitoring systems which allow the output of the sponsored projects to be evaluated. This, in turn, necessitates bibliometric information and reliable data on the research institutions (personnel, funding, etc.).

Similarly to the rating agencies in industry, the rankings in science are a kind of power factor. And as in industry, the relevant institutions in the field of science are also trying to control whether their competitive position is represented appropriately (cp. Hornbostel 2001, 139pp.). For re-search institutions this means that there has been a significant increase in the need for flexible Research Information Systems which can be used for very different purposes. So far, however, the various different CRISs have been completely unconnected; they also work with different definitions, do not collate information, and make new recordings of all data.

This seems paradoxical as greater pressure from competition, greater demands for accountability, and tight budgets increase all protagonists' need for high-quality information. In actual fact, CRISs are, however, usually designed from the perspective of the respective institution, without any consideration as to whether the institution's own system will in the future have to be „able to communicate".

The value of future CRISs will not lie simply in the information but also in their ability to assess information and to relate it to other information. This means that such systems must be suited for benchmarking, that the validity of the information must be ensured and that there must be further development of indicators which meet the requirements of new forms of communication and which can also be used in the humanities and social sciences.

2.3 Science-driven crossing of borders

Over the last few decades we have not only seen an economy-driven rise in the importance of applied science, but also the blurring of boundaries between applied and basic sciences and also the blurring of boundaries between disciplines. Whether this is something fundamentally new (a la mode 2)[7] or merely a transfer of emphasis is not so important in this regard (Gibbsons et al., 154). The decisive aspect is that this development results in new demands for CRISs:

In respect of the usability of CRISs, the information requirements from the R&D

[5] For more details: Deutsche Forschungsgemeinschaft (2003): Institutions – Regions – Networks. DFG Ap-provals and Other Basic Data on Publicly Funded Research, Bonn.

[6] For more details: www.wissenschaftsrat.de, s.a. Recommendations for rankings in the system of higher education and research Part 1: Research: http://www.wissenschaftsrat.de/texte/6285-04-engl.pdf

sector must be taken into account. Even more important perhaps is the information requirements of small com-panies which do not have any large research infrastructure but are searching for cooperation part-ners in science. Often, other public protagonists, such as the media, approach the science system with information requirements. Seen from within the science sector, these developments mean that scientists are also looking for information, people or literature beyond the narrow specialist community with which they are normally very well-acquainted.

This places demands on a CRIS from two directions: on the one hand it is meaningful to also develop products which do not belong to the classical repertoire of reports, such as technical re-ports, procedure manuals, learning materials, software, data etc. Expressed in more general terms: traditional CRISs and scientific repositories need more interoperability. On the other hand, such information demands can only be serviced if the individual CRISs have joint (multilingual) classi-fication. They must at least have a metadata model which makes it possible to carry out a consistent pan-European search via different CRISs.

3 Consequences for CRISs

At present, we note that there is an increase in the importance of the push factor: an increasing number of institutions are offering an increasing amount of research information for a diffuse public, or are recording and producing very specific information for their own requirements. At the same time, the pull factor is also becoming stronger: a large amount of information is being requested for evaluations, benchmarkings, accreditations, research reports or funding formulas. A large amount of energy is now devoted to making research information available, without, how-ever, there being any real coordination between supply and demand. The load from never-ceasing new requests for information is now so great that many universities in Germany are refusing to participate in evaluations. The mismatch between supply and demand is linked on the one hand to the fact that a market mechanism – such as that in economics - only exists to a very limited de-gree. On the other hand, however, to the fact that the information requirements differ greatly ac-cording to user group, vary over time, and in many respects even develop in opposite directions. The open access movement, for example, broadly stands for an expansion of information, if pos-sible without restrictions; rankings and evaluations, on the other hand, stand for compression, evaluation, comparability and simplification. CRISs which want to react to these challenges must therefore be very flexible and must above all use the potential which lies in the networking of various different CRISs. In particular it is the possibility of using structured information from CRISs for evaluation purposes and for decision-

[7] Please find the attributes of knowledge production in Mode 2 by Gibbons et al., 41ff.

making processes which makes the difference when compared with pure search engines. Some of the most important requirements for the fur-ther development of CRIS are as follows:

1. There is far too great a load on scientists from the many different data surveys. CRISs must be structured so that information on people, projects, publications, patents etc. can be used in different contexts. They must be just as usable for local management proc-esses as for reports which are addressed to a general public. They must be able to serve external data requirements and be suitable for internal analyses. They must on the one hand open up access to very in-depth, detailed information (for instance the documents in repositories) and on the other hand be able to provide demand-dependent aggrega-tions.

2. CRISs must be in the position to differentiate and make evident the relevance of the in-formation which it provides. Many CRISs today are already trying to network the litera-ture data they contain so that (assuming access entitlement has been granted) it is possi-ble to see very quickly whether the publication is contained in the SCI and how many ci-tations were attained. In other cases attempts are being made – similarly to the journal impact factor – to at least distinguish between A and B journals to mark the relevance of the publication. Measured according to elaborated bibliometric technologies, this doubt-less does not constitute progress. However, if several CRISs are joined together in a group, then this produces completely new possibilities. It is only in a group such as this that user-metrics can be developed which do not have the weaknesses and possibilities for manipulation which are common with WEB metrics. On the other hand, search en-gines are no substitute for CRISs. They can supplement one another very well.

3. External users usually have very specific information interests. In particular for the pur-pose of communicating with industry, CRIS or „Meta CRIS" should be able to support a search for experts or specialised institutions in respect of specific questions.

4. CRISs must provide structured information for evaluation processes. This means that certain standards must be maintained in the classification. This function, in particular, cannot be replaced by WEB search engines.

5. The information in a CRIS must be up-to-date and valid. Whilst, in the case of rankings, the publishers are obliged to check the validity of the information, a diffusion of respon-sibility occurs very quickly in the case of decentralised systems. The best method – and the most efficient in the sense of a multiple use of data –is to have cross-validation. If a research project is registered

with an outside funding agency and in a local CRIS, then it is a project which has passed through a peer review process and which really exists. If both are skilfully networked, then applicant, collaborators, abstracts, final reports and the literature which has been produced need only be entered once and are then available both locally and supra-regionally.

6. The quality of a CRIS depends decisively on the benefit the scientists can gain from such a system, both in providing data and in accessing and using data. The qualities which CRISs have to offer include firstly making work easier, e.g. creating a literature or pro-ject list at the touch of a button. That alone will, however, not be sufficient as most scien-tists have tools such as these. A further step in making work easier consists in avoiding multiple recording of data. This can be achieved by linking up with other information re-sources. But only when CRIS is systematically incorporated into research administration, decision-making processes, evaluation and research reporting at institutional level will the scientists have a vital interest in regularly supplying valid information. As a user of such a system, one expects a clear additional benefit over the use of search strategies on the Net. The advantage which CRISs offer is the clear structuring when certain standards (e.g. CERIF) are followed. This advantage can, however, only develop beyond the local level if the information contained can be easily exploited at national or European level. This presupposes a high degree of semantic and technical interoperability.

7. This would mean that there would also be the possibility of leaving the sector of „visi-ble" research products (white literature, patents) which are important for evaluations, and, with the Open Access Technologies, develop links to grey literature, preprints and post prints via CRIS (corporate data repository). This rapid exploitation of research communication does not assume quality assurance via peer review, but instead relies on rapid, unfiltered communication. Linked with a CRIS, this would enable people, pro-jects, institutions and also white literature to be assigned to the current communications.

4 References

Björk, B.-C. (2004): Open access to scientific publications – an analysis of the barriers to change. *Information Research*, 9 (2), Paper 170, (http://InformationR.net/ir/9-2/paper170.html).

Buttler, L.; Visser, M.S. (2006): Extending citation analysis to non-source items. *Scientometrics*, Vol. 66, No. 2, p. 327-343.

Deutsche Forschungsgemeinschaft (2003): *Institutions – Regions – Networks. DFG Approvals and Other Basic Data on Publicly Funded Research*, Bonn, (www.dfg.de/en/ranking/index.html).

Gibbons, M.; Limoges, C.; Nowotny, H.; Schwartzmann, S.; Scott, P.; Trow, M. (1994): *The New Production of Knowledge*, London: Sage.

Hornbostel, S. (2001): Die Hochschulen auf dem Weg in die Audit Society. Über Forschung, Drittmittel, Wettbewerb und Transparenz. In: Stölting, E.; Schimank, U. (Eds.): Die Krise der Universitäten. *Sonderheft Leviathan* 20/2001, p. 139 - 158.

Jeffery, K.; Asserson; A. (2004): Relating Intellectual Property Products to the Corporate Context. In: Farace, D.; Frantzen, J. (Eds.): *Grey Literature Conference Proceedings*, Text Release, Am-sterdam.

National Science Board (2004), Science and Engineering Indicators 2004, Vol. 1, Arlington, VA: National Science Foundation.

Pöschl, U. (2004): Interactive journal concept for improved scientific publishing and quality as-surance, *Learned Publishing*, Vol. 12, No. 12, p. 105-113.

Wissenschaftsrat (2004): *Recommendations for rankings in the system of higher education and research*, Part 1: Research. Hamburg.

Zimmermann, E.H. (2002): CRIS-Cross: Current Research Information Systems at a Crossroads. In: Adamczak, W., Nase, A. (Eds.) (2002), *Gaining Insight from Research Information*. Proceed-ings of the 6th International Conference on Current Research Information Systems, University of Kassel, August 29-31.

5 Contact Information

Stefan Hornbostel
IFQ Institut für Forschungsinformation und Qualitätssicherung
Godesberger Allee 90
D 53175 Bonn

Humboldt-Universität zu Berlin
Philosophische Fakultät III
Institut für Sozialwissenschaften
Universitätsstr. 3b
D 10099 Berlin
Germany

e-mail: hornbostel@forschungsinfo.de
www.forschunginfo.de