DINI AG Research Information Systems (AG-FIS)

**Research information systems**
**at universities and research institutions**
**- Position paper -**

# Inhalt

**Executive summary** .......... **3**

**About** .......... **3**

Definitions .......... 4

**A. For strategists and decision-makers: Positions** .......... **4**

I: What are the current challenges in research reporting? .......... 5

Distributed data storage at institutions .......... 5

Management systems fail to map research contexts .......... 5

Institutional research databases are only of limited use .......... 6

Disciplinary and funding databases are disconnected .......... 7

There is a lack of interfaces and exchange formats .......... 8

Standardisation opportunities are not taken .......... 8

II. What would be an ideal environment? .......... 8

Institutional research documentation is part of the IT infrastructure .......... 9

Institution research information is collaboratively built up and managed .......... 9

Value-added services rather than extra work: IT support for researchers .......... 10

Researchers are mobile – their profiles need to be, too .......... 10

Link with electronic applications .......... 10

III: How could these current challenges be solved? .......... 10

Clarify the legal situation .......... 11

Rigorously implement integrated information management .......... 11

Define institutional guidelines .......... 11

Develop reference models .......... 12

Use identifiers to link existing knowledge .......... 13

Simplify imports and exports .......... 13

Future trends: Creating value through Open Data .......... 14

Conclusion .......... 14

**In short: Expectations regarding research reporting** .......... **15**

**B. For practitioners: Guide to introducing research information systems ... 16**
System landscape in Germany ... 16
Which system for which use case? ... 17
Rollout and operation ... 19
Investment and choice of system ... 19
Project management ... 20
Operating concept ... 20
Acceptance of reporting obligations ... 21
Creating added value ... 21
Availability (Data are comprehensively available) ... 21
Reliability (Data are reliable) ... 22
Topicality and consistency (Data are up to date and consistent) ... 23
Efficiency (Collection is efficient) ... 24
Sustainability (Collection is sustainable) ... 25
Security (Data are secure) ... 26
Collaboration and other exchanges of experience ... 27
**Authors and advisors ... 27**
**Annex ... 29**
Glossary and list of links ... 29
**Literature ... 32**
Recommendations and publications ... 32
Books and compilations ... 32
Other literature references ... 33

# Executive summary

Reporting has become a regular part of science at every level. Researchers are required to report to external funding organisations and sponsors. Management needs an overview of the multitude of research information available in order to be able to make sound decisions and compete successfully for equipment and funding. Public accountability, particularly in terms of financing, has also grown in importance over time.

At the same time, universities and research institutions still face major problems when it comes to providing information on research performance. The causes of these problems are often very similar at each institution – distributed data storage without any interfaces, management systems that fail to map research contexts, and limited usability of existing systems when it comes to carrying out differentiated analyses: Specialist and funding databases are managed independently of one another, interfaces and exchange formats simply do not exist, and standardisation options are seldom used when developing such systems.

The development of financeable and functional research information systems and, above all, the exchange of existing information are of equal importance as campus management or suitable HR and finance systems when it comes to IT development in scientific institutions. It is difficult to imagine institutions being able to manage processes requiring manual input and annual data requests in the long term. Reporting requirements are also likely to increase over time.

This position paper describes specific strategic steps that need to be taken in order to develop long-term research reporting information management processes in German research institutions. Common standards need to be agreed on as they are a prerequisite both for reducing the considerable amount of work required to run systems and for enabling mobile researchers to transfer their portfolio to various applications and different research institutions. The working group also devised specific practical tips on designing, choosing, introducing and running a system as well as advice with regard to project management. These tips and advice are aimed at institutions wishing to introduce or develop a research information system.

# About

This position paper is the result of work carried out by the German DINI Working Group on Research Information Systems (DINI AG FIS) from May 2013-Nov 2014. The twelve members from different areas of research management and the library community have collected and reviewed requirements for sound research documentation in Germany. The positions set out in this paper are partly based on a series of workshops – sometimes including cooperation partners[1] – where views and previous experience were exchanged, and insights into the current situation in Germany gained.[2]

[1] These include annual meetings of research management associations as well as workshops and presentations at annual librarians' day and Open Access days.

[2] Survey conducted during a Library and Information Science Master's thesis at Humboldt University of Berlin. 51 out of 88 universities with the right to award doctorates provided responses. Source: Sticht, Kendra: Investigation into the use of research information systems at universities in Germany, results of an online survey, Master's thesis, Sticht, Kendra. (2015). Einsatz von Forschungsinformationssystemen an Universitäten und Hochschulen mit Promotionsrecht in Deutschland. Ergebnisbericht.. http://dx.doi.org/10.5281/zenodo.13841

DINI AG FIS members are involved in European networks, projects and initiatives engaged in the digital transformation of science.[3] These experiences have also been incorporated in this position paper.

We would also like to thank the colleagues who responded to the request for comments to the first version of this paper in January 2014.

The first section of the paper is primarily aimed at decision-makers and readers interested in strategic questions. The second is a practical section with guidelines on how to choose and introduce research information systems.

### Definitions

Research information means information pertaining to research activities, i.e. so-called metadata about projects, publications, published data sets, infrastructures and people/teams.

The research information systems referred to in the title collect administrative and scientific information from different sources in order to provide a structured view of an institution's equipment and achievements along with its organisational units. They also offer a data pool for value-added services, especially web applications. Research information systems can be implemented as specialised databases or modular applications that link simple project and publication databases with expert profiles.

This paper covers research information for the purpose of institutional research reporting (not on country or community level). We understand research documentation to mean the process of collecting research information.

The terms "research data" or "primary research data" must be distinguished from research information as they occur during the research process itself and are stored as values, tables or documents in specialist repositories in much the same way as full publication texts. In research information systems they are handled in a similar manner to publications.

## A. For strategists and decision-makers: Positions

Research reporting has become a regular part of science at every level. Individuals and research groups, institutes and sponsors collect and manage a multitude of information about expertise, publications, projects and third-party funding. The apparent wealth of information is however distributed over a large number of different systems, media and formats, and everyday information needs can often only be met with a great deal of work.

At management level there is also an increased demand for differentiated information in order to back uo strategic development plans. Research reporting to raise awareness, i.e. with more of a marketing angle, has also grown more important; firstly, as a way of making science accountable to the general public; secondly, to boost reputations in terms of competition between universities and

3 Detailed information about the authors is available in the annex

research institutions. On top of this comes a variety of regular reporting obligations and occasional information requests at every operational level.

All of this is enough to make universities and non-university research institutions think about more efficient and effective ways of handling research documentation. A carefully considered and, where necessary, joint approach should be adopted, since everyone involved has limited resources and shares mutual interests. The first section of this position paper summarises existing problems, outlines an ideal scenario, and drafts approaches that could be adopted in order to achieve this ideal scenario.

## I: What are the current challenges in research reporting?

In 2011 the German Council of Science and Humanities (*Wissenschaftsrat*) recommended improvements be made to reporting systems and the ability to provide information in the event of receiving requests.[4] The recommendation was triggered by the apparent practical problems institutes and faculties faced when asked to provide data for the Council´s disciplinary research ratings. The observations made were typical of the complicated research reporting processes in place at many institutions. Here are some of the main causes:

**Distributed data storage at institutions**

Universities and research institutions have a plethora of information related to research equipment and achievements. When specific requests can only be met with a great deal of administrative work, this is usually due to the institution’s non-standardised, distributed data storage. Requests range from personal bibliographies and academic careers that are only available to researchers through to internal project or patent directories and on to globally available publication metadata from repositories and university bibliographies. For reporting purposes, data about the various persons involved are collected in different formats and with varying levels of detail. Inconsistent structuring of these data compilations and the different systems, media and formats in use (databases, text files and tables) mean that they can only be reused to a limited extent. As a result, information within the institutions is often collected more than once, meaning that far more work is involved or that data validity is compromised. This kind of research reporting is neither satisfactory nor efficient for institutions and their management teams.

**Management systems fail to map research contexts**

A few institutions in Germany have already started to pool their data in integrated information systems. Even the HR and organisation master data often give rise to problems due to the organisational structure oriented towards cost centres that deviate from the “lived” structure, or due to dual memberships, personal associations, scholars, cross-organisational programmes and other “temporary structures” not being recorded in management systems. This naturally makes it difficult to introduce IT-based research reporting.

[4] Wissenschaftsrat 2011, Empfehlungen zur Bewertung und Steuerung von Forschungsleistungen (Drs. 1656-11), p. 48 et seq.

**Institutional research databases are only of limited use**

The vast majority of systems covered by the term “research database” are optimised for a certain usage scenario.[5] Figure 1 shows some typical examples of simple cataloguing systems at universities.

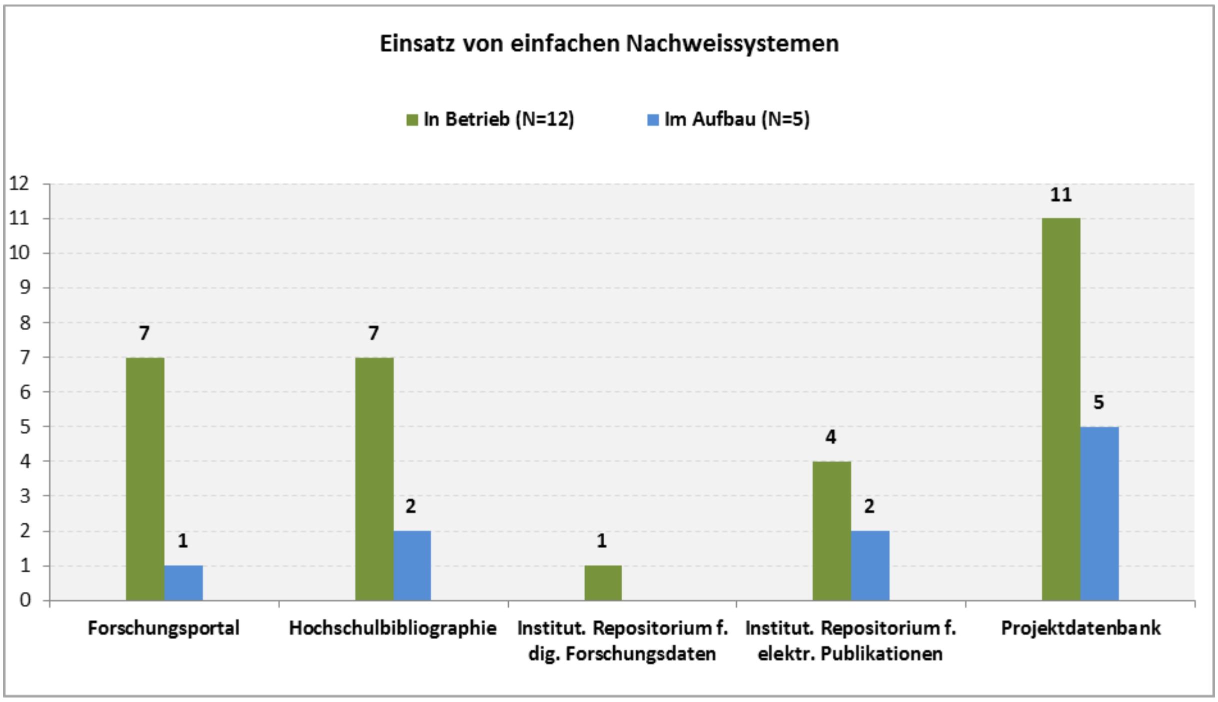


Figure 1: Use of simple cataloguing systems to document and report research activities at universities. These answers were provided by 51 out of 88 universities with the right to award doctorates. Multiple answers were possible.[6]

As they can seldom be linked via interfaces, most institutional research databases currently appear to be data silos of limited use for comprehensive institutional research reporting which, in turn is also to be used to supply information aimed at developing strategies.

However, more wide-ranging and diversified research documentation gives rise to a number of challenges: Firstly, for data protection and privacy reasons it is easier to introduce systems aimed at a specific and limited process such as resource allocation. The legal basis for documenting data on research performance for general “research reporting is currently not completely defined. Institutions determine or negotiate rules and regulations individually. Ideals like transparency and open data are areas of tension and currently out of balance with the need to protect personal data and the principle of data economy.

In addition, the introduction of an integrated research information system is an organisational development project which, at least according to current experience, is a major effort for any institution. It is therefore understandable that management is often apprehensive about pursuing this topic.

[5] Three usage scenarios were often stated during the course of an AG-FIS survey: 1) Public project and expert directories are used to support external online communication; 2) Collect and analyse parameters for internal allocation of resources or KPI systems (third-party funding, publications, etc.); 3) Manage university and institution bibliographies.

[6] Source: Sticht, Kendra, 2014, p. 4.

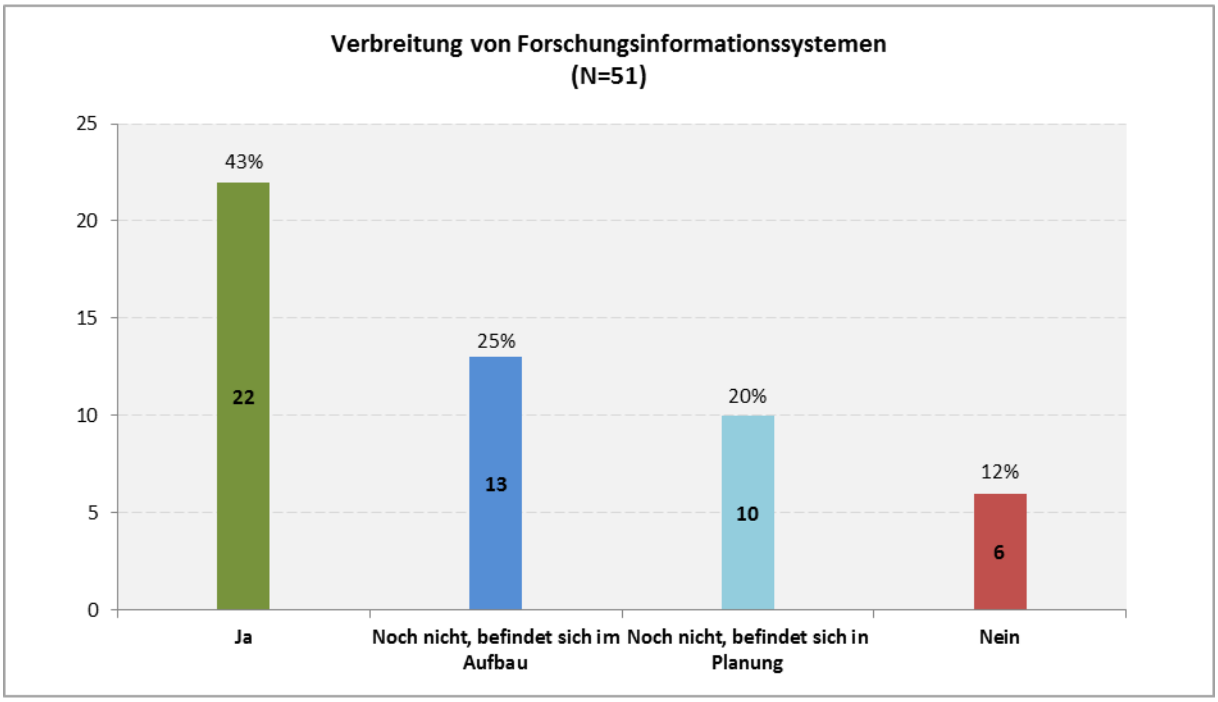


Figure 2: Use of research information systems in Germany - Out of the 88 universities in Germany with the right to award doctorates, only a quarter of them have a research information system. Another quarter of those stated that they were planning or in the process of setting up a research information system. Out of the 23 systems in planning or being set up, 13 are integrated research information systems and six are simple cataloguing systems.[7]

**Disciplinary and funding databases are disconnected**

Universities and research institutions can find high-quality information for their reporting in databases run by research-supporting institutions as well as commercial and non-commercial information service providers.[8] These sources offer a great deal of data for institutional reporting, but they remain largely disconnected from internal management systems and bibliographies.

Researchers generally collect additional research information such as personal bibliographies, project portfolios, reputation indicators and project results as and when required in order to gain funding. Interfaces that enable project coordinators or sponsors to reuse information that has already been collected are only actually established in a few cases.[9] By the same token, institutions are interested in using funding databases to complete their internal institutional system. Both of these observations also apply to the databases of national information institutions in Germany.[10] Their structure is primarily aimed at individual publication research, although they also offer special institutional assessment within the scope of the CHE University Ranking. There are no common usage scenarios where universities and research institutions can systematically correct these data or may use them themselves for research reporting purposes. As a result, most institutions currently have little influence as to the kind of information they can draw on from these information resources for use in their assessment.

[7] Sticht, Kendra, 2015, p. 2
[8] e.g. DFG-GEPRIS, CORDIS, Web of Science, Scopus
[9] The European OpenAIRE project links funding information from the CORDIS research framework programmes with publication information from repositories (cf. glossary).
[10] e.g. specialist databases SOFIS, FIS-Bildung and Psyndex

**There is a lack of interfaces and exchange formats**
In 2011 the Joint Science Conference recommended the introduction of open interfaces, standards and interoperability in order to ensure that digital research services can be reused.[11] These demands have largely been met when it comes to Open Access and literature management, but most research databases still do not have any standardised interfaces or exchange formats in place.

As a result, a lot of adaptation work or manual input is required to exchange research information. This of course means a major increase in the resources needed and used to continually manage institutional research information systems. This lack of transparency is also difficult to explain to researchers when they are unable to transfer their profile from one research database to another after changing jobs. A European research area that demands and supports mobility has to offer a certain level of transparency between research information systems. Action is urgently required in this regard so as not to lose the trust and acceptance of researchers.

**Standardisation opportunities are not taken**
The introduction of research information systems is currently linked to complex requirement assessments in which the wishes of a wide range of user groups need to be addressed and integrated.[12] So far there is no consensus with regard to basic reference models and functionality which could reduce and simplify this process, at least in German research institutions. The exchange of experience that has taken place to date has not led to any identifiable standardisation, although the recently conducted survey on research information systems does indeed indicate some kind of awareness for existing standards. One pragmatic step is the procurement of joint systems, e.g. at state level. However, a broader consensus would be needed to promote standardisation in general .

## II. What would be an ideal environment?

Well-organised research activity documentation reduces the burden for researchers and is supported by a service-oriented administration. An institutional research information system provides a structured view of all resources of interest to researchers, which can also boost their productive time. It provides management teams with information that can be taken as a basis for making strategic and operative decisions. It provides quality assured information for annual reports, regular reporting requirements and web applications, and routinely allows various requests for information to be met, such as the following:

- What cooperations does the institution have in place with country Y?
- What third-party funding for X has been received in the last three years?
- How many projects were carried out with international partners?
- How many scientists were involved in EU-funded projects?
- How large is the share of annual foreign-language publications?

[11] cf. GWK (2011), p. 50.
[12] An experience report published by the University of Hamburg impressively illustrates this. Stiehl (2011): Anforderungen an ein Forschungsinformationssystem am Beispiel der Universität Hamburg. The slides are available online: http://www.dini.de/fileadmin/workshops/forschungsinformationssysteme/Stiehl_Hans-Siegfried_DINIiFQ_WS_KIT_V2_3.pdf (last viewed on 21/10/2014).

**Institutional research documentation is part of the IT infrastructure**

In an ideal environment, research documentation is of equal standing with data processing when it comes to HR, finance and teaching (cf. Figure 3). IT landscape components are ideally connected to one another via interfaces, thus saving manual follow-up work (management systems, campus mangement, identity management). External information services can also be connected in a useful and effective manner.

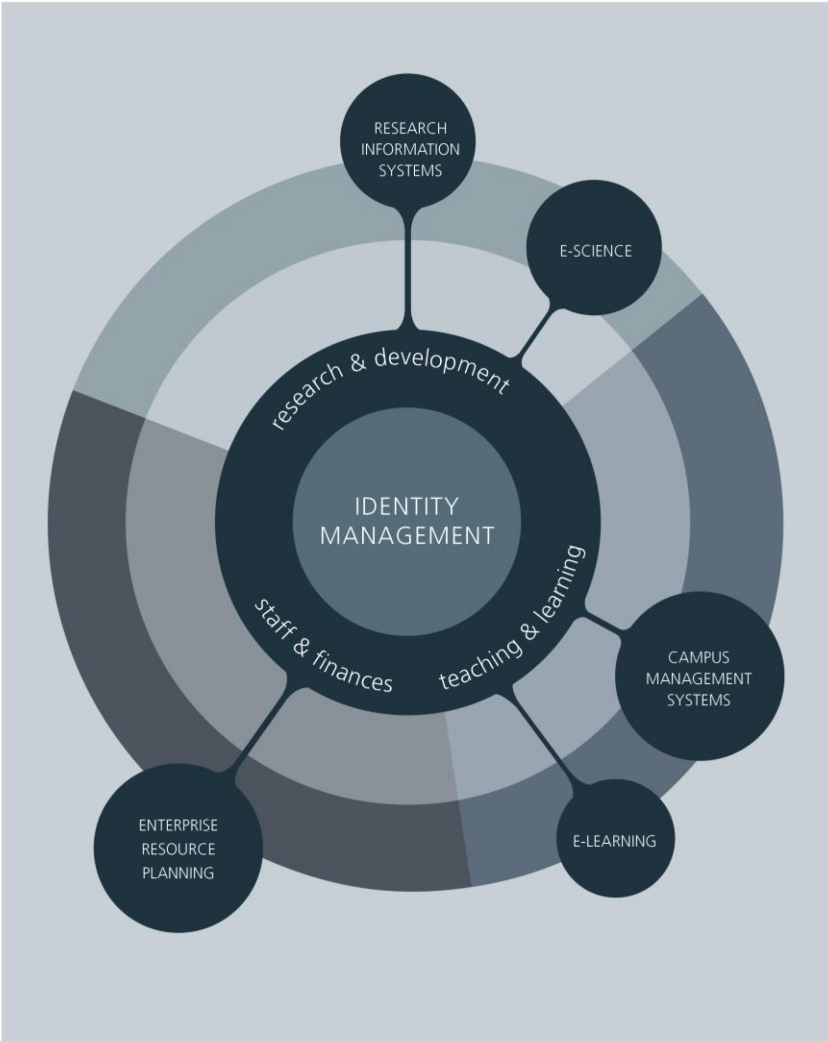


Figure 3: Research information systems and virtual research environments (e-Science) complement the integrated information management portfolio in terms of academia and teaching as well as their administration. They have interfaces to academia and teaching for upcoming academics (e.g. doctoral study programme) as well as to HR and financial administration in order to import relevant administrative data.

**Institution research information is collaboratively built up and managed**

Institutions have an institutional "body" of research information that is available for various research reporting purposes and permits an up-to-date and retrospective overview. By also logging organisational units and historical information, the research information system serves as an archive and differs fundamentally from the information provided on institutional websites.

Science and administration pool their information and then approve it within the scope of a comprehensive research reporting rights and roles concept in line with data protection and privacy law. Information is collected starting with individual objects (e.g. projects, publications, expert profiles, dissertations). The systems' output and analysis functions allow to aggregate the data as and when necessary, while also enabling a standardised display for internal and external reports. Centralised identity management sends organisational master data so that changes to employment relationships and organisational structures do not have to be made twice. This means that a cultural shift is taking place at universities and research institutions whereby they are moving away from distributed data storage to distributed management and use of a shared resource.

**Value-added services rather than extra work: IT support for researchers**

A service-oriented IT environment also provides research information to researchers' personal working environment, thus creating new kinds of scientific IT support. This gives rise to use cases such as the use of validated publication data on institutional websites or for personal CVs as well as the option for various data to be reused in virtual research environments. In an ideal environment researchers can play an active part in collecting such information. In order to simplify this task, the selected IT systems' working environments and user interfaces are both intuitive and up to date. Processes and structures are service-oriented and closely linked to science, while also easing the administrative burden on scientists.

**Researchers are mobile – their profiles need to be, too**

If we assume that data collection starts with each individual researcher, the exchange formats in particular have to be established on the basis of each individual portfolio. Agreement with regard to common standards would make it easier for mobile researchers to transfer their portfolio to their new employer's systems, thus obviating the need to re-enter this information manually.

Individual research information needs do not just involve an entry on an institutional website, they also extend to profiles on specialist portals and scientific social networks. This means that such an agreement on common standards is not just reached between universities and non-university research institutions, it also offers other providers an opportunity to investigate and put forward appealing solutions.

**Link with electronic applications**

In compliance with data protection and privacy law, the interaction between sponsors, institutions and applicants is transparent and enables existing information to be reused. Electronic applications are seen as a way of exchanging data between applicants, the institution submitting the application, and funding agency. Information of mutual interest such as project metadata, project results (data, publications, events), status, organisational affiliation and researchers' specialist fields are electronically exchanged and updated within this triangle.

## III: How could these current challenges be solved?

The research reporting impediments described above show that university and scientific management teams need to take action. The current interest is on information systems which are at best fully integrated and incorporated into the research environment. But even simple systems such as project databases and bibliographies can be used to improve information management if they can be modularly linked to provide overarching research reporting. Each institution should weigh up the amount of project work and financial investment required to set this up. DINI AG Research Information Systems (AG-FIS) looks at the practical aspects of system choice and introduction issues in a guide published in parallel (cf. Section B).

Irrespective of the individual system solution, we consider the following points to be in most urgent need of action:

**Clarify the legal situation**
The uncertainties that exist with regard to using mainly personal data with research documentation need to be tackled, especially if said data is to be used for external purposes and on the internet. Data protection and privacy are major issues when it comes to the internet and big data, meaning that the benefits of “transparency” and “improved accessibility”, which have frequently been requested, have to be weighed up. At the moment there is little specialist literature available on the subject and this gap in the triangle involving data protection authorities, legislation and users should be closed.

**Rigorously implement integrated information management**
Institutional research information systems are part of the scientific institution’s IT and organisational development agenda. They close the gap in terms of rigorous categorisation of research activities that is insufficiently mapped by management systems. The introduction and management of such systems inevitably entail a certain amount of work and financial expenditure. However, the work and expense involved can be weighed up against the clear cumulative benefits if several research reporting use cases can be included and various organisational units have access to improved data storage options. Users also stand to benefit from improved quality and services.

If research information systems are to reach their potential, they particularly need to be linked to master administration data. If individual components such as publication, project and patent databases are linked to HR and organisational master data, they can — depending on the resources and local pressure to come up with solutions — be concurrently used for research reporting. This reduces the amount of work involved while also increasing the chances of success. Adequately structuring the project phase and subsequent routine operation also increases the chances of reorganising research documentation in the long term, which will also make it easier to retrieve information. This can only be achieved if management teams are firmly committed to doing so.

**Define institutional guidelines**
Research documentation will always also be discussed within the context of the underlying reporting processes, thus giving rise to the need for acceptance within the institution.

It is easier to improve research documentation efficiency if the reporting purposes and therefore the use scenarios are known and in principle accepted. Trust will come about if the data are agreed upon, quality-tested and their intended use(s) transparent.

Such uses include value-added services for researchers. A research database has to swiftly convey to researchers the benefit they stand to gain during their daily research work at institutions or associations, e.g. reuse of data on website, and should also ideally generate a broader public impact, e.g. by being linked to relevant specialist portals or by disseminating selected details via research profile services.[13]

Institutions' management teams are responsible for developing and negotiating corresponding information management guidelines.

**Develop reference models**

Recognised reference models and standardised requirements catalogues for research information systems boost the level of transparency between systems and make it easier to introduce future reforms within the IT system landscape. Institutions can use reference models to simplify the collection of individual requirements and the development of new technical applications.

The following developments are recommended for developing reference models and requirements catalogues:

- The German "Research core dataset" project is working to define a minimum collection of research information that institutions should routinely provide for reporting purposes.[14] Such a core dataset represents a useful starting point for standardising an institution's research documentation and for exchanging data between institutions. Since August 2013 the specifications are been worked out in a project funded by the Federal Ministry of Education and Research (BMBF). First results are due to be published in mid-2015.

- The European CERIF (Common European Research Information Format) standard specifies the typical information objects collected in a research information system, including their temporal and contextual relations with one another, e.g. between projects, publications and infrastructures.[15] CERIF XML can be used as an exchange format.[16] For reasons of sustainability, this standard should be used as a reference when developing internal data models and definitions. Due to its complexity, CERIF was initially rather slow in terms of its dissemination, but is now growing in importance on an international level.[17] The German research core dataset also refers to CERIF.

- The VIVO Group, based in the US, has developed Open Source software for research profile services that is based on linked data technology. This Open Source software links publicly available research information and also offers the option to serve data such that they can be used by different institutions.[18] Taking account of the necessary standards in terms of reference models for research information systems leads to a major increase in the transparency and reusability of research information.[19]

---

13 A research profile service links and optimises research information for online retrieval. Relevant terms within this context include semantic web, Internet of Things (IoT) and linked (open) data – cf. glossary. There are currently few research profile services available in Germany.

14 Project website: http://www.forschungsinfo.de/kerndatensatz/en/index.php?home - cf. glossary.

15 CERIF is developed and disseminated by the European Organisation for International Research Information (euroCRIS): http://www.eurocris.org – cf. glossary

16 A CERIF XML interface was developed for the OpenAIRE European research database, cf. Open AIRE Guidelines for CRIS Managers (https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS , last visited on 26/10/2014)

17 Germany, the UK and the Netherlands all have official recommendations on using CERIF.

18 http://www.vivoweb.org/about

19 VIVO data model support examples: http://www.elsevier.com/online-tools/research-intelligence/research-initiatives/vivo and http://cns.iu.edu/docs/presentations/2014-borner-vivo-reuters.pdf and https://github.com/Symplectic/vivo

There needs to be more than just local agreement on reference models. Universities and research institutions need to launch joint initiatives where the people involved reach binding agreements and continue to develop standards. In Germany, for example, an upcoming task will be to use specific examples to check how the German research core dataset can be managed.

**Use identifiers to link existing knowledge**
Research information is often collected and disseminated more than once. If clear and permanent identifiers are used consistently, these items of information can be allocated and linked to one another. A similar situation applies to data stored in internal management systems and required in various contexts, e.g. personal records.

Standards are already available on an international level, meaning that, for example, the Open Researcher and Contributor ID (ORCID) can be used to import and export personal bibliographies as many publishing houses already use these identifiers to identify authors.[20] Uniform Resource Name (URN) and Digital Object Identifier (DOI) are also common for documents, as is GeoNames for locations. Universities and research institutions can use persistent identifiers to link their internal systems with one another and publish machine-readable links to working groups, research plans, buildings or other objects from their research information systems. This approach makes it easier to design interfaces, enables research profile services to be supported, and promotes the “mobile scientist” idea.

Institutions should also agree on the standards to be used in this regard.

**Simplify imports and exports**
The amount of work institutions and scientists have to put in to manage and update research databases is higher than actually necessary given the current state of technology. Standardised exchange formats and interfaces would make it easier to import and export information to and from other systems. Functional import and export formats have already been established for publication databases and catalogues, e.g. RIS, BibTex. By default, publication repositories have open interfaces[21] that allow bibliographies to be harvested, in turn allowing publications to be disseminated via external services.

Something along those lines should be developed for profiles stored in research information systems. Applications for such formats are value-added services for researchers who want or have to be present in several databases and social networks, as is the use of funding databases as an attractive source of data for research information systems.

This is another area where institutions should collaborate with one another. In order to use, for example, CERIF XML as an interface standard, a common vocabulary or data profile first needs to be agreed on. Such a vocabulary for OpenAIRE has already been defined in Europe.[22] A “Research core dataset” is developed in Germany, while on an international level the CASRAI transatlantic initiative is currently working on a common vocabulary.

---

[20] http://orcid.org/ - cf. glossary

[21] OAI Protocol for Metadata Harvesting, developed by the Open Archives Initiative. There are several thousand OAI-PMH providers worldwide as well as national research hubs that provide publication metadata.

[22] cf. OpenAIRE text in the glossary

Initiatives such as euroCRIS, OpenAIRE, VIVO and ORCID will increasingly add these developments involving their member institutions to their activities and work together to improve the exchangeability and standardisation of research information on an international level.[23]

**Future trends: Creating value through Open Data**

The above recommendations already outlined that research information is not just required within an institutional context, to some extent it is also a public good and should therefore be disseminated.

A number of databases are already available online, but the lack of interfaces makes it difficult to ensure a flow of knowledge. However, reusability models for existing research information help to make information easier to find and also help to disseminate current scientific knowledge.

Access to existing knowledge is expressly demanded and promoted within the context of Open Science[24]. In an ideal environment, research information systems should also form a link in the Web of Data[25] value chain and enable open access to publications and primary research data as set out in the Open Access guidelines. Such effects have already been observed in European initiatives such as Open Library Data, Open Educational Resources and Open Government Data.

Future trends also include semantic web applications that create information offerings by aggregating information from the web. Such applications do not require users to collect data again, meaning that they ease the burden on the system, reduce the amount of work involved in collecting data more than once, and make research easier to find within the sense of Open Science.

**Conclusion**

The digital shift taking place in the world of science not only affects the way in which scientists conduct and document research for their own purposes, it also creates new digital ways to collect and reuse research activities while also acting as a modern form of reporting for academic institutions. However, little use is currently being made of these options.

Electronic processing in science and administration will continue to grow in importance, as will the development of regional and national information services. It is difficult to imagine institutions being able to manage processes requiring manual input and annual data requests in the long term. Reporting requirements are also likely to increase over time.

IT-based research documentation needs to adapt to this digital shift in good time. At the same time, it is a component of the institutional IT system landscape that makes it easier for scientific institutions to handle complex administrative processes.

Local pressure to come up with solutions must not, however, lead to individual solutions that would force researchers into transferring or re-entering their data every time they change location. Common standards need to be agreed on as they are a prerequisite both for reducing the considerable amount of work required to run systems and for enabling mobile researchers to transfer their port-

---

23 cf. euroCRIS Newsflash 67

24 cf. Science as an open enterprise", Royal Society report, 2012. https://royalsociety.org/policy/projects/science-public-enterprise/report/

25 cf. Stocker, A., Tochtermann, K & Scheir, P. (2010): Die Wertschöpfungskette der Daten. HMD Praxis der Wirtschaftsinformatik, 47 (5), p. 94-104. http://dx.doi.org/10.1007/BF03340517

folio to various applications and different research institutions. However, reusability models for existing research information help to disseminate current scientific knowledge, thus assisting research itself.

DINI AG-FIS sees itself as a specialist German-language community that brings together the people involved in such undertakings and provides them with a forum to exchange views and ideas. Through its collaboration with euroCRIS and various institutional partners, DINI AG-FIS is involved in discussions within the European research area.

# In short: Expectations regarding research reporting

| | |
|---|---|
| **Data are available** | Data for research reporting are available in sufficient detail and a limited number of formats so that they can be analysed for various purposes. Guidelines stipulate which parts of the research information are to be offered outside of the institution. |
| **Data are reliable** | The data used for research reporting are reliable and can be verified. In an ideal scenario, data should not have to be stored twice or more in centralised or decentralised systems. |
| **Data are up to date and consistent** | Information needs to be up to date if it is to be used for internal and external research reporting. (Outdated information tends to be of more harm than use.) Information about completed projects and former staff is required for retrospectives and time-series analyses, i.e. data should be stored and readily available. |
| **Collection is efficient** | Research reporting is organised such that researchers are relieved of the burden of having to collect data more than once, which also prevents any incompatible data being collected on different occasions. The efficiency of the collection processes can be boosted by integrating IT systems and standardising the collection of data. |
| **Collection is sustainable** | Standard data collection processes can be reliably carried out, i.e. with realistic staffing capacities and resources. National and international metadata standards are taken into account when storing data in order to ensure a certain level of compatibility between systems. |
| **Data are secure** | Basic data ownership requirements are met as researchers retain control over their own data and are responsible for authorising anything related to the use of their data. Privacy is ensured by regulating access for internal and public purposes. |

# B. For practitioners: Guide to introducing research information systems

If a research institution is planning to introduce a research information system, it has a number of organisational and process-based issues to clarify and also needs to choose a suitable IT system. Depending on the size, equipment and expectations, this gives rise to the question of “how” to go about such an undertaking. This section of the position paper provides an overview of the basic expectations and conditions linked to the introduction of systems aimed at improving research reporting which need to be observed when introducing and managing systems.

## System landscape in Germany

Over the last few years, a few institutions in Germany have already started to introduce integrated information systems. As is the case with other systems in the IT environment, there is a wide range of possible solutions, such as inhouse developments, commercial products and services, and linkages between administrative systems.

A recent survey of universities provided an insight and showed that most systems document people and organisations along with their corresponding publications and projects. In some cases other information objects such as patents, expert profiles, external partners and internal university data about doctoral students, research funding and awards are also collected (cf. Figure 4).

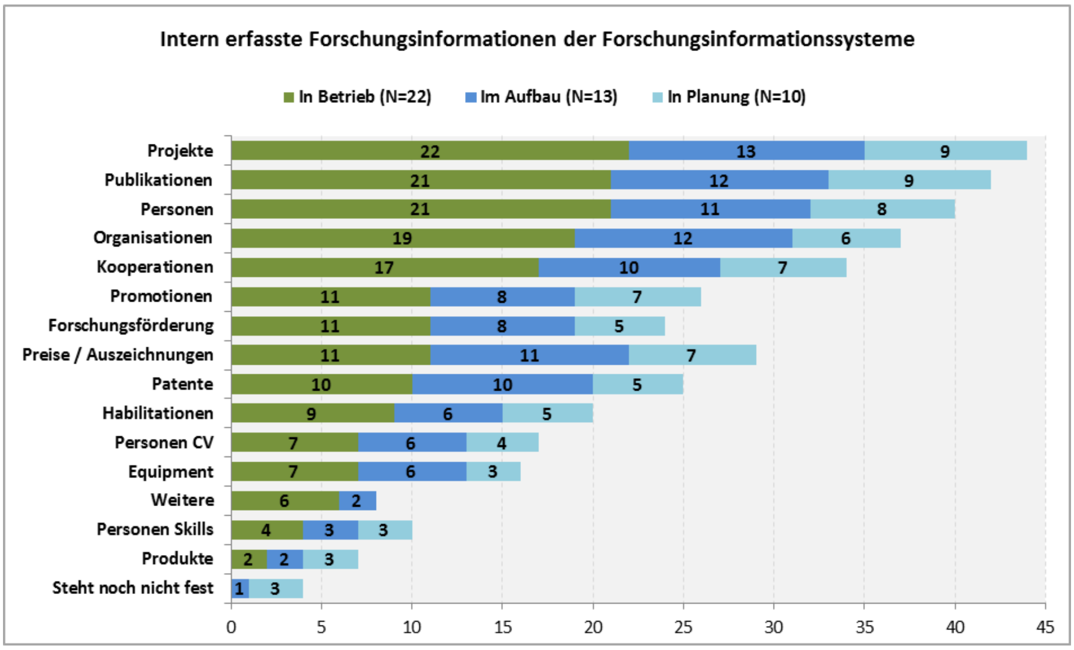


Figure 4: Which items of research information are collected in internal research information systems? The answers shown here were provided by 51 out of 88 universities with the right to award doctorates.[26]

[26] Sticht, Kendra, 2015, p. 7

Internal applications focus on reporting methods and decision-making processes, including annual and activity reports, reports used for evaluation purposes, internal allocation of resources, audits, rankings as well as management of project and publication directories on official institution and personal websites.

External user and interest groups (stakeholders) include the (specialist) public, institutional marketing, transfer and research centres, academic management teams, administrations and libraries as well as researchers themselves.

Each research institution governs the various applicable responsibilities. In many cases it is governed by the planning and research departments, although libraries are becoming increasingly involved due to their skilled staff who also act as information service providers. The research institutions surveyed stated that they use internal developments as well as commercial products in their systems.

## Which system for which use case?

As part of the institutional information systems, research information systems provide information for research reporting purposes. On an international level, the Current Research Information System (CRIS) is used to describe research information systems.

In view of the variety of systems also covered by the term "research database", three different types of research information system should be distinguished from one another in order to clarify the situation:

- Simple cataloguing systems (such as university bibliographies or research portals)
- Research profile services (such as Linked Open Data applications)
- Integrated research information systems with multiple output and analysis features (cf. Figure 5)

*Simple cataloguing systems* collect two or more information objects, but are optimised for individual use scenarios and not linked to one another. Examples include traditional research portals, university libraries, patent databases or publication and research data repositories. Reporting and analysis features are generally of secondary importance to the data collected here.

The term *research profile services* refers to information systems that use linked data concepts to pool portfolios from institutional and publicly available data sources. These approaches do not focus on supporting process-oriented administrative research reporting, they add value by linking and processing research information aimed at multiple institutions. The pioneer in this field is the Open Source software VIVO, which provides harvesting tools to aggregate research information from various online sources in a standard manner.

*Integrated research information systems* are combined database and reporting systems that enable an institution to comprehensively document, evaluate and continue developing their research activities. Integrated research information systems focus on setting up and building up a quality-assured body of metadata consisting of as much externally and internally processed information as possible that is managed collaboratively. Value-added services whereby institutions communicate externally via research portals and web services for institutional websites can also be connected. Integrated research information systems have the following characteristics:

- The information objects and their chronological order are described in a data model.
- Information from different data sources within an organisation and from external sources are pooled and semantically enriched.
- The IT solutions use a roles and rights concept to support distributed data management and quality assurance beyond contextual, hierarchical and organisational levels.
- The systems enable multiple output and analyses features, while also enabling data to be reused.

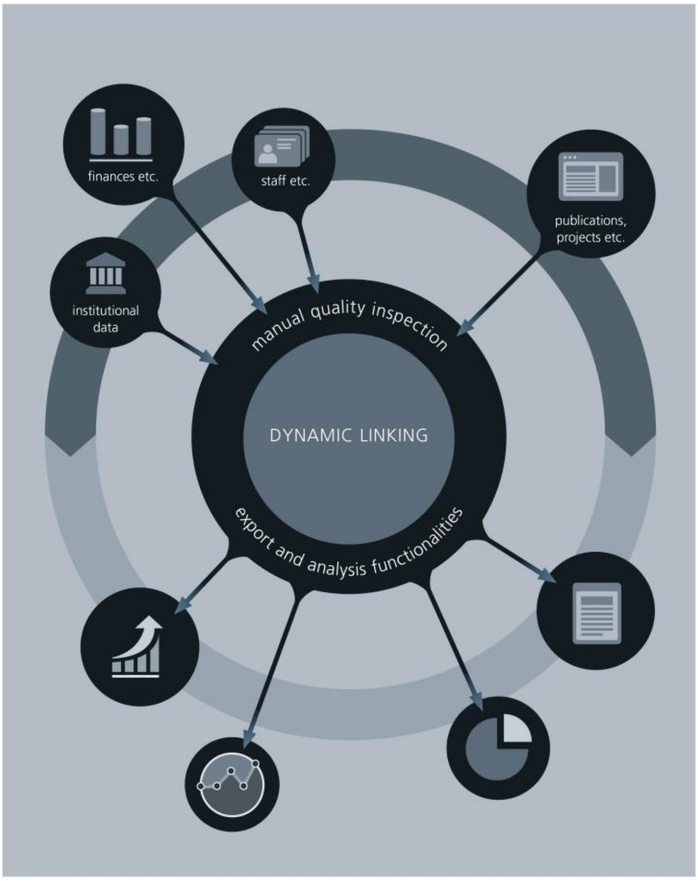


Figure 5: Diagram of an integrated information system that dynamically links various source systems and data pools while also offering multiple output and analysis features.

## Rollout and operation

The rollout of a research information system is tantamount to an organisational development project which may involve a lot of work and considerable investment. The number and diversity of stakeholders in science and administration whose needs have to be addressed may lead to complex communication processes.[27] Some main guidelines for introducing and managing a research information system have been collected from institutions sharing their experiences with one another and are described below.

**Investment and choice of system**

The scope and aim of the information service and/or research information system depend on an institution's individual cost-benefit considerations. The investment in an integrated research information system or simple cataloguing system needs to be weighed up against existing research documentation management processes as well as the anticipated boost in quality, e.g.

- How the availability and quality of information collected and stored by previous systems is rated (project and third-party-funding databases, university bibliographies, etc.);
- Which resources are used to manage current data sources;
- What cost will this incur for existing IT systems;
- How much work is anticipated in order to gain the consent of everyone involved.

A research information system needs to be linked to administrative information in order to be used sustainably for research reporting purposes. This is why integration of organisational, personnel and financial information is a basic requirement. That way, depending on the resources and local pressure to come up with solutions, individual systems such as project and patent databases and their accompanying data management processes can be concurrently linked. Such a strategy can also help to manage the number and diversity of requirements and processes while also increasing the chances of achieving success.

As things stand, there is no out-of-the-box solution with only a few integrated research information systems currently available on the market.[28] Some commercial products are aimed at international standards such as CERIF (cf. glossary) and provide established interfaces, analysis features and implementation process features, but their implementation and adaptation to the specific needs of a university or research institution may involve a lot of work and considerable cost.

Internal developments from research institutions generally involve research portals and project databases.[29] Internal development and, given the complex integration and value-added-service needs, maintenance are major challenges that only locations with well-equipped data centres or collective developer groups can handle.

[27] The example involving the University of Hamburg illustrates just how complex a basic needs assessment can be (cf. Stiehl, 2011): The results of the needs assessment carried out in 2011 initially returned 1,200 needs which were later boiled down to 212 split into 18 different categories.

[28] Examples (by no means exhaustive): Pure (Elsevier), Converis (Thomson Reuters), FactScience (QLEO), Symplectic Elements (Symplectic).

[29] cf. Sticht, 2015

Alongside commercial research information systems, some ground is being gained by IT developments in the Open Access community. Initially designed as open source publication repositories, additional information objects are currently being added and a transformation into research information systems can be observed. Such Open Source systems are only just starting to be built.[30]

Other open source solutions are research profile services that use Linked Open Data (LOD) technologies. The main example of this is the VIVO software[31] which collects research information freely available online and processes it into new research profiles, e.g. for an institution or discipline. The data can be processed and visualised. The added value of this system is that users can find and use information across institutions.

**Project management**

The rollout of an integrated research information system means linking data silos, setting up multidisciplinary data management processes, establishing data quality assurance processes, analysing reporting requirements and standardising assessments. Basic project management requirements are similar to other types of IT projects and organisational development processes:

- The project management design should reflect the size of the institution and the scope of the goals.
- The project management team should be competent in terms of the required process analyses and also have good social and communication skills in order to be able to effectively introduce process discipline and reach compromises.[32]
- The project should have enough leverage to encourage the people involved within the organisation to open their systems and adjust to the new processes while also providing the necessary standing during crisis situations (e.g. by appointing a prorector of research or Chief Information Officer).
- The project management team should include people involved in areas related to the interfaces, in particular those who support existing database systems, e.g. the library as operator of a centralised document server and/or university bibliography.
- Data protection and privacy officers, IT security and staff committees should be involved from an early stage.

**Operating concept**

No aspect of introducing a research information system makes organisational development as clear as the stipulation of an operating concept. The operating concept ensures mid- and long-term use and, with it, the sustainability needed to operate the application in a smooth and stable manner. Several people at the research institution should generally be made responsible for this.

[30] cf. the cooperation project involving the Forschungszentrum Jülich (JuSER), DESY (pubDB), GSI (GSI Repository), RWTH, MLZ (iMPULSE) based on Invenio: http://invenio-software.org/. Italian consortium CINECA developed a research information system called IRIS based on DSpace: http://www.cineca.it/it/content/IRIS and is involved in the development of the new DSpace-CRIS module (http://cineca.github.io/dspace-cris/index.html, last visited on 27/10/2014). Extensions to EPrints for research information: http://bazaar.eprints.org/154/, last visited on 16/10/2014.

[31] http://www.vivo.org, cf. glossary

[32] cf. specification of a skills profile for process managers in Groening & Schade (2011), p. 32 et seq.

User support ensures user involvement. Application support, operation and management focus on supporting both software and hardware. Acceptance of the research information system is bound to be enhanced by good constant on-site support. Training courses will boost user skills, which will then positively impact on high data availability. Decisions as to whether hardware and software support as well as comprehensive system skills are handled and built up externally or within the organisation are, however, more of a strategic nature.

Organisation of the constant flow of current data also plays a key role, which is why there needs to be clear organisational and technical requirements along with documentation of the processes that collect data relevant to the research information system. These need to be designed such that they permit the maximum possible quality of data collection without additionally burdening researchers, at least not permanently or unnecessarily. Processes and structures should be designed in a service-oriented manner and closely linked to science.

**Acceptance of reporting obligations**

Researchers may worry that data built up collaboratively in the institution's research information system may be unduly used for assessments and evaluations. Data protection and privacy are often brought as counterarguments if there is deemed to be a threat to transparency, e.g. if goals and use purposes of a research information system cannot be clearly conveyed.

Resistance to the introduction of research information systems should be adressed by a transparent approach to the underlying reporting processes and specialist evaluations. In the end the institution has to deal with the data processing requirements on a comprehensive scale. Referring to a reference model such as the German research core dataset (cf. glossary) may help in this regard. Research reporting guidelines can be drafted along with a framework for data use. Trust can be built and fostered if the data are agreed upon (quality assured) and their purposes transparent.

In order to generate acceptance, clear added value must be conveyed, e.g. a research database has to provide relatively swift initial benefits for use cases during everyday research work at institutions or consortiums with their scientists and administration staff (value-added services could include the reuse of data on websites). Ongoing quality assurance is required to make sure that scientists' needs are met in terms of adequate database-generated content.

## Creating added value

Investing in a research information system comes with high expectations. Some specific technical and organisational aspects can lead to added value for both researchers and management. The recommendations below are modelled on an integrated information management system in which science and administration pool their information within the scope of a comprehensive research reporting rights and roles concept in line with data protection and privacy law.

**Availability (Data are comprehensively available)**

The database forms an institutional "body" of research information that is available as a mutual resource for various research reporting purposes. Information about equipment and achievements is no longer spread across different offices and departments. Instead a roles and rights management system governs availability at offices and departments that process research information (distributed access rather than distributed data storage). Information is collected starting with individual objects, e.g. projects, publications, expert profiles, dissertations. Output and analysis functions then aggregate the data as and when necessary while also enabling a standardised display for reports.

Implementation of the “availability” requirement can be supported by the following measures:

- The data model should be shaped in sufficient detail so that, e.g. third-party-funded projects also contain financial details for allocating funds based on achievement, project details provide an overview of cooperation partners and publication data are linked to co-authorships. Standards such as CERIF (cf. glossary) help to manage and minimise the amount of work required to stipulate the individual attributes.
- The database should collect data from individual scientists.
- The information objects such as people, projects, organisations and publications should be linked to one another with a temporal reference. This means that data can be selected and aggregated based on specific needs, e.g. by organisational unit.
- Availability for internal and external purposes should be governed based on different visibility levels, e.g. “public”, “campus”, “personal access”.
- Research information should be protected by a roles and rights management system applied to data processing and output and analysis features.
- Centralised collection and distributed access should increase availability, but may pose particular challenges in terms of complying with data protection and privacy law. The institutions responsible for such issues should therefore be included at an early stage.

**Reliability (Data are reliable)**
A lot of people are involved in setting up and building up a body of information for research reporting purposes. The research information system pools existing information from science and administration, thus increasing data reliability. The system has to be integrated into the existing IT landscape and should import organisational master data in as automatic a fashion as possible. Internal user groups tend to overestimate how important data quality is to such a system in terms of its success, yet they underestimate how their involvement is contingent to the system’s success. This means that uniform data quality requirements and standards should be agreed upon by those involved. The integration of other internal systems gives rise to requirements that have to be met, some of which will represent new additional tasks for the people in charge. Potential sources of error need to be identified and integrated into overarching validation processes. Comprehensive quality assurance processes for which staff are required should be put into place to ensure compliance with standards. IT-based workflows may help to make work easier.

Implementation of the “reliability” requirement can be supported by the following measures:

- Master data from the organisation’s management systems should be regularly updated and linked to the collaboratively entered research information such as publications, projects and activity.

- Uniform data management quality standards and basic database quality assurance principles involving those responsible for linked systems and the research information system users should be stipulated.[33]

- Feedback about errors in the database should be agreed on for evaluations and reporting purposes. Corrections have to be made in the source, i.e. the research information system itself or the database that supplied the data.

- Quality assurance measures should be collaboratively planned and continuously implemented. Data controllers should be appointed to do this. With publications, the library can take on quality assurance duties for bibliographical data.

- Technical validation workflows should be implemented to allow collected data to be checked. This is particularly advisable if the data sources used only provide parts of the required information, e.g. third-party-funded projects, or if sources of error are known (allocation problems during publication imports).

- The status of research information collected collaboratively and at various stages of completion should be immediately identifiable, e.g. "being checked", "checked".

- Identity management and connecting databases with one another should help to avoid duplicates.

**Topicality and consistency (Data are up to date and consistent)**
As and when needed, research information systems can update and output data on certain reporting dates. Such data should ideally be up to date, although this is very difficult to achieve in real-life scenarios. More realistic is the varied use of the research information system together with a combination of different added value and routines that are as close to the desired data topicality as possible. Data management processes that ensure that data is routinely up to date should be distinguished from those to which special reporting requirements apply and for which the information can be manually entered and, in some cases, only entered on certain reporting dates. (Selective) storage and provision of historical data is required, also in order to be prepared for retrospectives.

Implementation of the “topicality” requirement can be supported by the following measures:

- Data management processes should be primarily optimised for use with up-to-date areas such as general research portals.

- Additional relevant information should be entered and checked on certain reporting dates.

- As many information needs as possible should be directly handled by the research information system. Use cases that (still) apply independently within an institution will be concurrently integrated into research information system management. In the long term, this gives rise to a process where changing parts of the database are updated as and when needed, which helps to ensure that all the data is up to date.

[33] The “Principles of Good Data Management” developed for British universities are a useful guide, as referred to in Clements and Reddy, cf. 69 et seq.

- Added value should be created for scientists such that they can use the research information system's data for their individual needs. This includes, for example, adding publication and project lists to personal or institutional websites, CV management features as well as indexing and dissemination of publications for relevant publication services.[34]

- As part of an institution's internal data storage and provision policy, measures for dating entries should be taken in order to provide information about completed projects and former staff. Retrospectives should cover a period of at least 5 to 10 years so they can be used for evaluations.

**Efficiency (Collection is efficient)**

The benefit of integrated information systems is the fact that data collection processes can be (partially) automated. Existing information from external and internal sources are used and set in relation to one another. By collaboratively pooling, adding to or updating data as and when needed, similar information can be provided for different purposes in a more efficient and organised way with an IT solution often supporting this on a day-to-day level. Multi-level input collection processes that build on one another can help to reduce discrepancies in terms of the database and multiple entries. Each research institution needs to decide for itself how much of a researcher's work it should take on and to what extent it should include centralised services such as the library in the planning and implementation processes. The more detailed a research institution's reporting requirements, the higher the need for researcher collaboration as in some areas they are the only people with the necessary skills.

Implementation of the "efficiency" requirement can be supported by the following measures:

- By connecting it to centralised master data and integrating several services, the research information system should be suitable for use as a key tool by people and institutions involved in science and administration.

- Data should be collected on several different levels. The input processes should ease the burden on individual researchers and allow entries to be added to. This preparatory work performed by the scientists themselves and their expertise are crucial in terms of quality assurance (cf. "availability").

- External systems should be integrated as data sources. Automated data request routines continuously add to the database.[35]

    - Working automated mapping of datasets from different sources are a major factor in achieving enhanced efficiency, e.g. publication types for publications. Unfortunately interfaces to external databases contain a lot of sources of error and generally cause a lot of correction work.

    - To this end, consistent duplicate checks and process-related data management and mapping to personal and organisational master data are imperative.

[34] e.g. current practice with Open Access repositories (OAI-PMH, cf. glossary)

[35] In terms of publications this could be, e.g. a connection to large bibliographical databases such as PubMed, Web of Science or Scopus.

- The user interface should be modern and ergonomic while also enabling users to quickly navigate to every level of data collection, processing and evaluation.

- Additional added value can be achieved by integrating processes already in place between science and administration, e.g. the option to process publication fees from institutional publication funds from within the system or the option to support invention disclosures and notifications of third-party funding.

**Sustainability (Collection is sustainable)**
A research information system cannot be operated in the long term without clear commitment from the research institution's management team. This includes the willingness to deploy resources needed to maintain operation and to appoint people responsible for such tasks. Key factors for long-term success include integrating related documentary processes and structurally interlinking other centralised management processes to as large an extent as possible. The sustainability of a research information system also depends on the use of national and international standards to ensure long-term data interoperability and strategically enable data for use in national and international scientific applications. In the future, this will also include features that support Semantic Web applications (Linked Open Data).

Implementation of the "sustainability" requirement is more likely to occur by meeting as many of the following requirements as possible:

- The institution should agree on an internal operating concept that also comprises long-term financing for the system and the resources required to achieve this.

- "Personal" prestige projects are likely to fail if the figureheads are replaced during or after their introduction.

- Observance of international standards such as CERIF makes research documentation flexible in the long term while also simplifying a potential change of IT system. This also promotes competition among standard software providers.

- The research institution's processes should be connected in as integrated a manner as possible.

- Individual local data model solutions should be avoided; instead there should be compromises that are compatible with standards. The amount of work involved to achieve this should always be taken into consideration.

- Standard identifiers such as the Open Researcher and Contributor ID (ORCID) can be used to simplify the import and export of research information.

- Research information should be served according to open standards and provide connectivity rather than just being managed in closed research portals. Data should be provided based on the Linked Open Data paradigm to make it easier to disseminate and exchange research information all over the world.

- Agreement on the conditions under which publicly available research information can be reused should serve as the basis for this. Open licences should be used to facilitate open access.[36] The research information system also needs to support Linked Open Data formats from a technical perspective.

- The option to automatically allocate data to various profiles should enable connectable databases to be created. Data profile examples include the German research core dataset, CASRAI profiles and OpenAIRE database.
- As of version 1.4, the CERIF standard enables several such data profiles to be added and managed in a structure very similar to that of the Semantic Web (semantic layer concept) while also displaying temporal relations between information objects and standardised formalism.

Electronic data transfer and the linking of local and national research databases will become increasingly important in the future, which in turn will drive the standardisation of information systems. Any questions that arise in connection with the quality assurance of imported publication or project data, or with feedback mechanisms in the event of corrections or error messages need to be discussed and solutions found.

**Security (Data are secure)**
A research information system offers a comprehensive summary of public and internal data that are largely of a personal nature. A research information system therefore needs to meet data ownership requirements. Researchers can access their data and are involved in collection processes. The data to be made available in public research portals or only for internal reporting purposes can be controlled. The intended purpose of data collection is adequately specified. The IT environment is protected against spying, monitoring and manipulation by third parties.

Implementation of the “security” requirement can be supported by the following measures:

- A rights and roles concept in line with data protection and privacy law should govern access to a research information system’s data from a technical perspective.
- Researchers should be granted the power of disposition and can authorise actions applicable to data. This applies both in terms of external visibility and in general when it comes to approvals for certain purposes.
- If quality assuring instances are involved in the processes, e.g. libraries in the event of publications, researchers should be able to determine who subsequently processed the data and can arrange for changes to be made to the dataset.
- The system shows the current stage of data collection and processing.
- A procedures directory, service agreement and, where necessary, employment regulations should help to create transparency in terms of intended purpose, rights and duties.

[36] cf. http://opendefinition.org/od/

## Collaboration and other exchanges of experience

IT-based research reporting is still in its infancy. In view of the high level of suffering reported on a regular basis, as a university or research institution it makes sense to focus on feasible local solutions. However, many of the value-added services and potential ways to ease the burden of managing a research information system can only be developed collaboratively.

The following action areas require a collaborative process among institutions:

- Clarification of legal issues
- Agreement on reference models and essential features for research information systems, e.g. interfaces and exchange formats, be they in the form of file formats for importing and exporting data or in the form of interfaces
- Agreement on the use of identifiers

Continued exchanges of experience between the pioneers in this field will help to drive the options outlined here and contribute to the development of common standards. DINI AG FIS will continue to exchange with universities, research institutions and other involved parties.

# Authors and advisors

The following people were involved in this position paper as authors (in alphabetical order):

- Daniel Beucke, *Niedersächsische Staats- und Universitätsbibliothek Göttingen* (Göttingen State and University Library)
- Andreas Bliemeister, *Leibniz Gemeinschaft* (Leibniz Association)
- Barbara Ebert[37,38], *Leuphana Universität Lüneburg* (Leuphana University of Lüneburg)
- Eiken Friedrichsen, *Leuphana Universität Lüneburg* (Leuphana University of Lüneburg)
- Lambert Heller, *Technische Informationsbibliothek Hannover* (German National Library of Science and Technology), Open Science Lab[39]
- Sebastian Herwig[38], *Westfälische Wilhelms-Universität Münster* (University of Münster)
- Najko Jahn[38], *Universität Bielefeld* (Bielefeld University)
- Matthias Kreysing[38], *Universität Göttingen / Universität Hildesheim* (University of Göttingen / University of Hildesheim)
- Daniel Müller, *Fernuniversität Hagen / Universität Göttingen* (University of Hagen / Göttingen)

---

37 Barbara Ebert was a member of the euroCRIS Executive Board at the time of writing this paper

38 Barbara Ebert, Sebastian Herwig, Najko Jahn and Matthias Kreysing are members of the German research core dataset project groups.

39 Together with other partners, the Open Science Lab connects European institutions interested in using VIVO.

- Mathias Riechert[40], *Institut für Forschungsinformation und Qualitätssicherung* (Institute for Research Information and Quality Assurance)

- Regine Tobias, *Karlsruher Institut für Technologie* (Karlsruhe Institute of Technology - KIT)

We would also like to thank the following people for their advice, critical reviews and discussions (also in alphabetical order):

- Katja Barth, *Universität Kiel* (Kiel University)

- Bettina Brand, *Forschungszentrum Borstel* (Research Center Borstel)

- Prof. Dr. Andreas Degkwitz, *Universitätsbibliothek der Humboldt-Universität zu Berlin* (Humboldt University Library Berlin)

- Karin Eckert, *Universität Mainz* (University of Mainz)

- Josef Hüvelmeyer, *TU Dortmund* (Technical University of Dortmund)

- Brigitte Jörg, JeiBee Ltd. (until 06/2014)

- Prof. Dr. Gudrun Oevel, *Zentrum für Informations- und Medientechnologien (IMT) der Universität Paderborn* (University of Paderborn Centre for Information and Media Technologies - IMT)

- Heinz Pampel, *Helmholtz Open Science Koordinationsbüro* (Helmholtz Open Science Coordination Office)

- Markus Putnings, *Universitätsbibliothek Erlangen-Nürnberg* (University Library of Erlangen-Nürnberg)

- Birgit Schmidt, *Niedersächsische Staats- und Universitätsbibliothek Göttingen* (Göttingen State and University Library)

- Frank Scholze, *Karlsruher Institut für Technologie* (Karlsruhe Institute of Technology - KIT)

- Hans-Siegfried Stiehl, *Universität Hamburg* (University of Hamburg)

- Holger Wendler, *Wuppertal Institut für Klima, Umwelt, Energie GmbH* (Wuppertal Institute for Climate, Environment and Energy)

[40] Mathias Riechert is employee in the German research core dataset project

# Annex

## Glossary and list of links

Below is an overview of the terms and initiatives covered in this paper. No warranty is made for the completeness of this list of databases and initiatives as it acts solely as an additional guide to readers of this position paper.

**DINI Working Group on Research Information Systems (DINI AG FIS)**
Established in 2012, DINI AG FIS works on documenting good practice of introducing and managing research information systems. Areas covered by DINI AG FIS include networking data management processes, handling personal data in research reporting, exchange formats for research information and acceptance of research information systems.

http://dini.de/english/

**CASRAI**
CASRAI stands for Consortia Advancing Standards in Research Administration Information and is an international standardisation initiative established in Canada with the aim of developing definitions and data profiles for research documentation.

http://casrai.org/

**CERIF**
CERIF stands for Common European Research Information Format. From a technical perspective, CERIF is a data model and metadata format for scientific information objects that defines organisations, people, infrastructures, projects and funding as well as results such as patents, publications and prizes while permitting temporal and contextual relations. Although powerful, CERIF uptake has been slow in the past, probably due to its level of complexity. In a number of applications, only part of the model or older versions are used. The information objects, links and vocabulary are selected by precisely defining the context (form follows function). These steps need to be simplified further for programmers and developers in order to encourage CERIF uptake. As of version 1.4, several data profiles can be added and managed in a structure very similar to that of the Semantic Web (semantic layer concept) while also displaying temporal relations and standardised formalism. Data profile examples include the German research core dataset, CASRAI definitions and OpenAIRE European research database profiles. CERIF is developed and disseminated by the non-profit organisation euroCRIS, and its members include research institutions, research sponsors, libraries and associations.

http://eurocris.org

CERIF and Linked Data

https://code.google.com/p/cerif-linked-data/wiki/InstructionsLDfromCERIF

CERIF XML for OpenAIRE

https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS

**CRIS**
CRIS stands for Current Research Information System.

**DSpace**
The Open Source software DSpace was developed by the Massachusetts Institute of Technology and HP Labs in 2002. It is used to set up institutional Open Access repositories and is developed and managed collaboratively.

http://www.dspace.org/

DSpace-CRIS module

http://cineca.github.io/dspace-cris/index.html

**EPrints**
The Open Source software EPrints has been in development by the University of Southampton since 2000. It can be used to set up and build up institutional Open Access repositories.

http://www.eprints.org

Extensions to EPrints for research information

http://bazaar.eprints.org/154/

**FIS**
FIS is a German abbreviation of *Forschungsinformationssystem* (research information system)

**Research profile services**
Research profile services are Semantic Web applications. These approaches do not focus on supporting process-oriented administrative research reporting, they add value by linking and processing research information aimed at multiple institutions. The pioneer in this field is the Open Source software VIVO, which provides harvesting tools to aggregate research information from various online sources in a standard manner.

**Research core dataset (Kerndatensatz**
The German "Research core dataset" project is working to define a minimum set of research details that institutions should routinely be able to provide for reporting purposes. Since August 2013, the specifications are being worked out as part of a project funded by the Federal Ministry of Education and Research (BMBF). First results are due to be published in mid-2015. Universities and non-university research institutions are involved as pilot institutions.

http://www.forschungsinfo.de/kerndatensatz/en/index.php?home

**Linked Open Data**
Linked Open Data are part of the Semantic Web and lead to a global network. The data are identified by Uniform Resource Identifiers (URIs) and can be accessed directly via HTTP while also using URI to refer to other data.

http://en.wikipedia.org/wiki/Linked_open_data

**OAI-PMH**
OAI-PMH (OAI Protocol for Metadata Harvesting) is a standardised interface format developed by the Open Archives Initiative. There are several thousand OAI-PMH providers worldwide as well as national research hubs that provide publication metadata.

http://www.openarchives.org/

**ORCID**
ORCID stands for Open Researcher and Contributor ID and is an international non-profit initiative that assigns identifiers to authors. Members and users of ORCID include research institutions, publishing houses and information service providers.

http://orcid.org/content/about-orcid

**OpenAIRE**
OpenAIRE stands for Open Access Infrastructure for Research in Europe. European Union funding is used to establish an open access infrastructure for publications with contextual information about funding and primary data and any related services. The aim is to visualise the results of European research funding.

OpenAIRE has interfaces to connect institutional systems. Research information systems can send data to OpenAIRE via a CERIF XML interface.

https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_CRIS

https://www.openaire.eu/

**Semantic Web**
The Semantic Web is a concept aimed at further developing the World Wide Web. Online information should be given a clear description of its meaning (semantics) that can also be understood or at least processed by computers. This enables information to be automatically served to interested users when they perform a search query.

The Semantic Web standard is used as a basis for structuring information in RDF (Resource Description Framework) triples format which enables information to be aggregated and used in database applications on the public web.

http://en.wikipedia.org/wiki/Semantic_Web

**VIVO**
The Open Source software VIVO is a Semantic Web application that links data such that it can be read and linked by machines and reused online. VIVO provides profiles that map relations between researchers, e.g. co-authorships or institution affiliations, which in turn allows multiple institutions and disciplines to be browsed (cf. research profile service). From a technical perspective, VIVO is based on the Linked Open Data paradigm and ontologies used in this field such as FOAF. The data can be processed and visualised. Institutions can install VIVO or provide VIVO-compatible data for Semantic Web applications. In an institutional setting, VIVO complements other management systems, but is not designed for collecting and processing (confidential) administrative information.

VIVO was established in the US with funding from, e.g. the National Science Foundation (NSF) and the National Institute of Health (NIH). Today it is collaboratively developed under the auspices of the DuraSpace Community (-> DSpace).

Short tour: VIVO in an information ecosystem

http://vivoweb.org/about

VIVO in Germany

http://blogs.tib.eu/wp/opensciencelab/vivo-fuer-scientific-communities/

# Literature

Below is a list of publications used in AG-FIS' work and which were directly or indirectly used in this paper.

## Recommendations and publications


- DFG (2010): "Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme." Empfehlungen der Kommission für IT-Infrastruktur für 2011–2015 (online: http://www.dfg.de/download/pdf/foerderung/programme/wgi/empfehlungen_kfr_2011_2015.pdf, last viewed on 8/8/2013)

- GWK (2011) Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland (online: http://www.gwk-bonn.de/fileadmin/Papers/KII_Gesamtkonzept.pdf, last viewed on 28/10/2014)

- Wissenschaftsrat (2011): Empfehlungen zur Bewertung und Steuerung von Forschungsleistungen (Drucksache 1656-11). Halle, 21/11/2011 (online: http://www.wissenschaftsrat.de/download/archiv/1656-11.pdf, last viewed on 29/8/2013)

- Wissenschaftsrat (2013): Empfehlungen zu einem Kerndatensatz Forschung (Drucksache 2855-13). Berlin, 25/01/2013 (online: http://www.wissenschaftsrat.de/download/archiv/2855-13.pdf, last viewed on 6/8/2013)


## Books and compilations


- Bittner S, Hornbostel S, Scholze F (ed.), Forschungsinformation in Deutschland: Anforderungen, Stand und Nutzen existierender Forschungsinformationssysteme - Workshop Forschungsinformationssysteme 2011, iFQ-Working Paper Nr. 10, p. 7.

- Degkwitz, A.; F. Klapper (ed.): Prozessorientierte Hochschule. Allgemeine Aspekte und Praxisbeispiele. Bad Honnef: Bock + Herchen, 2011, 217 S.
  Online: http://www.dini.de/fileadmin/docs/Prozessorientierte_Hochschule_2011.pdf (last viewed on 28/8/2013)

- Data Science Journal, Special Issue 2010: CRIS for European e-Infrastructure. Online: http://www.codata.org/dsj/special-cris.html (last viewed on 21/8/2013)

## Other literature references

- Clements, Anna & Reddy, Helen (2010): How a CRIS can drive improvements in information management. In: Stempfhuber, Maximilian & Thidemann, Nils (ed.): Connecting Science with Society. The Role of Research Information Systems in a Knowledge-Based Society (Proceedings of the 10th International Conference on Research Information Systems, Aalborg, Denmark, 2-5 June 2010), Aalborg University Press, p. 65-72. Online:

  http://www.eurocris.org/Uploads/Web%20pages/cris2010_papers/Papers/cris2010_Clements.pdf (last viewed on 27/8/2013).

- Groening, Yvonne & Schade, Ann Katrin (2011): Die Herausforderungen des Prozessmanagements an Hochschulen. In: Degkwitz, Andreas & Klapper, Frank (ed.): Prozessorientierte Hochschule. Allgmeine Aspekte und Praxisbeispiele. Bad Honnef: Bock + Herchen, p. 23-38.

- Sticht, Kendra: Untersuchung zum Einsatz von Forschungsinformationssystemen an Hochschulen in Deutschland. Ergebnisse der Online-Befragung, unveröffentlichte Masterarbeit, Berlin 2014

- Stiehl (2011): Anforderungen an ein Forschungsinformationssystem am Beispiel der Universität Hamburg. DINI ifQ Workshop "Forschungsinformationssysteme in Deutschland" held on 22/11/2011. The slides are available online:

  http://www.dini.de/fileadmin/workshops/forschungsinformationssysteme/Stiehl_Hans-Siegfried_DINIiFQ_WS_KIT_V2_3.pdf

- Stocker, A., Tochtermann, K & Scheir, P. (2010): Die Wertschöpfungskette der Daten. HMD Praxis der Wirtschaftsinformatik, 47 (5), p. 94-104. http://dx.doi.org/10.1007/BF03340517