



Establishing a Research Information System as Part of an Integrated Approach to Information Management: Best Practice at the Karlsruhe Institute of Technology (KIT)

Frank Scholze

Karlsruhe Institute of Technology (KIT), KIT Library,
Karlsruhe, Germany,
scholze@kit.edu

Jan Maier

Avedas, Karlsruhe, Germany,
j.maier@avedas.com

Abstract

The Karlsruhe Institute of Technology (KIT) is one of the largest research and higher education organisations in the world focusing on engineering and natural sciences. At present KIT, under the chairmanship of its executive board, is installing an extensive current research information system (CRIS) covering all institutes and facilities of the organisation.

The assumption underlying the project is that a consistent overview of research performance has become fundamental for the international competitiveness of research institutions and is increasingly important for strategic decisions at the executive level. Ultimately, it also leads to better data and control in rankings at higher education assessments.

The new research information system systematically maps all of KIT's processes and instruments to obtain, connect, present and utilise the research metadata of active researchers. This reduces the documentation workload for researchers, for the executive level and central units such as the library, and at the same time allows for and facilitates an overall view and the aggregation and visualisation of research metadata.

Our vision is to build a federally structured network of systems that gathers information on KIT's publications, research competence, research projects, patents and technological offers by retrieving data from external and internal sources as well as directly from the researchers. The network facilitates linking and aggregating of data and provides unique identifiers for individual researchers and organizational units. With its consistent data model the research information system also fosters the organisational development of KIT, which was formed in 2009 by the merger of a university and a national research centre.

The researchers and their activities are at the core of the research information system. The system substantially reduces their administrative burden in documenting project information and publications. Automatic data import from external online sources and repositories has the big advantage that publication data are acquired and validated only once and can be used subsequently for manifold purposes such as websites, CVs, publication lists and reference management software.

Due to the complexity of workflows, internal and external sources and the organisational infrastructure that is required for such an implementation, KIT has chosen a commercial software solution (by AVEDAS AG) as the basic installation, and in cooperation with AVEDAS will expand it and adjust it to a tailor-made system that covers the entire KIT publication and research life cycle. The CRIS at KIT is part of a larger integrated and service-oriented approach to information management (Karlsruhe Integrated Information Management — KIM).

Key Words: research; information system; information management

The Karlsruhe Institute of Technology (KIT)

The Karlsruhe Institute of Technology (KIT) is a German academic research and education institution resulting from a merger between the University (Universität Karlsruhe) and the Research Centre (Forschungszentrum) Karlsruhe. The university, also known as Fridericiana, was founded in 1825. In 2009 it merged with the former national nuclear research centre founded in 1956 as the Kernforschungszentrum Karlsruhe (FZK). KIT is a unique initiative as it combines the mission of a University funded by the State of Baden-Württemberg with the mission of a national research centre funded by the Federal Government. A special law had to be passed to make this happen. From a historical perspective, however, it was a logical step, as both institutions look back on a long tradition of research and education with many parallels in terms of content and organisation. KIT is one of nine German Excellence Universities and positions itself as an

institution with excellent research and teaching in natural and engineering sciences on an international scale, and with scientific excellence as a global cutting-edge university in research, teaching and innovation. As a member of the Helmholtz Association, the largest science organisation in Germany, KIT makes major contributions to top national and international research.

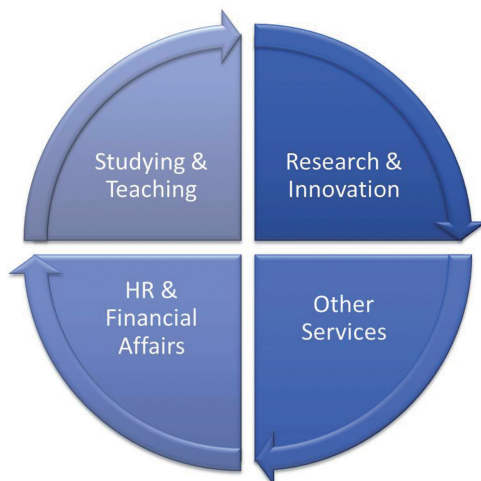
Karlsruhe Integrated Information Management (KIM)

The concept of 'Karlsruhe Integrated Information Management (KIM)' pursues the goal of increasing excellence in research and teaching at the Karlsruhe Institute of Technology (Hartenstein *et al.*, 2007). Therefore, it strives for a continuous and sound integration of relevant legacy systems and data as well as for increasing the accessibility and transparency of related business processes. The KIM project's technical realisation is based on a service-oriented Architecture (SOA) model which was termed 'KIM integrated Service Oriented Architecture (KIM iSOA)' (Freudenstein *et al.*, 2006).

KIM consists of a series of associated sub-projects (Figure 1). Among them are 'Management of courses, examinations and study assistance (KIM-CM)' and 'Identity management (KIM-IDM)'. The results of these projects have led to services which are accessible through the portals 'studium.kit.edu' and 'intra.kit.edu'. Furthermore, provisioning all KIT employees and students on the basis of their associated KIT accounts not only facilitates innovative IT services, but contributes substantially to giving people in the university and in the large-scale research sector a sense of solidarity.

In order to co-ordinate information provision and processing, and to ensure that organisational and cultural change associated with the KIM projects is managed properly, co-ordination structures have been established at KIT. Among these are the position of a 'Chief Information Officer (CIO)' as a member of the executive board in order to promote the integration and co-ordination of information provision and processing in all fields of the KIT at the top management level; and the 'Committee on Information Provision and Processing (IVA)', which was established by the senate and the executive board chaired by the CIO and gives recommendations to the president's office and the senate on all questions concerning the processing or provision of information at KIT, especially in matters of general importance. There are two sub-committees established by the IVA, one for information provision (A-IVS) and one for information processing

Fig. 1: Project Portfolio within Karlsruhe Integrated Information Management (KIM).



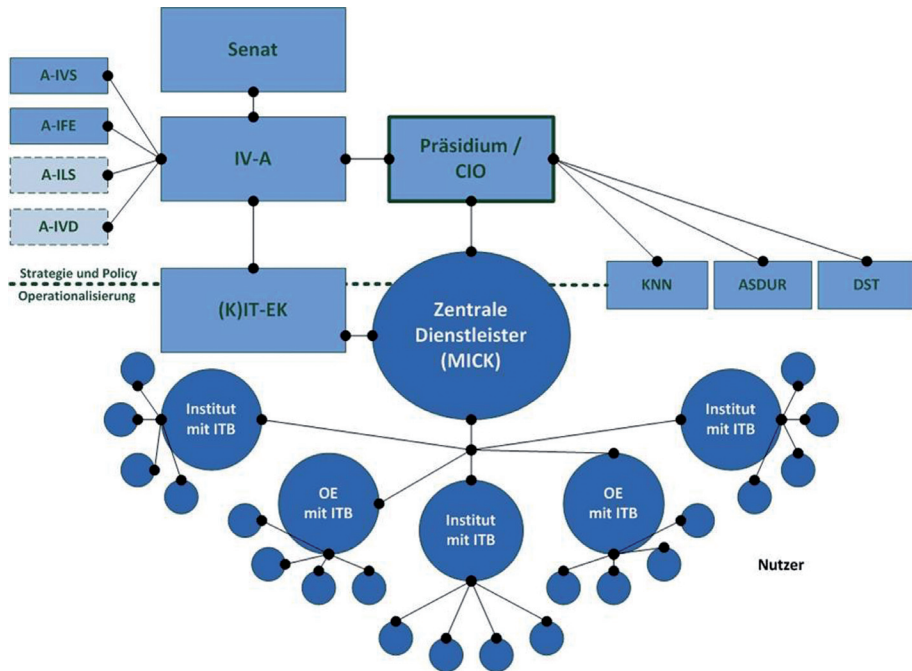
(A-IFE). Two more will follow: one dealing with the ‘application of new media in research, teaching, study, further education and administration (A-ILS)’ and one for ‘information processing in administration (A-IVD)’ (Figure 2).

Guidelines for these sub-committees form user governance: the sub-committees represent the interests of users in the respective fields, advise the operational officers and give appropriate recommendations for strategic development. Integration with the IVA is guaranteed by the fact that the respective chairmen of the subcommittees have to be members of the IVA. At the operational level, all relevant organisational units are represented in the ‘Media and Information Service Centre Karlsruhe (MICK)’. MICK is organised co-operatively by the Steinbuch Centre for Computing, KIT Library and the administrative units dealing with information provision and processing. It pools all competencies and services in this domain. On the other hand, it does not break up or absorb the computing centre, library or administration as distinctive organisational units, as they have unique tasks and missions which do not overlap.

CRIS and Repositories

CRIS (Current Research Information Systems) have existed since the 1960s and have been designed to manage information related to research. This

Fig. 2: Governance within Karlsruhe Integrated Information Management (KIM).



includes input and output data like publications, patents, products, prices as well as projects, equipment and facilities. At the European level, there exists a standardised data model called CERIF (Common European Research Information Format) which is maintained and developed by euroCRIS.¹

Repositories were developed in the 1990s as part of the movement to secure open access to scholarly publications. Therefore they concentrated on providing free and unrestricted access to *publications* and not to the contextual data resulting from the research processes which led to these publications. As discussions and developments have matured, repositories have become much more aware of the importance of contextual information. The discussion now focuses on publication management, virtual research environments and research data.² Critics say this is due to the fact that, despite numerous initiatives and incentives,³ repositories are not as well filled with full-text publications as promoters of the idea had hoped they would be.⁴

After initial discussions within the Knowledge Exchange (Razum *et al.*, 2007), there is now a wider uptake of the question as to how and to what purpose CRIS and repositories should integrate or interact. Three main lines of discourse can be distinguished. Some think that a repository will develop into a CRIS (Carr, 2010), adding more contextual and standardised meta-data about persons, projects, funders etc. The majority within both communities think that a parallel development will continue, and consequently there has to be some kind of interface between the systems depending on local requirements — be it one or bi-directional (Sheppard, 2010).⁵ As a third perspective, some argue that a CRIS already has the ability to contain full-text publications and has to be enhanced accordingly to act as a repository. Their main argument is that a CRIS already has a rich and contextual data model which can express temporal roles and relationships between entities (Jeffery *et al.*, 2010).

KIM-FIS — the Current Research Information System at KIT

Information about the research output and the context of research at KIT is currently held in numerous systems run by different organisational units using different formats and data models. It is not possible to combine, aggregate or integrate this rich information, whether it regards scientists, institutes or other characteristic features. This became all the more evident when the University and Research Centre Karlsruhe merged to form KIT. Both organisations had developed their own policies and daily practices in collecting, managing and evaluating research information.

These heterogeneous ‘cultures’ make it difficult if not impossible to gain a consistent overview of and insight into the research performance of KIT. However, such an overview is of crucial importance for KIT because it is competing with the top research institutions on an international level.

Therefore, in March 2010, the executive board of KIT initiated a project to build an integrated research information system as one of the sub-projects of KIM called KIM-FIS. This means that it will be part of the service-oriented concepts developed within iSOA. In a tendering process, KIT decided to choose Converis (by AVEDAS AG)⁶ as a software platform for building its CRIS. Converis fulfilled most of the requirements which had been defined and both parties were interested in a strategic partnership promoting agile

development on the one hand and long-term development of the product on the other. After the decision to acquire Converis in July 2010, a demonstrator was built which focused on managing information about publications and projects. KIT library is responsible for the former, the research department for the latter. The project is co-ordinated by the CIO and there is a steering committee in which the library, research department and the department for planning and control are represented. The researchers and their activities are at the core of both sub-projects within KIM-FIS. A key challenge is to automate data collection and integration as much as possible. This includes migration of the existing publication databases, repositories and project databases within the former University and the former Research Centre, which are not integrated in researchers' workflows very well. In KIM-FIS, data is acquired and validated only once and then they can be used for manifold purposes, such as websites, CVs, publication lists and reference management software. In December 2010 the demonstrator was presented successfully to different stakeholder groups within KIT, namely researchers, the executive board and the Committee on Information Provision and Processing (IVA).

Publication Management

Publication management within KIM-FIS is also part of the open access strategy of KIT. As researchers strive for impact and visibility for their published results, they are willing to provide full-text files as long as they can be sure that copyright issues and other administrative burdens are taken care of. This is underpinned by a rigorous publication policy within the former research centre. It states that all publications have to be approved by the chair of the respective institute before they can be handed over to a publisher. The approval workflow ensures that coverage of the publication database of the former research centre is nearly 100%. The database has existed since 1956 and it holds about 75,000 records. At the former University, a publications database has existed on a voluntary basis since 1968 and it contains about 60,000 records. Currently a general publication policy for KIT is in the making which tries to combine the best of both worlds. The central idea is that the chair of an institute can decide if he/she wants to approve publications planned, as he/she is responsible for the scientific output of his/her unit. KIM-FIS will support the workflows which are currently being tested and rolled out to selected researchers and institutes. A first production release of the system was scheduled for August 2011. For the remainder of 2011, more

institutes will gradually use KIM-FIS and give feedback, allowing for one or more iterations in development if necessary. The general roll-out of the system is scheduled for 2012.

From the very beginning, it became obvious that KIM-FIS was strongly linked to the identity management system already established in KIM-IDM.⁷ It provides single-sign-on facilities for a wide variety of internal and external services at KIT. Roles and rights management must be clearly defined and the unique identification of researchers, funders and organizational units must be warranted. Internationally standardised persistent identifier systems for these and other entities are a natural requirement within projects like KIM-FIS, as they will greatly help the mobility of researchers' portfolios. However, there is not even a consensus about which persistent identifier systems should be deployed on a global scale for publications and which services should be supplied by them (Hakala, 2010). ORCID⁸ and other activities in this domain are carefully monitored; for the time being a pragmatic 'bag of identifiers' approach seems most promising for KIT. This means that maybe more than one (persistent) identifier will be attached to a person or a publication.

In addition to the publication databases, two open access repositories have been developed locally, one at the former University and one at the former Research Centre Karlsruhe. KIT is a strong actor in open access. It has passed an institutional policy endorsing the green and gold routes. It has signed the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities and has established a fund to support open access publication charges.⁹ It also runs KIT Scientific Publishing,¹⁰ an institutional open access publishing house which actively engages in the Association of European University Presses (AEUP).

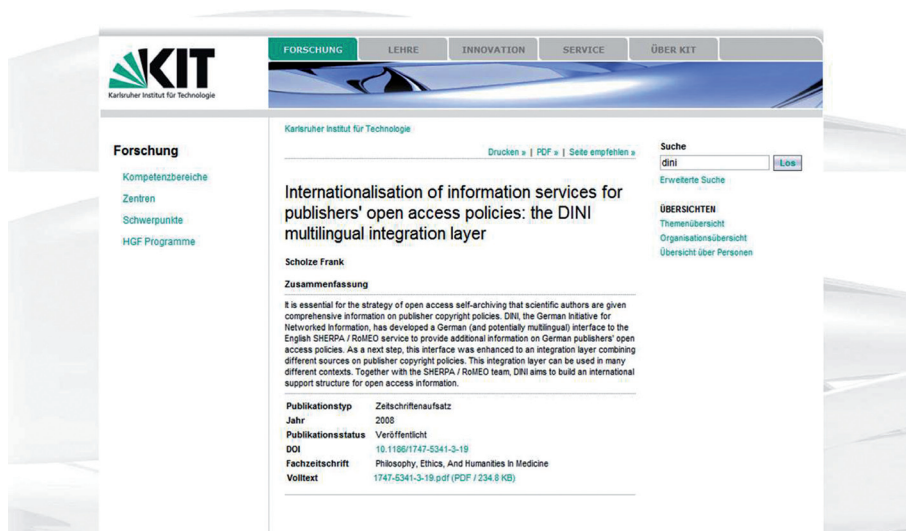
It soon became obvious that none of the locally developed repositories would be the long-term solution to supporting the green road to open access at KIT. Two options remained: to migrate to a new repository platform like dspace or fedora, or to expand the possibilities to handle full-text publications within Converis. As has been said before, one of the main arguments in favour of full integration is that Converis already supports a rich and contextual data model which can express temporal roles and relationships between entities. Therefore, it seems natural to follow an integrated approach which avoids the need to develop and maintain interfaces between the CRIS and a repository.

As KIM-FIS is gradually being deployed at KIT during the second half of 2011, a separate project (KIM-FIS Repository) will develop additional functionality to turn Converis into an open access repository. The benchmark will be the DINI Certificate for Document and Publication Services (Dobratz and Scholze, 2006).¹¹ DINI, the German Initiative for Networked Information, is a coalition formed by German Higher Education infrastructure and service institutions, such as libraries, computing centres and media centres, as well as by scientific learned societies. The DINI Certificate aims at networking document and publication repositories by promoting the use of standards, interoperability and co-operation between Higher Education institutions. It formulates requirements and recommendations covering technical, organisational and policy aspects like visibility, author support, legal issues, indexing and interfaces, access statistics and long-term availability. Looking at the technical parts of the DINI Certificate, the main focus in development will be on external interfaces, like OAI-PMH, SWORD, OpenAire and others, adding to the visibility of publications which is not a traditional feature of a CRIS. From a researcher's perspective this makes perfect sense. A research information system is the back end in which his/her whole research portfolio is available for re-use and external communication. It should provide publication lists for his/her website, for project proposals as well as for his/her CV or for scholarly communication with peers. From this perspective, separating a CRIS from a repository seems is much like erecting an artificial barrier, although the historical development is clear. A CRIS will always have an internal side which is used for evaluation and reporting. This does not prevent its rich and networked data from being displayed externally to achieve maximum visibility. When it comes to publications, it enables storage of all full-text articles and then display of links to the publisher website or internal or external repositories depending on the publisher's open access policy and the user's rights and roles (Figure 3).

Perspectives

The motto for developing KIM-FIS is 'think big in small steps'. Therefore a CERIF-based rich data model has been implemented, although, as first steps, only the publication and project information modules will go into production service. Patents, prices and other assets of research output will follow later in 2012 and 2013. By this time, publication management including repository functionality will be operational for some time and will play its part in a

Fig. 3: KIM-FIS demonstrator with links to the published and the repository version of a publication.



large-scale evaluation exercise within the Helmholtz Association of Research Centres, the so-called Programme Oriented Funding (POF).¹² Indicators based on publications will be calculated for comparing the Research Centres (KIT being one of them). KIM-FIS will then be the backbone for this exercise at KIT.

Acknowledgements

The authors would like to thank Karlsruhe Institute of Technology (KIT) and the Ministry for Science, Research and the Arts of Baden-Württemberg (MWK) for supporting this work.

References

Björk, B-C., P. Welling, M. Laakso, P. Majlender, T. Hedlund, *et al.* (2010): 'Open Access to the Scientific Journal Literature: Situation 2009'. *PLoS One*, 5(6), e11273. doi:10.1371/journal.pone.0011273.

- Carr, L. (2010): 'EPrints: A Hybrid CRIS/Repository'. In: *Workshop on CRIS, CERIF and Institutional Repositories*, 10–11th May 2010, Rome, Italy. <http://eprints.ecs.soton.ac.uk/21048/>.
- Dobratz, S. and F. Scholze (2006): 'DINI Institutional Repository Certification and Beyond'. In: *Library Hi Tech*. pp. 583–594.
- Freudenstein, P., L. Liu, F. Majer, A. Maurer, C. Momm, D. Ried and W. Juling (2006): 'Architektur für ein universitätsweit integriertes Informations- und Dienstmanagement'. In: *Proceedings of INFORMATIK 2006 — Informatik für Menschen*, 36. Jahrestagung der Gesellschaft für Informatik. Dresden. pp. 50–54.
- Gaedke, M., J. Meinecke and M. Nussbaumer (2005): 'Aspects of Service-Oriented Component Procurement in Web-Based Information Systems'. *International Journal of Web Information Systems (IJWIS)*, 1(1), 15–24.
- Hakala, J. (2010): 'Persistent identifiers — an overview'. *KIM Technology Watch Report* 2010. <http://metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/>.
- Hartenstein, H., W. Juling and A. Maurer (2007): 'Karlsruhe Integrated Information Management — KIM'. In: *Changing Infrastructures for Academic Services. Information Management in German Universities*. Göttingen: DINI — Deutsche Initiative für Netzwerkinformation, <http://www.dini.de/fileadmin/docs/DINI-Informationeninfrastrukturen.pdf>.
- Jeffery, K.G., A. Asserson and D. Luzi (2010): *State of the Art and Roadmap for Current Research Information Systems and Repositories*. euroCRIS White Paper. <http://www.irpps.cnr.it/it/system/files/PositionPaperRoadMap20100610.doc>.
- Razum, M., E. Simons and W. Horstmann (2007): Institutional Repository Workshop Strand Report — Exchanging Research Information. <http://www.driver-repository.be/media/docs/KEIRstrandreportExchangingResearchInfoFINALFeb07.pdf>.
- Sheppard, N. (2010): 'Learning How to Play Nicely: Repositories and CRIS'. In: *Ariadne Issue 64* <http://www.ariadne.ac.uk/issue64/wrn-repos-2010-05-rpt/>.

Notes

¹ <http://www.eurocris.org>

² See the talks at the DINI-Helmholtz-Workshop: *Repositorien — Praxis und Vision*, Berlin 2010, <http://www.dini.de/veranstaltungen/workshops/dinihelmholtz-workshop-repositorien-praxis-und-vision/programm/>

³ <http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1294&lang=1>

⁴ (Björk *et al.*, 2010) show that repositories in the STM field cover about 12% of the published Journal articles. Cf. Paul Jump: Research intelligence — Empty buckets or pure gold? In: *Times Higher Education* 17 March 2011.

⁵ Workshop on CRIS, CERIF and Institutional Repositories. 10–11th May 2010, Rome, Italy. <http://www.irpps.cnr.it/it/eventi/workshop-on-cris-cerif-and-institutional-repositories>. 2nd WORKSHOP ON CRIS, CERIF AND INSTITUTIONAL REPOSITORIES 'Integrating Research Information: CRIS+OAR' 23–24th May 2011, Rome, Italy. <http://www.irpps.cnr.it/it/eventi/2nd-workshop-on-cris-cerif-and-institutional-repositories-integrating-research-information-crisoar>.

⁶ <http://www.avedas.com/en/converis.html>.

⁷ <http://kim.cio.kit.edu/138.php>.

⁸ <http://www.orcid.org>.

⁹ <http://www.bibliothek.kit.edu/cms/english/publishing.php>.

¹⁰ <http://www.ksp.kit.edu>.

¹¹ <http://www.dini.de/dini-zertifikat/english/>.

¹² <http://www.helmholtz.de/en/about-us/programme-oriented-funding/>.