E-Science related Information Services in the Context of Digital Libraries

Achim Osswald

Institute of Information Science
Cologne University of Applied Sciences
50678 Cologne, Germany
achim.osswald@fh-koeln.de

Abstract. E-Science and Grid-Computing are supposed to enable synergetic and collaborative workflows in the Digital Library context. Only a few enhanced information services adapted to the new circumstances in science and research have been realized so far. Project based acitivities focus on IT-related premises and results of collaborative workflows. This paper emphasises the relevance of library and information science (LIS) based information services and know-how in e-Science activities. They could improve and ensure the quality of information services provided. Projects funded by authorities in the EU and Germany are analyzed to see the coverage of LIS-based services. Only very few of these e-Science projects realize LIS-based services. This could reduce the quality of research in the long run and LIS experts might loose their role and expertise in serving scientists.

Keywords: e-Science, library and information services, digital library, Grid-Computing, EU, Germany, GB, USA

1 Introduction

Since years scientists use e-based communication tools and exchange their research results via the internet. Pre- and postprints of publications are delivered digitally via institutional and subject based repositories. Peer-reviewing and other ways of science related communication takes place via the internet. But compared to todays practice e-Science (=digitally enhanced science) promises to provoke more efficiency and comprehensiveness in these academic activities. These advantages will be enabled by Grid-Computing and services based on it. E-Science is based on “highly distributed network environments” (http://en.wikipedia.org/wiki/E-Science) providing the software tools and computer power necessary to process large sets of data – by interconnecting computers and tools whereever they are available. This

is supposed to enhance information exchange and intensify communication resp. cooperation between researchers. Their competitiveness as well as their ability to cooperate on the international level will be improved.
E-Science programmes in several countries have initiated projects which try to investigate these benefits promised by e-Science in theory. It seems that information services adapted to e-Science do not play an important role within these projects and plans. Nevertheless such kind of services will – or at least should – become relevant. Without information services adapted to the e-Science environment scientists will not be able to benefit from e-Science as proposed. If libraries and other information service providers[1] will not realize these services others may overtake their role inadequately and LIS institutions might loose there connection to scientists.
This is the background scenario for analyzing projects related to e-Science. The objective of this paper is to investigate if and how these projects are realizing LIS-based information services by applying the specific know-how and experience of librarians and information scientists. The paper will focus on e-Science projects in Germany and the European Union (EU) which have been funded by national and EU-support programmes. Preliminary research indicates that e-Science projects have a focus on a broad range of newly developed and designed IT-applications but do not focus on LIS based information services so far.[2]

2 E-Science and Scientific Working Methods

2.1 Scientific Working Methods today

Today scientists mostly work net-based. From their desk they access special hardware, software, data, and applications – all of them realized as distributed systems. Typically access to them has to be made explicite by keying in specific data. Typically cooperation with other scientists e.g. exchange of interim or final results of their research for annotation and further use takes place via the Internet. Problems occuring in this context are incompatibility of software, insufficient

[1] Which will be abbreviated as LIS institutions further on.
[2] See for the need of a service framework Lavoie, Geneva & Dempsey, 2006.

knowledge of the software applied and missing or ignored standards as well as licensing hurdles and accounting problems.

2.2 E-based Scientific Working Methods in the DL-Context

E-Science – as described – tends to an unconstrained working environment based on Grid-computing[3] and collective use of resources. Knowledge about the resources available relies on the semantic web concept based on machine readable metadata. Users do not have to know where and how the heterogeneous data is available. This all needs Grid-compliant application software as well as middleware supporting those Grid-applications.
The advantages of Grid-based workflows refer as well to the advantages of e-Science (see also BMBF, 2005):

- Organisations and teams are grouped virtually related to projects
- scientists work mobile whenever they like
- Cooperation on a national level as well as on the international level will be eased, sometimes facilitated at all
- Results are getting transparent, synergies and enhancements will be accomplished when cooperating
- Parallel or duplicated research activities can be avoided due to knowledge about activities of others.

Implications in the context of hardware and networking infrastructure are also relevant: A lot of money has been invested in such kind of projects on the national and international level.[4] But there are as well implications regarding information services and the LIS institutions providing them. This topic will be analyzed further on.

3 Digital Library Services and e-Science

3.1 Data Capture, Reference and Access

Since several years funding programmes by the EU, USA and several European countries have been implemented to provide the conceptual and network-related needs of e-Science. Preliminary projects in

[3] See the interesting animation on Grid-based working procedures at http://gridcafe.web.cern.ch/gridcafe/animations.html.
[4] See e.g. the related funding programmes in Great Britain, the Netherlands and Germany.

scientific disciplines generating and using mass data like physics, meteorology or space science have shown the importance of implementing complementary infrastructure for storage, managing and provision of primary data. In the e-Science discussion the term *Digital Library* has been used frequently. Unfortunately a very limited understanding of the term Digital Library is applied (see e.g. Heery 2006).[5] The concept of "Library service" is reduced to storing and providing data for research – embezzling that library based services are much more. In fact it is an expansion of classic library activities because such kind of services do not provide publications and information collected only, but also sets or collections of primary data.[6] All of them are described by metadata and made available for usage. These traditional library competencies will play an important role as a prerequisite for successful e-Science based research.

Technical information services are connected to these traditional LIS-based services. In most cases they will be provided in cooperation with partners in an university or research institution. Such kind of interconnected information services are e.g.

- dedicated or „single sign on" authentification for accessing the Grid
- authorisation of access to specific resources.

On top of these additional services should be applied, e.g.

- standardized description of all resources generally available
- detection of all resources currently available
- provision and presentation of resources for further usage.

3.2 Personalizing Resources by Value Adding Services

Personalizing information services in the Grid-based working environment will be an important part of adjusted services. This should be implemented according to the current needs of a group or a single researcher, e.g.:

- Semantic web-based selection and integration of resources according to the related metadata
- Modifying retrieval parameters related to the working environment and the resources to be searched in

[5] See general trends in Digital Library-related issues at Bollen et al., 2005.
[6] See this concept e.g. in the presentation of Lyon, 2006.

- Provision and output features well-adapted to the working environment
- proactive information services
- rating and recommender services
- documentation and preservation of data and research results.

These are information services not really new to the community. Nevertheless by embedding them into the Grid they are enriched by quality and service. This results from their customization to the needs of virtually built target groups (see BMBF, 2005).

3.3 Supporting Infrastructure for Academic Publishing and Resource Linking

Author and title approach as well as subject indexing of resources by extracting and adding metadata has always been a basic activity of librarians. The same applies to linking of publications and other resources with data – and vice versa. But there are at least several other options for supporting services like e.g.:

- Reference linking to citations
- Linking to domain related or other authority files
- Semi-automatic generation of metadata
- Generation of subject related descriptors / metadata to be registered at search engines
- Allocation of consistent identifiers
- Integration in procedures of long-term preservation.

Publishing is constituent for research and the development of ideas. E-science needs dedicated supporting infrastructure for publishing processes within the Grid. Therefore the development of an adequate publishing infrastructure is an important part within the supporting measures taken by governmental and research funds. Libraries can become an important part of the publication process at research institutions as shown in projects like DILIGENT (see chapter 4.1)[7] or those initiated by the *Initiative Digital Peer Publishing (DiPP)* in Germany.[8] DiPP supports several projects on digital publishing. In

[7] See as well slide 16 and 17 of the presentation of Castelli, 2006.

[8] *http://www.dipp.nrw.de/index_html?set_language=en&cl=en*
DiPP "supports novel forms of scholarly communication by providing technical, legal, and organisational frameworks and tools for scholars, libraries and academic institutions -- for better digital information sharing and distribution among peers in a

these cases libraries are an important and sometimes the main player in the creation of new academic e-Journals.

The more the activities around eScience are dominated by IT institutions in a non-converged environment the more important gets the role of libraries in the publishing process. Since centuries libraries have been involved in this process by University Presses but today this role is functionally enlarged and will get essential for keeping in touch with members of the research community.[9] A wide range of services can be implemented within this context. All of them offer the chance for proactive appliance of LIS-based competence and can tribute to the awareness of library based services.

Excursus: Mentality Change of Scientists is needed

Working in a Grid-environment will make a mentality change of scientists necessary. Their dependence on local IT-know how will decrease but knowledge about the features and potential of IT has to increase. Scientists should be team-oriented the more they work in the Grid. Openness to share not only knowledge but also the results of collaborativly-based research will get important.
Additionally science institutions will have to develop new concepts of acknowledging and rewarding to adjust to the collaborative working environment. And software will be needed to handle all these new procedures and activities. There is some evidence that Open Source software and the Open Access model are meeting these needs more than proprietary solutions.

4 Information Services in Grid-based e-Science Projects

4.1 E-Science oriented funded by the EU

Since 1998 the *European Union* has launched grid-oriented support programmes within the Fifth and Sixth Framework Programme (1998-2002; 2002-2006). In the context of "Information Society Technologies (IST)" about 400 Mio. € have been invested but only a small part of the projects funded cover the topic of information services in a broad sense. Projects like DILIGENT (“A Digital Library Infrastructure on Grid

---

given disciplinary field and beyond”. DiPP is funded by the Ministry of Innovation of North-Rhine Westfalia, Germany.

[9] See slides 28-31 of the presentation of Thomas, 2006.

enabled Technology")[10] and GRACE ("Grid Search and Categorization Engine")[11] have some connection with information services but none of them deals with the wide range of those services mentioned in chapter 3 (BMBF, 2006a).
Several other projects like SimDat (Data Grids for Process and Product Development using Numerical Simulation and Knowledge Discovery)[12], DIP (Data, Information, and Process Integration with Semantic WebServices)[13], DataminingGrid[14], OntoGrid[15] and BRICKS (Building Resources for Integrated Cultural Knowledge Services)[16] fit even less into this category.

4.2 E-Science oriented Projects resulting from Support Programmes in Germany

In Germany a support programme called "e-Science und vernetztes Wissensmanagement" ("e-Science and networked knowledge management") was launched in 2004 (BMBF, 2004). Topics in the call related to information services focused on

- Media-integrating and process-oriented knowledge representation and metadata-systems
- Context-oriented procedures for access to distributed, heterogenious data sets and other resources
- Efficient publication systems based on open standards, procedures and resources to communicate results of collaborative research.

A list of 9 projects has been initiated in the years 2005/6 with in total about 17 Mio € financial support by this programme. Just three of them have – in a broad view – an explicit focus on library based information services. These are:

[10] http://www.diligentproject.org/
[11] GRACE focuses on batch search, subject indexing and knowledge management in distributed documents and data sets; see http://www.ub.uni-stuttgart.de/grace/.
[12] http://www.scai.fraunhofer.de/simdat.html
[13] http://dip.semanticweb.org
[14] http://www.datamininggrid.org
[15] http://www.ontogrid.net/ontogrid/index.jsp
[16] http://www.brickscommunity.org/

SYNERGIE (http://www.dai-labor.de/index.php?id=878; no LIS institution involved)
”The paradigm of SYNERGIE is to link information and knowledge with the help of innovative information services for the better support of users. The goal is to provide these innovative services in the form of a knowledge-rich, collaborative platform for the formation and the support of scientific communities. The integration of different information sources and cooperative services enables a new level of synergies between currently distinct areas of research.” (http://www.dai-labor.de/index.php?id=878). This will be achieved by

- ”integrating existing information sources,
- Conditioning and enhancing the information with relations and semantic concepts,
- Providing scientists with a personalized information management system, and
- Offering a collaborative working environment for scientists with similar interests.”[17]

Im Wissensnetz (http://www.im-wissensnetz.de; only a patent division with LIS know how is involved)
The project will try to transfer the concept of proactive, context oriented knowledge representation of cooperative workflows into an e-Science environment; it will apply text-mining technology on the metadata which is made available.[18]
Contributions planed by the project are

- a platform providing intelligent social networking
- a community-ontology
- tools for collaborative creating annotations
- an editor for collaborativly creating ontologies

ESciDoc (*http://www.escidoc-project.de*; one LIS institution involved)
The project will “realize a platform for communication and publication in scientific research organizations” (http://www.escidoc-project.de/homepage.html):

---

[17] Very few more details are provided via http://www.dai-labor.de/index.php?id=878
[18] See for a few more details http://www.im-wissensnetz.de/Wissensnetz/CMS/Einfuehrung/; unfortunately there is not much more information provided about the project so far.

1. ".Ensure permanent access to the research results and research materials of the Max-Planck Society and seamless integration within eSciDoc as well as integration into an emerging, global, electronic knowledge space.
2. provide effective opportunities for access to information for scientists of the Max-Planck Society and their work groups.
3. Support scientific collaboration in future eScience scenarios."[19]

The information provided by the three projects so far is very poor. Hence no clear judgement can be made how close they are really related to library based information services. The most relevant connection to traditional library based services are those services creating and/or using subject indexing. In an e-Science environment this will happen at least semi-automated serving knowledge management activities related to the distributed environment of scientists.

The other six projects within the programme are more or less related to repository building or technical aspects of data or knowledge management like

- *STEMNET* (Development of a knowledge management system focusing on the subject of stem cells)[20]
- *HyperImage* (development of picture-based e-Science-networks based on subject indexing of parts of pictures)[21]
- *WIKINGER* **(**Wiki Next Generation Enhanced Repository)[22]
- *WISENT* - Wissensnetz Energiemeteorologie (IT-optimized cooperation in reseach and development of organisations in the area of energy meterology)[23]
- *Ontoverse* - Cooperative knowledge management in the life sciences network ("development of a new, internet-based

[19] http://www.escidoc-project.de/the-project-escidoc.html
[20] http://www.dl-forum.de/deutsch/projekte/projekte_2681_DEU_HTML.htm; no LIS institution involved.
[21] http://www.dl-forum.de/deutsch/projekte/projekte_2688_DEU_HTML.htm and http://www2.hu-berlin.de/hyperimage/content/idee.htm; no LIS instition involved.
[22] http://www.imk.fhg.de/sixcms/detail.php?template=&id=2563&_SubHP= UnsereAbteilungen&_Folge=&abteilungsid=974&_temp=PR; no LIS institution involved.
[23] https://bi.offis.de/wisent/tiki-index.php?page_ref_id=83; no LIS institution involved.

application for cooperative and interdisciplinary ontology building in terms of an ontology-wiki")[24]

- *FRESCO* - Fraunhofer e-Science-Cockpit ("provision of an integrated and personalized working environment for researchers and developers").[25]

An indicator for the degree in which know-how of librarians will be considered may be the number of LIS institutions participating. As shown in the footnotes only two LIS institutions are involved as project partners. None of them is a library. Unlikely the specific know how available in LIS institutions will come into action in these projects. Perhaps further research may show that some individuals with LIS experience have been involved. Whether or not the picture does not alter in terms of LIS experience applied in the design and creation of information services within these projects.

4.3 E-Science oriented support programmes in GB and the USA

The picture taken from the German projects should be compared with projects in GB and the USA. A report published by the German Ministry of Education and Science (BMBF, 2006b) shows that only two out of 51 projects/institutions related to the topic are focusing on data curation and data access, none of them has an explicit relation to the topic of information services.[26] Details of this might be shown in another paper.

Since 2006 the newly implemented "capital programme"[27] offers new "development areas" (ibd.) one of which is "Users and innovation: Personalising technology".[28] But even in that context LIS-based services are not mentioned primarily. Further research is needed to

---

[24] http://www.ontoverse.org/en/projektbeschreibung/projektbeschreibung_1.html; one LIS institution involved

[25] http://www.ipsi.fraunhofer.de/i-info/en/content/view/97/0/

[26] „In einer Liste des e-Science Centre mit 51 Projekten und Einrichtungen des UK e-Science Programme finden sich zwei Einträge speziell zu Datenpflege und Datenzugang, nämlich das Digital Curation Centre und das Projekt Data Portal, dagegen kein Eintrag, der eindeutig dem Informationsdienstleistungsbereich zugeordnet werden könnte." BMBF, 2006, p.2.

[27] http://www.jisc.ac.uk/whatwedo/programmes/programme_capital.aspx

[28] http://www.jisc.ac.uk/whatwedo/programmes/programme_users_and_innovation.aspx; "To create opportunities to transform practice by developing technologies and processes that support the user experience in improved and innovative ways." (ibd.)

analyze which part LIS services will play within the projects to be funded under this programme.
A comparable overview and report is available on e-Science related projects in the *USA* summerizing these actitivities and their focus (BMBF, 2006c). Like the projects in GB their focus is on IT-related activities in the context of distributed storing, retrieval and access to large data sets. Additionally activities in long term preservation and access to data are listed. LIS-based services and their adjustment to distributed virtual communities of researchers, questions of licencing, accounting and personalisation are not main topics of these projects.

5 Discussion and conclusions

Despite the fact that e-Science always has been mentioned in relation to Digital Libraries there is very less involvement of library know how in e-Science activities so far. The concept of Digital Libraries is mostly interpreted by IT experts. So far LIS based information services adapted to the e-Science environment do not play an important role – if at all – in e-Science projects implemented in the EU, Germany, USA and GB. The paper as well has shown which variety of information services LIS know how could bring into action.
Nevertheless LIS institutions should try to gain expertise in services related to e-Science – especially services providing an e-Science related publication environment to researchers because they are still their primary clientel. Libraries as institutions as well as librarians should try to gain a role in the e-Science projects which are getting implemented during the years coming. Otherwise libraries will loose an important part of their tasks within the research community. So far there is little hope that any other group of professionals will fill this gap by serving users with high quality LIS-based information services. There is time left to keep track and jump on the bandwagon! It should not leave without LIS experts!

References

[1] Bollen, J., Nelson, M., Manepalli, G., Nandigam, G. & Manepalli, S. (2005). Trend Analysis of the Digital Library Community. D-Lib Magazine. 1; (doi:10.1045/january2005-bollen) *http://www.dlib.org/dlib/january05/bollen/01bollen.html*

[2] Bundesministerium für Bildung und Forschung[29] (BMBF 2004). *Bekanntmachung des Bundesministeriums für Bildung und Forschung von Richtlinien für die Förderung zu "e-Science und vernetztes Wissensmanagement"*, Berlin; *http://www.bmbf.de/foerderungen/3179.php*

[3] Bundesministerium für Bildung und Forschung (BMBF 2005). *Grid-Anwendungen im Bereich Information Management. Fokus: fächerübergreifende und Endnutzer-orientierte Anwendungsszenarien;* Berlin 2005; *http://www.e-science-forum.de/de/79.php*

[4] Bundesministerium für Bildung und Forschung (BMBF 2006a). *EU*, Berlin 2006; *http://www.e-science-forum.de/de/86.php*

[5] Bundesministerium für Bildung und Forschung (BMBF 2006b). *Großbritannien*, Berlin 2006; *http://www.e-science-forum.de/de/85.php*

[6] Bundesministerium für Bildung und Forschung (BMBF 2006c). *USA: Bibliotheken, Digital Libraries und Informationseinrichtungen*, Berlin 2006; *http://www.e-science-forum.de/de/88.php*

[7] Castelli, Donatella (2006): *Digital Libraries of the Future - and the Role of Libraries*, 8th International Bielefeld Conference 2006, Bielefeld (Germany), *http://eprints.rclis.org/archive/00005821/*

[8] Heery, Rachel (2006). *Digital Library Goes e-Science (DLSci06).* Workshop Held in Conjunction with ECDL 2006, September, 17-22, 2006, Alicante, Spain; *http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/october06/heery/10heery.html*

[9] Lavoie, B., Geneva, H., Dempsey, L. (2006). *A Service Framework for Libraries*. D-Lib Magazine 12, 7/8, doi:10.1045/july2006-lavoie

[10] Lyon, Liz (2006). *E-Science and the Role of Libraries*, 8th International Bielefeld Conference 2006, Bielefeld (Germany), http://conference.ub.uni-bielefeld.de/2006/docs/docs_main.htm

[11] Thomas, Sarah E.: (2006). *Publishing Solutions for Contemporary Scholars: The Library as Innovator and Partner*. Paper delivered at 8th International Bielefeld Conference 2006, Bielefeld (Germany) http://eprints.rclis.org/archive/00005820/

[29] German Federal Ministry of Education and Research