iPRES 2009

THE SIXTH INTERNATIONAL CONFERENCE ON THE PRESERVATION OF DIGITAL OBJECTS

Proceedings

October 5-6, 2009
Mission Bay Conference Center
San Francisco, California

California Digital Library

e-Infrastructure and Digital Preservation: Challenges and Outlook

Reinhard Altenhöner

Deutsche Nationalbibliothek
Adickesallee 1
60322 Frankfurt am Main

Abstract

Undoubtedly, long-term preservation has raised a great deal of attention worldwide, but in a broader perspective the attention is out of proportion compared to the number of real operational solutions – not to mention the big picture of a comprehensive digital preservation infrastructure. Up to now, the existing digital preservation infrastructure mainly consists of a small number of scattered trusted long-term archiving data repositories, which have the control of and the responsibility for our digital heritage. The ongoing discussion on e-Infrastructure points out that we are still lacking reliable structures which support the expected integrated digital preservation infrastructure provided jointly by cultural heritage organizations, data centers and data producers. The slow development towards a global integrated digital preservation infrastructure in combination with adjacent pressing questions for the information infrastructure in general has led to an extended strategic discussion within the last years in Europe and the US: In studies, position papers and evaluation approaches we find elaborated building blocks for a roadmap and definitions for a more advanced landscape. With a focus especially on the German situation and on existing and ongoing practical experiences, the paper discusses reasons and strategic aspects which maybe can illuminate the prohibitive factors for the unassertive progress.

Digital Preservation

The impact of digital preservation in the portfolio of cultural heritage organizations and data centers is becoming increasingly important. Digital preservation (DP) describes the process of securing sustainable access for use and reuse of all kinds of digital data and publications. DP comprises in this sense not only bit stream preservation but also continuous activities starting with the ingest (validation of the objects, extraction of relevant technical metadata, etc.) and pursued by recurring interventions to migrate the object.

The issue is on the agenda of service providers in the area of information. If institutions that offer information services including DP fail in this area, customers and financiers will probably reflect their support. In this sense those institutions have to verify that they are able to fulfill the need for sustainable access to information. Practically, they are under pressure to have solutions at their disposal which allow them to care for digital objects in their responsibility.

And obviously there are some solutions available which announce themselves as DP, mostly dedicated to single institutions or specific types of objects. But in general there seems to be a disproportion between the effort spent in discussions and research and the relatively little number of operational (broad scale) solutions to the quantity of real operational (broad scale) solutions. Surely it is difficult to confirm this statement with figures, because exact metrics for that are missing – but the fact that even wide distributed journals like 'Nature' pick up the topic[1] shows the dimension of uncertainty in the scientific community.

Probably one of the reasons for the delayed development is the effort necessary to set up and maintain appropriate systems. Continuing activities are necessary to keep digital objects available for a long time - this may be in contradiction to working procedures in the area of science and research.

The topic does not lack awareness or support in the current debate on e-Infrastructure. Even more, it is astonishing that progress is rather slow. There must be some more reasons which handicap the initialization of a broader DP-infrastructure. They shall be reflected in this paper.

"e-Infrastructure"

DP has raised a great deal of attention worldwide in the strategic discussions – often addressed in the surrounding field of e-Infrastructure and GRID-services.

The expression "e-Infrastructure" has become a synonym for everything related to the future workplace of scientists and researchers. But what are the essential characteristics?

"Infrastructure can be defined as the basic physical and organizational structures needed for the operation of a society or enterprise, or the services and facilities necessary for an economy to function" is the notion Wikipedia offers.[2] Transferring this to the area of science and research we learn that apparently everything is

[1] See Editorial contribution "Data's shameful neglect" in *Nature* 461, 145 (9 September 2009) doi:10.1038/461145a and related articles in this issue.

[2] http://en.wikipedia.org/wiki/Infrastructure

involved which enables the day-to-day work of scientists in a virtual, globally connected network. Asking for roles in this network we can identify the producer / author and the results of his work on the one side. On the other side some facilitators for science and research can be identified, traditionally

- research institutions
- educational institutions (often combined)
- the (scientific) publication market
- facilitators (libraries, professional information providers and data centers)

While publishers traditionally organize and facilitate the scientific communication and exchange, libraries traditionally secure the availability of professional scientific information in their current and historic dimension. These roles become increasingly merged: In the past the role of publishers and libraries were separate—today we have some overlaps: libraries offer institutional repositories and provide a space for digital publication.

A good example for the broader understanding of science and research infrastructure is the current 7th framework EU-programme (Research infrastructures) 'ICT Infrastructures for e-Science', which is dedicated to these major directions:

- Extend and reinforce the high capacity communication infrastructure GÉANT
- Strengthen multidisciplinary grid and supercomputing infrastructures
- Expand scientific data infrastructure
- Encourage the adoption of e-Infrastructure by an increasing number of user communities
- Stimulate new organizational models
- Support the construction of new computation and data treatment facilities (petaflop supercomputing)[3]

Taking these general goals into a dedicated perspective for data the commission states:

> The objective of scientific data e-Infrastructures is to develop an ecosystem of European digital repositories, combining and adding value to national and discipline-based repositories to respond to Member State requests to improve access to scientific information. […] Europe needs to pay particular attention to the accessibility, quality assurance and preservation of key data collections. […] In a heterogeneous digital data landscape, where it is estimated that only 28% of research output is managed in digital repositories, a new strategy for the management of scientific information and associated policies needs to be developed, based on the path-finding activities of key research stakeholders (e.g. EMBL, ESA, ECMWF, CERN) as well as academic institutions and libraries.

In fact the EU has equipped the funding track 'e-Infrastructures' with a total of 1.8 billion EUR until 2012. These activities are based on former funding within the 6th framework, where some preparatory projects for DP were initialized.[4]

R&D, strategic implications and practical implementations

> Ja, mach nur einen Plan
> Sei nur ein großes Licht
> Und mach dann noch 'nen zweiten Plan
> Gehn tun sie beide nicht.
> (Bertolt Brecht, *Dreigroschenoper*)[5]

In 2006 Digital Preservation Europe, funded as a FP6 Coordination Action, has highlighted a significant mismatch between the scale of the problem and the level of effort being mobilized to address the problem through research[6]. Based on DPE's analysis of existing research agendas and current research trajectories the preservation community has not adequately come to grips with the digital preservation problem. Progress has been very limited, scattered and of limited applicability.

Systems for long-term preservation of data, information and knowledge are being created in many fields, from the various disciplines of science to the equally broad range of cultural interests. Progress was and is being made especially in the EU-funded projects like CASPAR, PLANETS and SHAMAN, but there is a significant lack of progress in establishing a common approach to solving the problems of preservation across the spectrum of memory institutions.

[3] http://cordis.europa.eu/fp7/ict/e-infrastructure/home_en.html

[4] Grouping these former and partly ongoing activities into networking & exchange-, evaluation & targeting- and tool-development-projects, it came out that most of them are concentrating on technical issues. Examples are CASPAR ("it will form the basis of a continent-wide preservation infrastructure"), as well as SHAMAN, PLANETS ("developing a practical infrastructure for digital preservation").

[5] "Go make yourself a plan / And be a shining light. / Then make yourself a second plan / For neither will come right." (Bertolt Brecht: *The Threepenny Opera* (Trl: Desmond Ivo Vesey))

[6] DPE: Digital Preservation Europe, see; http://www.digitalpreservationeurope.eu/

To name one example: There are a number of different approaches to auditing and certification of trusted repositories and there are at least as many different approaches to preservation-related metadata models. As observed in the DPE roadmap, these models are: "not yet interoperable and integrated within a technical framework, not yet standardized themselves, not operating on the basis of appropriate basic standards, not yet fully taking the current state of the art in computer science into account"[7].

Even for experts it is difficult to keep an overview of all the long-term preservation related standards and ongoing standardization activities. Action to resolve this fragmentation is urgently required so that the suppliers and customers that make up the marketplace for DP solutions are able to move forward with confidence that research results will be inter-operable. Once achieved, an integrative approach remains necessary to prevent researchers from losing sight of the need for inter-operability.

Some years later in 2009 the PARSE.Insight Draft roadmap comes to similar conclusions with special regard to research data: The roadmap assembles technical and non-technical components (or delivers ideas like the suggestion for a normalization institute) aimed at bridging the "islands of functionality, developed for particular purposes […] separated by discipline or time"[8]. Relevant aspects in the high-level roadmap are—besides of the technical challenges—organizational and financial components, attended policy infrastructure concepts and components.

The listed components are ambitious and cover a lot of important measures in the sense of an activity agenda for funding. On the other hand the top-down approach should be accompanied by pragmatic initiatives driven from given problems and existing data. The implementation of components from scratch is risky and needs the integration of data producers and data centers respectively memory institutions.

Most of the demanded tasks should be accomplished in a cooperative way by different stakeholders in order to integrate requirements on a generic level. Only the systematic identification of these practical driven needs leads to operational solutions. Maybe this will take more time, but the experience with different assets for format registries explains very well, where the difficulties are.

The pragmatic, data-driven approach may be helpful in another regard: Often DP is considered as an isolated task which runs independently from other activities of memory institutions. This leads to complete and often closed solutions with separate ingest procedures and preservation rules. But in the perspective of producers and data holders DP has to become an integrated part of an often existing local, regional or global infrastructure. The integration in existing workflow procedures, the inclusion of new components into an existing infrastructure of data sharing and exchange is the key to provide distinct DP functionalities into the e-Infrastructure. In this sense DP is not an isolated phenomena but a self-evident part of the restructuring of infrastructure. The wide spread discussion on e-Infrastructure reflects this insight increasingly.

In the same way the approximation to those demands has to be done with practical solutions for DP in memory institutions.

Research data: refocusing the DP - Insight view on the producer side

Beginning in 2006 / 2007 the discussion around an e-/DP-Infrastructure became more intensive. The working paper on digital repositories of ESFRI in 2007 presents a good example for the discussion on the European level: principal demands for sustainable access and availability were stated and help to prepare the new funding policy in the EU.[9]

On a national scale, nestor, the German network of expertise in digital long-term preservation, has deployed some activities to identify the state-of-the-art in data preservation and define the needs of scientists and researchers[10]. Nevertheless the results are preliminary to some degree; they can be considered as first steps towards

7 http://www.digitalpreservationeurope.eu/publications/reports/dpe_research_roadmap_D72.pdf

8 See http://www.parse-insight.eu/downloads/PARSE-Insight_D2-1_DraftRoadmap_v1-1_final.pdf

9 The working group paper can be find under ftp://ftp.cordis.europa.eu/pub/esfri/docs/digital_repositories_working_group.pdf

10 nestor has initiated two relevant studies: nestor - materialien 6 - Langzeitarchivierung von Rohdaten / Thomas Severiens, Eberhard R. Hilf, Frankfurt am Main: 2006 (nestor - materialien 6) see http://nbn-resolving.de/urn:nbn:de:0008-20051114018 and Anforderungen von e-Science und Grid-Technologie an die Archivierung wissenschaftlicher Daten / Jens Klump, Frankfurt am Main: 2008 (nestor - materialien 9), see http://nbn-resolving.de/urn:nbn:de:0008-2008040103 . Results are summarized 2009 in a nestor position paper: "Digitale Forschungsdaten bewahren und nutzen": publication of nestor-WG Grid/eScience and LZA, see http://nbn-resolving.de/nbn:de:0008-2009071031 . Similar activities from DINI, the german initiave for network information with a position paper on research data, see http://nbn-resolving.de/urn:nbn:de:kobv:11-10098082

operational best practice recommendations: DP-activities have to be planned and carried out in a community-specific way and a serious cooperation between different actors for the information infrastructure is important. The risk of dividing research data and linked publications between the different types of responsible organizations has to be avoided.

An interesting asset to cover the needs of a whole campus was made with the German project 'KOLAWiss", which recently published its results. The results based on former questionnaires show very well the different requirements of individual scientific faculties and the high need for cooperation[11].

Based on a faculty-oriented approach PARSE.Insight has evaluated specific needs by taking three case studies with the High Energy Physics, Earth Observation, Social Sciences & Humanities[12].

The PARSE project (co-funded by the EU under the 7th Framework Programme) is motivated by the collaboration of different stakeholders in the Alliance for permanent access[13], which plays a key role in order to join forces of data producers, providers and memory institutions.

As an intermediate result one can note that obviously the close connection to the distinctive requirements of different scientific communities is the most important step to go forward. Community-based care of digital preservation seems to be therefore the most promising approach. One precondition is a serious level of participation by scientists; they deliver the context information on the data itself, the production environment.

Beside those steps of progress some issues remain:

- Obviously the DP-terminology is not stable defined especially with the faculty members
- Lack of capacities and motivation of science and research to support the DP-activities by clearing and describing the data represents a risk.

Individual project results provide us with interesting insight-views; they give us a clear idea on the importance and value of research data (and adjacent objects like preprints and 'final' electronic publications). But there is probably a gap between the dedicated DP-institution (in case it is defined) and the data producing side.

Considering the operational situation we can resume that there are good examples of running data archives—mostly in a strong collaboration with scientists—but organizationally independent. But we are far away from a situation where we have a broad infrastructure for those data available for sustainable cross-domain access.
In comparison the situation for (electronic) publications seems to be settled: Defined institutions are responsible to collect and to archive this type of digital objects. But there is a serious amount of work to organize the ingest in an efficient way in order to ensure that important information from the production process finds its way into the archives. Some open questions here lead to some general questions on legal issues.

Legal issues

With respect to legal, the situation is notably complex. The discussion is dominated from tracks like copyright issues, Open Access, Digital Rights Management measures, Access Rights (and fee models). In general there are a lot of open questions and the situation in the context of DP leads to uncertainty in the whole scene.

Some institutions have a legal mandate to collect publications and assure permanent availability. In the case of the national library of Germany the respective law was amended in 2006[14]. This was a very important step to implement DP-procedures into the organization and fixed the need for additional funding in order to fulfill the extended mission.

In this sense the official assignment of DP-tasks to defined organizations helps to clarify roles and responsibilities. This includes standards for data care/curation within research organizations. The existence of appropriate policies should become a precondition for funding of research.

In this context often the question of costs arises, in addition the need to know more on the potentials.

Economics in DP

A lot of investigations were done on this area; LIFE undertaken by the British Library is probably the best known model to count the long-term costs in the lifecycle of digital objects[15]. With regard to research data JISC has assigned the study "Keeping Research Data Safe" which points out a lot of basic considerations for a pragmatic

[11] See http://kolawiss.uni-goettingen.de/?q=de/node/10 (unfortunately only in German)

[12] http://www.parse-insight.eu/downloads/PARSE-Insight_D3-5_InterimInsightReport_final.pdf

[13] http://www.alliancepermanentaccess.eu/

[14] http://bundesrecht.juris.de/dnbg/index.html

[15] See http://www.life.ac.uk/

methodological approach to tackle the problem[16]. Additional recommendations offer advice for funding organizations and proposals for further investigations.

Beside issues described in other studies[17] one of the most complex issues is the clear separation of dedicated DP-costs in order to claim for basic and additional needs of resources. Therefore it is difficult to compare the results of different evaluations. Moreover, the level of uncertainty about the real costs of DP is high. This means that beside a clear terminology we need accepted key indicators to describe and compare DP-services with each other.

In close conjunction with the question of costs (and funding) the question of market potentials of DP arises. The better the prediction on the future development of DP in the private sector is the better funding agencies are able to target their action to most promising fields of interest. The DPImpact study "Socio-economic drivers and Impact of Longer Term Digital Preservation", financed within the 7th framework programme, will deliver first deeper insights into this topic[18]. The study focuses especially on the demand in business, companies and enterprises and identifies the needs for change in legacy systems in the direction of DP.

The national perspective

In addition to the strategic discussions and activities on the European level, similar efforts are undertaken on the national level. In the following, the example of Germany shall be discussed. On the one hand there are some initiatives to implement operational solutions[19]; on the other hand there is an ongoing discussion on strategic issues.

Besides some preliminary papers in the DP-community there are different initiatives to address the needs of science and research for a sustainable infrastructure for data and long-term access on a national scale. Beginning in 2005/2006 the German Research Foundation put the topic of DP into their strategic founding guidelines for the timeframe up to 2015.[20] In parallel, the paper "Neuausrichtung der öffentlich geförderten Informationseinrichtungen".[21] has been published by a joint taskforce of experts commissioned by the national ministry for education and research and the "Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung" (BLK). Facing the reorientation of the whole information infrastructure, DP was only briefly addressed in a single chapter: The need to preserve electronic publications in the first place, organizational issues (like the legal deposit), international collaboration and a first step to define national responsibilities were stated. So the goal was addressed, but pragmatic recommendations for implementation were missing. Some years later in 2009 a new approach was started by the follow-up organization of BLK, the 'Gemeinsame Wissenschaftskonferenz des Bundes und der Länder'. This appointed group finished a state-of-the-art report of the situation of technical information infrastructure in Germany, identifies gaps and general needs of science and research. In the next phase of the working plan the relevant organizations will be identified and suggestions for organizational task sharing will be made.[22] Some preparatory tasks were done in the Priority Initiative "Digital Information", a joined activity of the leading science organizations in Germany. Initialized up to 2012 the alliance is focusing on core activities like National Licensing, a National Hosting Strategy, Research Data, Open Access, Legal Frameworks and Virtual Research Environments.[23] Activities of working groups concentrate on recommendations and studies, e.g. for a national hosting strategy.

[16] See http://www.jisc.ac.uk/publications/documents/keepingresearchdatasafe.aspx

[17] Broader in scope and highly relevant: The Blue Ribbon Task Force on Sustainable Digital Preservation and Access: Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation. (Interim Report), See http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf. The published paper describes mainly the deficits, the recommendations and conclusions will be published end of 2009. Major barriers for DP are identified in the area of funding, responsibilities, motivation and lacking awareness resp. fear to fail.

[18] Publication is still in preparation.

[19] See the collected project information of nestor, the German network for long-term-preservation, see www.langzeitarchivierung.de/modules.php?op=modload&name=PagEd&file=index&page_id=53

[20] DFG-Positionspapier: Wissenschaftliche Literaturversorgungs- und Informationssysteme - Schwerpunkte der Förderung bis 2015, Juni 2006, see http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/positionspapier.pdf

[21] Neuausrichtung der öffentlich geförderten Informationseinrichtungen, Abschlussbericht der BLK-Arbeitsgruppe "Zukunft der Fachinformation", Bonn 2006, ISBN 3-934850-85-5, See http://www.blk-bonn.de/papers/heft138.pdf (no English version available)

[22] The report is not yet published, for further information, see: http://openpr.de/news/325157/Forschung-auf-Dauer-publik-machen-Experten-erarbeiten-ein-nationales-Konzept-fuer-Informationsinfrastruktur.html (permalink)

[23] For more information see http://www.allianzinitiative.de/en/start/

Reflecting the state-of-the-art discussion on the national scale (and the international discussion is similar) it comes out that the topic DP is now a self-evident part of the discussion. This is done necessarily on a very conceptual level, but the stage is already prepared. The discussion has switched from electronic publications as a 'final' result of scientific work to the whole publication chain. Especially the long-term availability of research data is the booster for new initiatives.

In the German situation the key demands of cooperation and integration are reflected on the one hand in the nestor-initiative[24] and on the other hand in activities which follow the kopal-project.

Kopal & DP4LIB

For some years now, long-term preservation systems or archives have been in place, most of them—noted in statements by the projects themselves—use the OAIS-framework as a scale for their own ambitions and in order to describe the basic architecture of the system. Another characteristic of these systems is the high degree of centralization, which means that we often find solutions which were made for one single institution. Another precondition for those developments is that these archives are typically dedicated to a small number of deliverers of digital material.

The systems work more or less independently from each other; small areas of cooperation have been determined but are not yet operational. So the exchange of archived objects and work sharing is still an idea in order to make these activities more efficient and to save resources. For some deliverers the parallel approach (and their presence in both archival systems) is a dedicated goal in their strategy, because having two different systems is a better guarantee that the ingested objects will remain available over time.

In a systematic perspective it can be stated that:

- Operating long-term preservation systems are concentrated on big deliverers with automated data processing routines.
- Most of the solutions are proprietary in a technical sense as well as being tailored to very specific user groups.
- The number and relevance of well documented machine interfaces in this field is low.
- The systematic decision and processing of digital preservation processes is hidden inside the "black box" of the archive solution: deliverers (producers like publishers or libraries as licensing partners) have no possibility to influence e.g. dedicated migration steps.
- The need to normalize workflows and objects is high.
- The interoperability between digital preservation archives is poor.

In this sense, the existing digital preservation infrastructure is dedicated to a small number of trusted long-term archiving repositories, which have the control and the responsibility of our digital heritage. The idea to establish a network of interacting systems—the safe places network—was announced by the National Library of the Netherlands (KB) in 2006 and some steps have already been made towards an implementation.

The German project 'kopal' (Co-operative Development of a Long-term Digital Information Archive)[25] had the mission to practically prove and implement a co-operatively built and used long-term preservation system for digital publications. Within kopal, the partners have developed a technological solution for long-term archiving that includes not only the archiving and bit stream preservation of digital documents, but also the implementation of preservation planning mechanisms (especially migration) for digital documents to ensure their accessibility in the future. Kopal is based on the DIAS solution of IBM, originally developed for the KB. Kopal leverages the commercial system DIAS with an underlying commercial software set of IBM-standard software, which was extended especially for remote access, enhanced metadata administration and extended machine-readable interfaces. Additionally, the open source software JAVA library was implemented in the project, used for automated ingesting routines (extracting of metadata, quality control, ingest and retrieval). Moreover, object validation and metadata extraction software was integrated and amended. The kopal Library for Retrieval and Ingest, koLibRI, is therefore important in that sense that the reuse and the possibility to integrate the features in other system-environments have a crucial impact on the success of the complete solution. Furthermore, the software is used for the migration of defined objects in the system in an automated workflow by governing the validation and access mechanism.

As of June 2006 the Deutsche Nationalbibliothek and its project partner, the State University Library Göttingen (SUB), have been ingesting parts of their digital collections into the system. In mid 2007 the project was finalized and

[24] Nestor in its new situation after the publicly funded period is presented at IPRES 2009 by S. Schrimpf. Specific effort on new stakeholder groups and the increasing relevance of "social" components like education / professionalization and standardization is the characteristic of the new approach.

[25] See for more information on kopal: http://kopal.langzeitarchivierung.de/index.php.en. For the following activities R. Altenhöner: Implementing a cooperative long-term preservation infrastructure solution for heterogeneous institutions – report on activities in progress in Germany, paper for the 74th World Library and Information Congress (2008), see: http://ifla.queenslibrary.org/IV/ifla74/papers/084-Altenhoener-en.pdf

now the operative phase of the project has started. In a cooperation contract between the libraries, IBM and the data center GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung), which drives the operational service of the archive, all partners have agreed to continue their work and to enable other institutions to join the consortium. The kopal archival system has been transferred into practical use and is about to be adopted by further partners from the library and heritage community.

As mentioned before one major goal in the project was the processing in a cooperative environment, which allows multiple institutions to participate on different levels of involvement. This means the work sharing (data center being responsible for the operational service and the bit stream preservation, the libraries resp. memory institutions being responsible for all the aspects of preservation planning and the ongoing functional extension of the solution). The transparent integration into existing library systems and the reusability through memory institutions plays a critical role. Considering the aspect of flexible reusability international standards for long-term archiving and metadata were adopted. In this way, both sustainability and the ability to further develop the system are guaranteed.

The digital preservation solution therefore needs to be embedded in the working environment and dedicated workflows in which cultural heritage organizations collect, share, disseminate and present digital objects. Basically in kopal a distinction of system users between "clients" (in different stages) and those who are responsible for the complete system was made. In this sense the possibilities to reuse kopal in a productive environment reach from the complete outsourcing to the inhouse system solution inside an institution. The last possibility is rather expensive in terms of funding and staffing. Therefore a more differentiated client-oriented solution was adopted:

Clients in the sense of account owners rent an account similar to the bank accounts we are familiar with. In kopal this account is a virtual area on top of the system that is independent from other participating institutions. This means that every participating institution has its own dedicated account which can be administered for its own purposes. In consequence those organizations assume the responsibility to curate the digital content they collect e.g. from the Web. So the role is extended to the obligation to run the ingest-service and especially the data curation activities in their own responsibility. An account-holding member uses the platform and additionally it is responsible for the normalization and evaluation of data. Even the presentation of the archived objects is part of the task and also, the planning, the conceptual preparation and the implementation requires additional steps such as the systematic migration e.g. of dangerous (or at least difficult) formats. This will take place together with other account-holders and needs investment in know-how, permanent monitoring and qualification of personal.

Clients in the sense of kopal-participants assign the cited tasks to other institutions with the status of an account owner, who is responsible for the data curation of digital objects and the services all around digital preservation. The participants are obliged to describe the policy which should be followed in the system for their own ingested entries; they select and describe the objects for the purpose of long-term preservation. From the perspective of a participant the solution makes sure that the amount of effort is reduced in comparison to the needs of a full archival system. At the same time it is possible to influence the rules and regulations in the archival system in order to participate in the discussions and to take over dedicated responsibilities for specific tasks in the whole process of digital preservation in a cooperative working scenario.

This model of operation and organization describes the range of possibilities and the potential options, where long-term preservation with this dedicated background could take place. The cost model basically developed within the kopal project is dependent on the degree of measures / services the leading organization (the "account owner") is willing to assign.

Advantages of this approach of sharing the tasks and the degree of responsibility are:

- shared resources
- shared licensing costs
- optimized use of distributed knowledge

The kopal project has developed in its life time the basic functions to implement and fill in the roles and responsibilities described before. But especially on the area of automated communication between different systems and the practical level of operational organization kopal still needs some additional development. This will be done in a second step with DP4LIP (Digital Preservation for Libraries).

Technically this means that there is a need for an enhanced rights management in order to provide different levels of ingest and retrieval. There is also a need for a seamless integration. Another goal is to obtain more information about the costs and amounts of work for the introduction of long-term preservation processing into different types of organizations. In the end, it will be possible to generate valuable estimates for a funding and investment model for a complete infrastructure solution.

The partners DNB and SUB will offer a package of services, which allows re-users to choose between different levels of service and to customize the existing solution to their specific needs. Furthermore the partners deliver dedicated consulting and operational services. Identified positions / factors in the cost model are:

- Consulting & support
- Detailed planning
- Hardware extension, licenses
- Adoption / customization of SW-components
- Ingest
- Operational service

The planning for the next steps started after it became obvious that the technical solution in kopal is not detailed enough to address the different needs of potential partners - especially smaller institutions, which bring the demand for simple and integrated solutions.

In the discussion with several organizations and institutions it has become clear that there is a need to have different models of implementation with a high degree of customizable options. Moreover, it was recognized that these institutions are motivated to become involved in the basic principles of DP. The requirements for this engagement are very different and this means that the single services must be applicable to the user needs. Therefore a consortium structure will be set up with a documentation center, different types of libraries and library service providers as well as a virtual consortium of other institutions basically located in the information infrastructure of science and research. This guarantees a broad range of potential requirements to make sure that the project covers most of the needs of those organizations in a prototypic way.

Therefore the existing services should be enhanced to a real cooperative solution—not only in a technical sense, but in an operational / organizational sense. So the starting point is identified to initialize the next step of development for the cooperative kopal solution:

Based on the kopal results up to now, the partners wish to improve the practical reusability of the software development. In order to create a generic solution that can be implemented in many heterogeneous environments and integrated as a part of the working policy of cultural heritage organizations, there is a need to develop an open concept with modularized service packages.

These are the general goals:

- Creation of a flexible long-term preservation infrastructure adapted to the needs of (smaller) cultural heritage organizations and their service providers
- Technical enhancement of the existing solution, conforming to the partners' requirements
- Implementation of a reusable process model and preparation of a handbook to introduce long-term preservation in (smaller) cultural heritage organizations

Conclusions

- The awareness of DP as a challenge for the whole information infrastructure is widely shared. This includes not only the acting stakeholders in the scientific production and publication chain, but also funders, research organizations and politicians.
- The pressure from funders and customers on memory institutions to provide DP-services is increasing.
- e-Infrastructure is much more than DP. But DP becomes a self-evident part in the discussion.
- Especially the need to preserve research data has extended the focus.
- Infrastructure is much broader than technical solutions can be. Financial, legal and organizational aspects have to be considered.
- The need to integrate much more information from data production domains into the archived documents (requirements for reuse, production environment, technical framework, semantic context) is apparently increasing. On the other hand the disposition of data producers and authors to cooperate by supporting DP-activities is unsure.
- Nevertheless DP-activities need to be in a close cooperation to the data-producer or –distributor.
- There is a serious need for a comprehensive terminology of DP.
- We need a clear understanding of mandates in the area of DP – related to institutions and to data.
- Common metrics for DP in addition to ongoing standardization activities will help to compare different initiatives and offers. This should include even workflow routines.
- Cooperation is crucial in the light of the huge dimension of DP. Everything appropriate to encourage cooperation and to share know-how should be supported.
- The development of efficient models to integrate DP into existing workflow processes is evident. Besides advanced DP-solutions broader participation structures depend on the level of complexity DP-providers offer to both: memory institutions and data producers.
- The integration of data curation into the missions of research and research related institutions is important.
- Initiatives for cooperation need to be sustainably funded.
- International funding collaboration is needed to initialize cross-border actions.
- Consistent and sustainable funding related to the basic tasks of an institution or to the production of data is needed. Projects should be linked to innovative approaches.
- Dedicated support for global needs like preservation planning tools, file format registration, migration strategies, registry of emulation measures is needed.
- Support of a trusted performance measurement structure (DP-services, certification / accreditation) is needed.