

Conference report Annual Conference 2009

24 November 2009
Koninklijke Bibliotheek
The Hague, The Netherlands

Alliance for Permanent Access
Prins Willem-Alexanderhof 5
2595 BE The Hague
The Netherlands
t: +31 70 314 0425
f: +31 70 314 0424
e: communications@alliancepermanentaccess.eu
www.alliancepermanentaccess.eu

← Alliance for Permanent Access →

Index

Welcome and Introduction of the conference by Chairman

- Dr. Peter Tindemans, Global Knowledge Strategies and Partnerships Page 3

Keynote Speaker PARSE.Insight

- Dr. David Giaretta, Science and Technology Facilities Council Page 3

Science Community Insight I II III

- **High Energy Physics** by Dr. Salvatore Mele, CERN Page 4
→ **Earth Observation** by Veronica Guidetti, European Science Agency Page 5
→ **Social Sciences** by Dr. Matthew Woollard, UK Data Archive Page 5

Cross Communities Insights I II III

- **R&D and technical tools** by Prof.Dr. Andreas Rauber, Vienna University of Technology Page 6
→ **Funding and sustainable business models** by Chris Rusbridge, Blue Ribbon Task Force on Sustainable Digital Preservation and Access Page 6
→ **Certification / accreditation mechanism** by Barbara Sierman, Koninklijke Bibliotheek Page 8

Sharing experiences from Parallel Sessions

- **R&D and technical tools** hosted by Reinhard Altenhöner, NESTOR Page 8
→ **Funding and sustainable business models** hosted by Peter Doorn, Nationale Coalitie Digitale Duurzaamheid [NCDD] Page 9
→ **Certification / accreditation mechanism** hosted by William Kilbride, Digital Preservation Coalition Page 10

Keynote Speaker EU

- Konstantinos Glinos, European Commission Page 10

Conclusions by Chairman

- Dr. Peter Tindemans, Global Knowledge Strategies and Partnerships Page 11

Photo impression Page 12

Download all presentations at our website: www.alliancepermanentaccess.eu

← Alliance for Permanent Access →

→ Welcome and introduction to the conference

Dr. Peter Tindemans, President Global Knowledge Strategies and Partnerships



Taking the global view, the digital divide is deep and wide. But even though we live in a digitally privileged world, we still need to go forward in building and developing a sustainable data infrastructure.

Experiences from European collaborations show that just co-ordinating national activities is not enough, we need to bring parties together on a EU-level to really initiate hands-on collaboration and from there, stimulate implementation of the wanted situation within the various member states. It is clear that no matter what activities we have ongoing right now, in the end the community model will turn out to be predominant.

Funding for data infrastructures should become an integral element of EU research policies and programmes, not in the least because continuous progress in R&D is essential for Europe and sustainable data infrastructures are a key requirement. Also important is to get a quality model on EU level off the ground.

Overall aim of this conference is to formulate an action list for the Alliance for 2010 and beyond. Already two points are clear:

- Build up a framework for the policy side and the technical aspects
- Keep on focusing on communication and creating awareness

→ Keynote Speaker PARSE.Insight

Dr. David Giaretta, Science and Technology Facilities Council



PARSE Insight objectives:

- Gather information
- Make inventory
- Formulate roadmap
- Identify gaps

Experiences with other types of infrastructures (phone, railways, etc.) have taught us that they go beyond mere technological needs and technical details.

We know that each infrastructure contains certain 'choke points' that need to be resolved before the infrastructure can be fully deployed: it is therefore essential to focus on identifying those choke points with respect to data infrastructures and tackle them right away. Another important lesson learned: focus on services rather than technology is crucial to the success of any infrastructure.

When it comes to data infrastructures, trust is extremely important for researchers. There is still a high level of reluctance to store one's data 'elsewhere'. We need long-term funding, but that is not a popular request with funding agencies. Furthermore, it is absolutely clear that the funding and organisation needed for data infrastructures cannot be born by one organisation alone.

← Alliance for Permanent Access →

Survey among scientific communities resulted in unexpectedly large response. Researchers' main motivation for data preservation:

- Research is publicly funded; data should be accessible for society at large
- Allow re-analysis and re-use of data
- Advancement of science
- Support validation purposes in the future

Do researchers actively share their data? Only 25% states that their data is open to all. Majority only has data accessible to own group. Constraints for data sharing are mostly related to fear of misuse and legal issues.

Based on the survey results, a list of threats was defined and was coupled to potential solutions. Here, the need for Digital Rights pops up. Countering these threats seems crucial for progress in this field, sharing will only take off once scientists feel comfortable in doing so.

Often, the need for data preservation is motivated by stating that we need to preserve 'for future generations', but that is not a very useful motivation. Future generations don't pay taxes and they don't vote, therefore it is a weak argument. Instead, we should focus on infrastructures that are useful and offer benefits to current generations.

Survey results showed high similarity across the various scientific communities. The general nature of concerns, threats, motivations etc. is apparent.

→ Science Community Insight I: High Energy Physics (HEP)

Dr. Salvatore Mele, CERN



The real issue in HEP when it comes to data preservation concerns the data that was generated in six facilities during the 80s and 90s, which that have been shut down in the meantime. According to researchers involved, this data is 'rotting away'. Not because of physical loss as tapes are still there and are being copied regularly, but because the (oral) knowledge on what is actually on those tapes, or in other words the human factor and the metadata, all that is gradually getting lost. We have the tapes but nobody knows what the data is about.

In the survey among the HEP community, roughly half of the respondents clearly stated that better science would have been possible if the 'old' data had been available for use.

Knows from own experience while still active in research that even as recently as 2004, the only way to store and preserve data was to write a 140 page articles filled with old fashioned tables in print, in spite of all the advanced ICT infrastructure that was present. And that sums up the real question for researchers in HEP: where do I put my tables of data?

Considering the way ahead, an important problem is that data preservation is not sexy. It does not get you tenure or grants and it doesn't advance your academic career.

Question from audience: what about ownership of data, how is that defined? Who owns the tapes?

Mele: So far, the issue of ownership hasn't emerged, because physicists are not interested in looking back to data. They build a facility, collect data, use that data to the point where nothing new can be extracted, throw them away and build a new facility to collect new data and the whole cycle is gone through again. Nobody would go back to previous data, but now the question of ownership is coming up, but we need to address that, no clear answers yet.

← Alliance for Permanent Access →

→ Science Community Insight II: Earth Observation (EO)

Veronica Guidetti, European Space Agency (ESA)

Following the survey among the Earth Sciences (ES)/Earth Observation (EO) community, ESA was surprised to find out that the community was already largely aware of data preservation issues and needs.

Survey showed that majority of researchers in ES/EO needs access to historical data. Important note: data like this cannot be reproduced, are not experimental.

ESA's view on the role of the Alliance in 2010 and beyond:

- Address dilemma: focus on accessibility of historical data or on setting up sustainable infrastructure for data preservation now and in the future?
- Evaluate whether there are successful models to be exported to other communities
- Put focus on operational requirements
- Stimulate the European Commission to get more involved; EC should act as prime coordinator
- APA can be the driving entity to move partners into a pre-operational phase
- APA is and will be fundamental in tackling operational challenges, they should also actively take up the role as coordinator of new initiatives
- More active contributions from partners are needed, as well as a clear mandate for APA and partners to really get things off the ground

→ Science Community Insight III: Social Sciences

Dr. Matthew Woollard, UK Data Archive



Social sciences have been facing data preservation issues for 40 years. A lot of expertise has been built up that may be useful to other communities.

We should realise that not everything can be saved with the same level of future access without an increase in expenditure. Gives example of a 'personal favourite'; a study dating from 1969 on 'personal attitudes towards soft cushioning'. This study hasn't been accessed since 1972 and is a clear example of the need for selection. Do we want to save everything and do we need everything to be accessible on the same level?

Standards are necessary, but we need to keep the workability in mind as well. Refers to an ISO protocol from 2004 on how to code for male and female subjects in a study. "It is a 17-page document that explains we should use '1' for men, '2' for women and '9' in case the sex is unknown."

← Alliance for Permanent Access →

→ Cross Communities Insight I: R&D and technical tools

Prof.Dr. Andreas Rauber, Vienna University of Technology



There is already a range of broad, comprehensive systems (for storage, sharing etc) available. There are also several metadata extraction tools available, but it is surprising to see how they differ in the answers they generate. Up till now, lots of effort have been and are being put into bringing together repositories and registries, but these are mostly geared towards broad, general formats. When it comes to tools for specific collections and/or communities, the situation quickly gets more complicated.

Still, in spite of all that is readily available, we have the feeling that something is missing, but what? Why do we not simply use the services that are out there? We shouldn't forget that data preservation is still a young discipline and most of the present tools are still in the prototype stage and are the results of proof-of-principle projects. The real issue is the final development; that is the hard part and it is usually not taken up by anyone. Who takes over once the project has been completed? That is a serious obstacle. Who's the problem owner? The result is that we now mainly have (a lot of) demo's and prototypes, but no actual workable solutions.

There is a need to educate the IT field on data preservation and it is essential to realise that most of the present tools were not developed for data preservation as such.

Remark from David Giaretta: We should be careful that the strong focus on tools might blur the vision on the infrastructure.

Remark from the audience: Cultural change is needed to accept final development as publishable work; for example a systems biologist makes a workable system out of a prototype, but the respective computer scientist will not accept this as new work, so it cannot be published. That is not a good incentive for the actual user to put efforts into really making prototypes work.

→ Cross Communities Insight II: Funding and sustainable business models

Chris Rusbridge, Blue Ribbon Task Force on Sustainable Digital Preservation and Access



The particular focus of the Blue Ribbon Task Force concerns digital information that ultimately serves the public interest (e.g. official and historical records, scientific research data, private data that may become part of public records, etc.). The task force's goal is to:

- Sample and understand best and current practices for digital preservation and access, and to begin to synthesize major themes and identify systemic challenges.
- Identify and develop useful economic models for digital preservation and access, and map these models to common institutional, enterprise and community scenarios. This part of our activities is the topic of the task force's Final Report, which is due early 2010.

← Alliance for Permanent Access →

The goal of the report is to provide actionable recommendations for decision makers seeking economic models for access and preservation that promote reliability, cost-effectiveness, trustworthiness and compliance to relevant policy and regulation.

Some intermediate conclusions and experiences from the BRTF activities so far.

- Benefits and incentives should be clearly articulated. When it comes to benefits of digital preservation, we have found that we need a clear 'value proposition and that the benefits should emphasize the outcomes of digital preservation. Also, we should cultivate a sense of value and promote a 'willingness to pay'. When it comes to articulating the incentives for decision-makers to act, we should convince them to accept the responsibility to undertake preservation activities, which can be stimulated by focusing on institutional 'self-interest', in the form of e.g. business opportunities, policy compliance or mission-driven tasks. It is also important to orchestrate incentives over the complete digital lifecycle.
- It is or otherwise should be made perfectly clear that we cannot preserve everything for all time and therefore we should prioritize our needs and allocate resources where they generate most value. We need to set realistic preservation goals and make sure expectations and capacities are aligned. Ongoing, efficient allocation of resources should continued to be supported as well as a transfer of resources from those who are willing to pay to those who are willing to preserve. Productive use of resources is essential.
- There is variety of organizational forms available that can deal with preservation activities, including organizations with no private interest in preservation, organizations with a private interest that act on behalf of others (e.g. research library) or organizations that have mandate to preserve (e.g. national archives). Whatever form is chosen, governance is key and should focus on strategy, responsibility and accountability. In the end, trust in both the organization and the governance mechanisms will prove the ultimate success factor.

Looking ahead, we have identified a number of challenges that need to be tackled:

- Currently, long-term preservation activities are funded by short-term resource allocations
- There is lack of clear responsibility for digital preservation and a general sense that it is 'someone else's problem'
- The incentives for those who are in a position to preserve and those who benefit are misaligned
- There is little coordination of preservation activity across diffused stakeholder communities
- Valuing and monetizing the benefits of digital preservation is needed to attract funding and investments

The economic analysis is underway, but we feel that market mechanisms are unlikely to emerge as the preferred models. The desire for openness will hamper full-fledged market-based activities. Questions to be addressed concern amongst others the balance between quantity and quality and the feasibility of economies of scale. What is clear is that research is a zero-sum game; instead of pointing to others (organisations) to supply funding for data preservation, we all should ask ourselves how much of our research funding we are willing to sacrifice to enable sustainable data preservation.

The final report will be ready January 2010 and will be considered successful if the recommendations are of practical use to decision-makers and will catalyze further work on economically sustainable digital preservation. Note: economic sustainability is all about risk management. There are many threats to the long-term future of digital materials and the economic threat is pervasive and immediate. Mitigating the economic risk should be at the heart of any sustainability strategy.

← Alliance for Permanent Access →

→ Cross Communities Insight III: Certification and accreditation mechanisms

Barbara Sierman, Koninklijke Bibliotheek



This introductory lecture focussed on auditing and certification of digital repositories. Barbara Sierman from the Koninklijke Bibliotheek gave an overview of the global attempts to establish mechanisms for auditing & certification.

She reflected on the fact that some twenty years of digital preservation has brought a wealth of information, many papers available on the internet, lots of interesting discussions are being held, several projects have been started and already published results. But what have all these activities achieved in providing a solution for big and small repositories was one of the questions she raised.

There seems to be a gap between these achievements and actual implementation in practice, she quoted Reinhard Altenhöner from the Deutsche National Bibliothek who said at a recent IPRES-meeting that progress is being made in Europe, especially through projects like among others CASPAR, PLANETS and SHAMAN which are funded by means of European funding mechanisms but that there is a significant lack of progress in establishing a common approach to solving the problems of preservation across the spectrum of memory institutions.

So are there standards related to digital preservation available which can facilitate the community in the much needed A&C process. Although they are available, again there is a gap between availability and usage in practice. According to Barbara Sierman pooling best practices and put those in a central accessible database would be one of the future solutions. In line with this is creating clarity on what level of organisation e.g. tasks, responsibilities and mandates is needed. This also includes the roles of all stakeholders.

→ Sharing experience Parallel Session I: R&D and technical tools

Reinhard Altenhöner, NESTOR



During the session, a lively discussion took off right away; some of the main observations/conclusions/ recommendations: First, concerning the technical issues it was mentioned that we have to explore how far we can standardize datasets. The need for standardization should not overlook specific characteristics or needs of the different communities and should not hamper workable solutions.

On the other hand, we have to accept that standardisation always implies a certain loss of information, but as long as that is outbalanced by the overall general benefits, it should not be considered a problem. Working out real-life situations is needed now and it was proposed to set up a study to test ten generic formats for data delivery and gather experience with different actors at the same time.

The group suggested that APA organises a metadata framework for scientific data sets, focusing on structure, rights and provenance. It was also suggested that APA should take the lead in creating guidelines for metadata, in which particular emphasis is put on the need to separate data from processes. Concerning maintenance issues, APA could play a role in requesting tool support involving commercial players. Finally, the need to educate young researchers on data management should gain priority.

← Alliance for Permanent Access →

Second, concerning organisational issues, the participants feel that persistent identification/citation tools are needed that link data to owners. Furthermore, guidelines should be set up on how to share work between communities and data institutions, which also pay attention to the different layers of terminology (community insight, cross links to a generic layer (other communities / times)) involved.

Some participants mentioned the need for a redefinition of the stakeholders in digital preservation focusing on their respective roles, demands and potential. Specifically, technology providers should be included here as an important party. The group sees a role for APA as a mediator / broker between research communities, memory institutions and technology providers. Finally, it was suggested to apply for a Network of Excellence-grant, which focuses on research communities.

→ Sharing experience Parallel Session II: Funding and sustainable business models Peter Doorn, Nationale Coalitie Digitale Duurzaamheid (NCDD)



Session hosted three presentations.

1st presentation by Michael Jubb, Research Information Network

Gave a fairly pessimistic outlook; generally researchers do not trust other researchers' data and do not expect others to work with their data as well. Another obstacle is the lack of career opportunities and prospects for digital curators, their occupation is presently unclear. We should involve the researchers and users more in this type of activity; also at this conference the majority of the participants is active on the supply side. Input from user perspective is lacking.

2nd presentation by Neil Beagrie, Charles Beagrie Ltd.

Presented project on keeping research data safe. During phase 1 of the project, an activity-based cost model was developed and tested. Phase 2 has recently been started, which focuses on the cost and benefits of digital curation. Preliminary results indicate that the benefits are primarily apparent on the short term and they decline over time. Furthermore, this decline appears to be steeper for processed and published data than for raw data.

3rd presentation by Anna Palaioyk - Data Archiving and Networking Services (DANS)

Case study DANS. Development of an activity-based cost model and 'score card', which can serve as a tool for analysis to determine whether resources are indeed dedicated according to priorities set. Turns out that in many cases, we are not spending the time and money according to where the priorities are (or should be).

During the discussion (in the session, not plenary), the importance of taking notice of the alarm bells set off by the research community was mentioned. Cost models are mostly concentrated on the storage and archive side of the issue and not on the access side. It would certainly help to put researchers in a more central position when discussing these issues. It was also mentioned by Salvatore Mele that ideally, the costs of preservation will amount to only a percentage of research funding, but the majority of these costs applies to human resources (according to the present models). What will happen when the amounts of data increase, will these costs scale or do they pose a serious obstacle? Finally, someone mentioned that the final declaration of the International Polar Year was signed by all parties involved and they all agreed on opening up their data, but so far only very few parties took the necessary steps. A frightening notion.

← Alliance for Permanent Access →

→ Sharing experience Parallel Session III: Certification and accreditation mechanisms William Kilbride, Digital Preservation Coalition

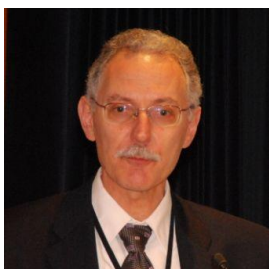


Participants started by examining experiences with current accreditation tools, which was followed by a discussion on the merits of self-certification. It was noted that some of the tools are difficult to use and implement and can be threatening to organisations. On the other hand, perhaps there is merit in setting a tough system, it is good to show to potential funders and users that you've passed such a difficult test. The major threshold mostly concerns the 1st round, but the benefits emerge during the 2nd, 3rd and 4th rounds; at that point the services will start to improve.

It is very important that we develop trust in these mechanisms, but we should also be realistic. No methodology for establishing that trust will give us 100% reliability.

Questions on certification and accreditation are particularly prominent in the context of outsourcing. When certain tasks are put into the hands of third parties, it is crucial to convince users of their reliability as well. Currently, there is a gap in service; we need a party that manages relationships between those who want to be audited and those who perform the audit (mainly consultants). Perhaps this a role the Alliance could take up? Another question the Alliance could address is whether we need a minimum set of standards?

→ Keynote speaker EU Konstantinos Glinos, European Commission



First time that e-science infrastructures are really on the political agenda (recent report / communication on ICT infrastructures for e-science COM (2009) 108).

Important to note: it's not only about data, Data is just one piece of the puzzle that is facing the researcher. What we need are integrated services – network use, computing tools etc. Scientists just want the simplest way to do their job, are not concerned with other issues.

We need to agree on some basic point concerning how we view the digital community in 20 years from now. It is no use to try to gain consensus on a fixed roadmap; that simply will not happen.

Ultimately, all this is about moving towards e-science workspaces: from the old situation of having all books in your office and separate rooms and units for all tasks to using your computer as the library, the lab, the network, the communication device, the archive etc all in one.

In analogy to the US task forces, an expert group is being set up. Their task is to interview key actors and key bodies in order to provide input on the way forward to reaching an overall e-science vision.

Future directions:

- Strengthen services and user orientation
- E-infrastructures go beyond scientific research (data problem is everywhere)
- Reflect on governance, ensure sustainability
- Strengthen global dimension

← Alliance for Permanent Access →

→ Concluding remarks

Dr. Peter Tindemans, Global Knowledge Strategies and Partnerships

Recommendations to the Alliance for Permanent Access (APA).

1. APA should take up, or rather continue, its role as a coordinator in the field and each APA-partner organisation should clearly assess its own role and added value. Today's conference made it clear that preservation should only be undertaken with a well-defined goal in mind:

- Advancement of science
- Future of mankind
- Democratization of data

APA can act as a clearinghouse for arguments in its ongoing tasks in communication and raising awareness. Another important message to bring across: focus not only on future, but also strongly on current benefits.

2. A mandate is needed to (top-down) set up repositories; we should get to work. APA is in the perfect position to lobby on a high level within the EU, on the level of national states, to get this mandate in motion. APA should organise high-level meetings for ensuring such a mandate. APA should also actively lobby for funding for data preservation and position of data preservation in future calls by the EC. The example of ESA should be followed, they succeeded in getting the mandate to start concrete actions.

3. Funding for data preservations needs to be sustainable and as science is a zero-sum game, data preservation should be funded through science funding. APA should play a role in establishing a target; what percentage of research funding should be dedicated to data preservation? Apparently the Wellcome Trust is willing to set up a data repository for the life sciences community. APA should follow that initiative and also adopt a few communities to get the work going.

4. With respect to technology, we should now move from prototyping to real world tests. APA should take the initiative to define the agenda for the road ahead and act as initiator to bring parties together (including technology providers, commercial parties) to try out those real life situations.

5. The discussions on certification and accreditation should be kept going. Even though the process is not easy, in the end it will prove worthwhile to have a tough system of auditing because that will create trust with users, partners and external relations. APA should take the lead working out a European mechanism for certification and accreditation.

General conclusion: APA is very well positioned to develop a strong agenda for action and should use that position actively.

← Alliance for Permanent Access →

Conference dinner | 23 November 2009



Conference dinner | 24 November 2009



← Alliance for Permanent Access →



Delegation
of the Finnish
Academies
of Science
and Letters



www.alliancepermanentaccess.eu