CASE STATEMENT PROPOSAL FOR THE RDA PRESERVATION e-INFRASTRUCTURE WORKING GROUP

14 February 2013

Contents

---

1 WG Charter

The purpose of the PeIWG is to reach wide agreement on the e-Infrastructure services which are needed to help repositories to preserve their data holdings, to ensure the interoperability of service implementations, and to build trust of service providers. Such distributed services supporting interoperability, including those that support continued usability, authenticity, accessibiliy, retrievability, visualization and replication, should allow the repositories to simplify, share the cost of, and improve, their preservation activities.

Many aspects of preservation need the support of the e-Infrastructure services which this working group intends to address, but one particular aspect should be emphasised. The Open Archival Information System preservation standard (ISO 14721) [1] requires that a repository's specified community maintain the ability to use (and re-use) the digitally encoded information. This implies that the repository provide whatever knowledge, software, and other means necessary for the community to use such information. Many repositories currently operate essentially independently, providing their user base with adequate capabilities for use: the e-infrastructure services we wish to define not only would allow for this to continue, but would enable a much broader set of users to use each repository's data holdings. One way of achieving such a result is expanding the specified communities for repositories, a significant step change, which can be enabled by the services defined by this working group, as part of the focus on community knowledge and usability.

The long-term vision is a standardization of preservation services and their application programming interfaces (APIs). The implementation of these services is outside the remit of this working group but we understand that many such services, or component services which can be brought together to produce the required results, already exist or are under development.

In addition we seek to guarantee that trust in the quality of the services is quantified using reproducible preservation metrics.

1.1 Months 1 to 6

The objective of the initial six months of work is to identify and compare the services which are available or being developed, including those from on-going projects (EU: APARSEN [2], SCIDIP-ES [3], EUDAT [4]; US: D-P-N [5], DataOne [6], CDL [7]; International: ESA LTDP [9], DPHEP [10], Records in the Cloud, InterPARES……..). In addition we will use results from existing surveys such as PARSE.Insight [11], and new DELPHI-based questionnaires to identify aspects not yet being addressed. Examples of these services include: various registries e.g. for Representation Information, for services, accessibility etc.; alert services to collect and distribute information about changes that affect preservation; services to assist understandability; virtualisation and visualisation services which facilitate data combination. There may be overlaps with some other RDA working groups and these will be identified in this period and specific preservation aspects discussed. As part of this we will examine evidence of the efficacy, scalability and areas of applicability of the various potential services.

1.2 Months 6 to12

During the following six months we will reach consensus on a minimum set of core services, their APIs and service discovery. We will also identify options for service interoperability where similar service offerings are available, and identify preservation quality metrics. The documents resulting from this effort will be published for a first public review around month 12 but interim versions may be useful for funding bodies.

1.3 Months 12 to15

We will update documents based on the received feedback and submit them to a second public review.

Following the second public review we will produce targeted explanatory material on exemplars of use of the proposed services; targets include high level decision makers, managers, developers and end users to support the take up of the services.

1.4 Long-term goals (months 15 to 18 and beyond)

We will finalise the e-infrastructure services document on the basis of the 2nd public review and support interoperable implementations of at least some of the services via planned developments outside RDA.

2 Value Proposition

2.1 Individuals, communities, and initiatives that will benefit from the RDA Preservation e-Infrastructure

- Preservation service providers – will be able to align their offerings and ensure interoperability
- Tool developers – will be able to integrate the use of the services into their tools
- Data centres, data archives and other data repositories – will be able to improve their preservation capabilities and also make their data usable by a wider community
- Vendors – will be able to enhance their storage systems to make preservation systems
- Data managers and data scientists – will be able to ensure the data they produce will be used longer and more widely
- Publishers – will be able to preserve the data with which they are entrusted and those that are related to their publications more easily.
- Researchers and other data users – will be able to have seamless access, use, reuse, to trustworthy data from a wider range of repositories.

2.2 Key impacts of the RDA Preservation e-Infrastructure

Use and re-use of data over time and across disciplines will be improved, addressing several of the issues highlighted by the High Level Expert Group on Scientific Data [12].

Society's valuable digitally encoded intellectual capital will be

- more easily and more cheaply preserved through sharing the effort and best practices,
- more reliably preserved, and
- more broadly used,

thereby increasing its value and contributing to society's wellbeing and wealth.

3 Engagement with existing work in the area

There are several projects currently underway in many parts of the world which aim at producing services which can be used to help preservation and re-use. These include:

- EU: APARSEN [1], SCIDIP-ES [3], EUDAT [4], DARIAH [19];
- US: D-P-N [5], DataOne [6], CDL [7], Data Conservancy [20], iRODS [21];
- Australasia: ANDS [8]
- Africa:
- India: Centre of Competence in Digital Preservation
- China:
- International: ESA LTDP [9], DPHEP [10], Records in the Cloud, InterPARES

We have members of the working group associated with or leading most of these initiatives and we believe that most of the remaining projects and organisations will be added to the working group shortly.

4 Work Plan

4.1 Preservation e-Infrastructure operation

- **Form and description of final deliverables**
  Deliverables include a widely reviewed and agreed definition of preservation services together with implementations of many of them.

- **Milestones**
  a. Summary of source material M6
  b. Candidate recommendations for services M12, with first public review
  c. Second public review M15
  d. Final versions, prototype implementations of some core services M18.

- **Mode and frequency of operation**
  Email plus virtual meetings every month with face to face meetings aligned with preservation conferences where possible. We recognize the difficulties which the spread of time zones introduces and we will vary meeting times to maximize participation.

- **Achieving consensus, addressing conflicts, and staying on task and within scope**
  - Consensus will be reached via open discussion, voting, and majority considerations informed by evidence where possible.
  - Conflict will first be addressed by WG leaders. An escalation procedure will be drafted, for example the RDA Council will be consulted, and an independent person not in the WG will be brought in to mediate the conflict.
  - Staying on task and within scope: we have considerable experience in projects and standards development. The key mechanism for reaching consensus will be by examining evidence and

identifying limitations of applicability of competing ideas. In addition of course we will agree on a detailed schedule and track action items.

- **Operation parameters**
  The work is voluntary, and not every WG member will be able to contribute equally, therefore we will aim to organise the work focusing efforts on members' specific interests but also to ensure that all members can contribute to internal reviews. The WG will hold internal assessments every 3 months to ensure we are track.

- **WG Assessment**
  The 3 monthly assessments will involve work group members and also external reviewers who have expertise in this area, including those who declined membership of the working group because of pressure of other work.

- **Broader community engagement and participation**
  At around months 6, 12 and 18 there are a number of preservation related conferences including APA [13], PV [14], IDCC [15], iPRES [16], PASIG [17] and IEEE [18]:
  - We anticipate the connections listed below will grow over the course of the WG's activities.
    1. EU: APARSEN, SCIDIP-ES, EUDAT, DARIAH;
    2. US: D-P-N, DataOne, CDL, Data Conservancy;
    3. Australia: ANDS
    4. Africa:
    5. India: Centre of Competence in Digital Preservation
    6. China:
    7. International: ESA led LTDP, DPHEP, ICA……..

  We recognize that services need users and we will work with the many repositories, and their users, with which WG members are associated. In addition we will work with vendors and solution providers to obtain their feedback and early involvement in implementations.

5 Adoption Plan

The Working Group comprises some of the major players in digital preservation who have good connections with repositories and user communities – the potential users of these services. The initial question they will address is: "why would users be interested in adopting the agreed services?". For example each repository will have their own community of users, often discipline specific, in the form of repositories and end users. Each of these will have its own national funding and user base, and associated justification. However each of the service providers and repositories recognize the increasing demands for preservation made on them in terms of complexity of objects, cost, scalability and interoperability. In addition there will be demands in terms of a greater, cross-disciplinary, community of users. All these will result in the implementation and take-up of one or more of the common service interfaces.

Beyond the immediate connections of the working group we believe there is a demand in associated stakeholders which, given a critical mass of initial users will result in growing adoption. Critical to this will be the involvement of tool developers and vendors, so we will make strenuous efforts to involve as many as possible. Some of the initial members are already involved in preservation projects and are interested in establishing the PeIWG's recommendations in their communities and centres.

Thus we believe that the agreed services will be implemented and used.

In summary the PeIWG provides the opportunity to spread the usage of each other's data and services, either directly or through externally agreed services,

6 Initial Membership

Confirmed members:

David	Giaretta	Alliance for Permanent Access	UK/NL
Andreas	Rauber	Institute of Software Technology Vienna University of Technology	Austria
Christoph	Becker	Institute of Software Technology Vienna University of Technology	Austria
Beniamino	Di Martino	Professore Ordinario - Professor Dipartimento di Ingegneria dell'Informazione Seconda Universita' di Napoli	Italy
Carlo	Meghini	Istituto di Scienza e Tecnologie della Informazione [ISTI] Consiglio Nazionale delle Ricerche [CNR]	Italy
Dinesh	Katre	Associate Director and HoD Human-Centred Design & Computing Group	India
Gillian	Oliver	School of Information Management Victoria University of Wellington	New Zealand
Heila	Pienaar	Deputy Director: Innovation & Technology Department of Library Services University of Pretoria	South Africa
Jamie	Shiers	DPHEP and CERN	International
John	Faundeen	U.S. Geological Survey, EROS Center	USA
Liuba	Shrira	Professor of Computer Science Brandeis University	USA
Luciana	Duranti	Chair and Professor | Archival Studies School of Library, Archival, and Information Studies The University of British Columbia	Canada
Martie	van Deventer	CSIR's Information Services	South Africa
Rainer	Stotzka	Karlsruhe Institute of Technology (KIT) Institute for Data Processing and Electronics	Germany
Reagan	Moore	RENCI, UNC, School of Information and Library Science	USA
Steve	Hughes	JPL, NASA	USA
Steven	Morales	Digital Preservation Network	USA
William	Underwood	Georgia Tech	USA
Milena	Dobreva	Univ. of Malta	Malta
Sam	Fineberg	HP	USA
Matthias	Hemmje	FernUniversität Hagen	Germany
Helen	Glaves	British Geological Service	UK
Ross	King	AIT Austrian Institute of Technology	Austria
Cal	Lee	UNC	USA
Grigoris	Antoniou	Univ Huddersfield	UK
Yuri	Demchenko	UVA	Netherlands
Sergio	Ruiz	DataCite	International
Tim	Smith	CERN	International
Mustapha	Mokrane	World Data System	International
Mike	Hildreth	Univ Notre Dame	USA

(other members will likely be added in the course of the working group startup)

7 References

[1] Reference Model for an Open Archival Information System (OAIS), ISO 14721, available from http://public.ccsds.org/publications/archive/650x0m2.pdf
[2] APARSEN project, web site: http://www.aparsen.eu
[3] SCIence Data Infrastructures for Preservation project (SCIDIP-ES), web site: http://www.scidip-es.eu
[4] EUDAT project, web site: http://www.eudat.eu
[5] Digital Preservation Network (DPN) web site: http://d-p-n.org/.
[6] DataOne project web site: http://www.dataone.org/
[7] California Digital Library see http://www.cdlib.org/ and microservices at https://wiki.ucop.edu/display/Curation/Home
[8] ANDS service: http://www.ands.org.au/
[9] Long Term Data Preservation activities in Earth Observation, see http://earth.esa.int/gscb/ltdp/
[10]DPHEP project web site: http://www.dphep.org/
[11] PARSE.Insight project – see www.parse-inisght.eu
[12] Riding the Wave, available from http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
[13] Alliance for Permanent Access (APA) web site http://www.alliancepermanentaccess.org, which includes the ODE (www.ode-project.eu) web site and maintains the CASPAR (www.casparpreserves.eu) and PARSE.Insight (www.parse-insight.eu) web sites.
[14] PV conference series – see http://www.alliancepermanentaccess.org/index.php/community/event/pv-conferences/
[15]IDCC conferences – see http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc
[16]iPRES conferences - see http://ipres-conference.org/ipres/
[17]PASIG conferences – see http://sun-pasig.ning.com/
[18]IEEE storage conferences – see http://storageconference.org/history.html
[19] DARIAH, Digital Research Infrastructure for the Arts and Humanities, web site: http://www.dariah.eu
[20] Data Conservancy, web site http://dataconservancy.org
[21] iRODS, web site: https://www.irods.org