OVERVIEW OF PARSE.INSIGHT: CONTEXT, OBJECTIVES, RESULTS SO FAR

21 September 2009
# Who we are

<table>
<thead>
<tr>
<th>Organization</th>
<th>Code</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science and Technology Facilities Council</td>
<td>STFC</td>
<td>UK</td>
</tr>
<tr>
<td>Koninklijke Bibliotheek</td>
<td>KB</td>
<td>NL</td>
</tr>
<tr>
<td>Deutsche Nationalbibliothek</td>
<td>DNB</td>
<td>DE</td>
</tr>
<tr>
<td>Max Planck Gesellschaft</td>
<td>MPG</td>
<td>DE</td>
</tr>
<tr>
<td>International Association of Scientific, Technical and Medical Publishers</td>
<td>STM</td>
<td>NL</td>
</tr>
<tr>
<td>European Space Agency, ESRIN</td>
<td>ESA</td>
<td>FR</td>
</tr>
<tr>
<td>Fernuniversität in Hagen</td>
<td>FUH</td>
<td>DE</td>
</tr>
<tr>
<td>European Organization for Nuclear Research</td>
<td>CERN</td>
<td>CH</td>
</tr>
<tr>
<td>Georg-August-Universität Gottingen Stiftung Oeffentlichen Rechts</td>
<td>UGOE</td>
<td>DE</td>
</tr>
</tbody>
</table>
Alliance for Permanent Access

The Alliance aims to develop a shared vision and framework for a sustainable organisational infrastructure for permanent access to scientific information

- The European Science Foundation
- European Space Agency
- CERN
- Helmholtz-Gemeinschaft Deutscher Forschungszentren
  - Max Planck Gesellschaft
- Science and Technology Facilities Council
  - The British Library
  - Koninklijke Bibliotheek
- Deutsche Nationalbibliotheek
- Joint Information Systems Committee
- International Association of Scientific, Technical and Medical Publishers
- Digital Preservation Coalition
  - NESTOR
- Netherlands Coalition for Digital Preservation
  - Portico

http://www.alliancepermanentaccess.org/
Connections to other EU projects

- CASPAR (http://www.casparpreserves.eu)
- Planets
- SHAMAN
PARSE.Insight aims to provide:

- **Insight and understanding** into the capabilities and practices within the various research communities
- An **inventory** of current and planned research and development relating to e-infrastructures and permanent access
- A **roadmap** for a support e-infrastructure for maintaining long-term accessibility and usability of scientific and other digital information in Europe
- Identification of **gaps** in the existing and planned infrastructure
- Progress towards a standard for **evaluating** the sustainability and trustworthiness of digital repositories
Motivation

• The problem: long-term access to the digital resources created by scientific endeavour

• Recognised importance of digital preservation
  – OAIS Reference Model
  – Multitude of initiatives, very worthwhile, but lack of “big picture”
Motivation

• Concern with data and documents

• Need for supporting e-infrastructure
  – What should this look like?
  – How can it be developed?
  – What timescale?

• The role of the Alliance for Permanent Access
Things change/disappear

- Software
- Hardware
- Environment
  - E.g. Network links to related information
- People
  - What is “common knowledge”

How can we ensure that the information trapped in the “bits” remains understandable despite all these changes?
Infrastructures for preservation

• Social / Legal / Financial / Organisational
• Agreements / Trust / Standards
• Costs/ Benefits/ Rewards
• Technical components
Lessons from other Infrastructures

• Need to “grow”, “encourage”, “foster” rather than “build”
• include organisational, financial, legal & marketing
• Provide services rather than specific technologies
• Tackle “choke points”
• Various phases of development
Encouraging Organisational and Social change

• Policies: mandates for depositing research data and funding agencies requirements:
• Robust and reliable deposit places, where researchers can be sure their data will not get lost, be corrupted or misused with correct right access mechanisms.
• Elements that increase comfort levels so that new users will know how to use and interpret the available data.
• Communication and awareness around these issues.
• Have publication of data as valued and as referencable as is a publication of a paper in a journal.
Infrastructure

• No organisation can do everything that is required for digital preservation forever
• Need to share the cost/effort
• Need to identify commonalities
  – None will be a perfect fit for all purposes
Preservation and Re-use
Unfamiliar information

• Preservation
  – Digitally encoded information which must be usable and understandable
  – Unfamiliar because of separation in time

• E-Science/GRID/CyberInfrastructure for data
  – Digitally encoded information which must be usable and understandable
  – Unfamiliar because of separation in discipline or location – even if created yesterday

Support automated usage where possible
Approach

- Approach based on evidence from community insight …
- … while taking full account of current work on digital preservation
- Coverage of disciplines: wide and deep
- Coverage of resources: data and documents
Approach

Phase 1 Preparation

Phase 2 Information gathering

Phase 3 Analysis

Phase 4 Gap

Phase 5 Impact

Milestone 1

Milestone 2

Milestone 3

Milestone 4

Milestone 5

Survey platform

Draft Roadmap

Survey structure

Carry out: Survey(s)

Case studies

Interviews

Analysis

Update roadmap

Identify gaps between “ideal” vs “actual”

Identify which gaps are most important

Approach

Survey platform

Draft Roadmap

Survey structure

Carry out: Survey(s)

Case studies

Interviews

Analysis

Update roadmap

Identify gaps between “ideal” vs “actual”

Identify which gaps are most important
The way it might have been

- Compilation of existing roadmaps
- Few hundred responses to surveys
  - A few tens of interviews
- "Mechanical processing"
- Project delivered as described but without major impact
...Changes

- Request for evidence and vision for preservation infrastructure
- Roadmap as a *Roadmap for Science Data Infrastructure*
- ..........
Achievements so far

• Successful large-scale surveys
  – Investigating attitudes and practices in digital preservation
  – Researchers, publishers, data managers, policy makers
  – Several thousand responses: a very valuable base of evidence
  – Based on threats to preservation

• Roadmap
END