Preservation at Scale

Proposal for a Workshop, iPres 2013, Lisbon

Subject: This workshop on preservation at scale will address the practical problems organizations run into when they have to deal with very large archives

Program Committee:

- Marcel Ras, Koninklijke Bibliotheek (Marcel.Ras@KB.nl)
- Sheila Morrissey, ITHAKA (sheila.morrissey@ithaka.org)
- Amy Kirchhoff, Portico, JSTOR (amy.kirchhoff@portico.org)

Length of Workshop: preferably a full day

Program:

This workshop on preservation at scale will address the practical problems organizations run into when they have to deal with very large archives. While tools, software, policies, methodologies for preservation are available and becoming "ready to use", the majority of organizations in the digital preservation space are not wrestling with the type of issues introduced by large scale preservation. The presenters in this workshop do. For example, Portico and the KB e-Depot have each archived over 20 million e-journal articles in the last decade. Throughput of content increases annually, and therefore processing power in the form of hardware and software must increase annually. Preservation at scale requires continual development of new infrastructure. In addition to preservation of digital documents and publications, data archives are collecting a vast amount of content which must be ingested, stored and preserved. Whether we have to deal with nuclear physics materials, social science datasets, audio and video content, or e-books and e-journals, the amount of data to be preserved is growing at a tremendous pace.

In the workshop we will address problems of large scale preservation from managing formats to managing incredible throughput. The workshop will consist of a mix of presentations and discussions around specific topics. The workshop will bring together a group of people having knowledge and experience and <u>doing</u> preservation at scale. The main topics are:

- Ingest, storage, replication, access, validation at scale
- Strategy and implementation of preservation at scale
- Operational issues (in central and distributed preservation)
- Development of infrastructures for preservation at scale
- Migration at scale
- Audit and certification
- Economics
- Access at scale
- Data exchange
- Collaboration

Speakers (note, we have spoken with all speakers below and they are all willing and interested in speaking):

Speaker	Topic
Maureen Pennock, Digital Preservation Manager at The British Library	Organizational perspectives on strategy and implementation
David Rosenthal, Info Systems Project Manager, LOCKSS System, Stanford University	Economics and operational issues in distributed preservation at scale
Peter Burnhill, Director of Edina	Keepers and collaboration in e-journal archiving
Lars Bjornshauge, Managing Director at DOAJ	Challenges in preserving OA content and the long tail of very small publishers
Vinay Cheruku, Director of Content Management Systems, Portico & JSTOR	Scaling up – issues and solutions
Marcel Ras, Program manager International e-Depot, Koninklijke Bibliotheek	<not yet determined>
Andrea Goethals, Manager of Digital Preservation and Repository Services, Harvard University	<not yet determined>
Tim Dilauro, IT Architect, Digital Research & Curation Center, The Johns Hopkins University	Lessons learned at the Data Conservancy and Johns Hopkins University Library
David Minor, Chronopolis Program Manager and Director of Digital Preservation Initiatives, UCSD Libraries	Ingest, validation, replication and pricing
Ross King, Chairman of the Board of the Open Planets Foundation	The SCAPE project and its results
Tobias Steinke, Deutsche Nationalbibliothek	Ingest levels for handling preservation of different object types in a large collection
Tim Gollins, Head of Digital Preservation, UK National Archives	Implementation of the New Digital Records Infrastructure that brings their Parsimonious Preservation approach into practical reality at scale