

Is the Future of Preservation Cloudy?



Schloss Dagstuhl - Leibniz Center for Informatics (German: Schloss Dagstuhl - Leibniz-Zentrum für Informatik GmbH) is the world's premier venue for informatics. World-class scientists, promising young researchers, and practitioners come together to exchange their knowledge and to discuss their research findings.

An interesting seminar is going to be held at the Conference Center in November 2012:

Is the Future of Preservation Cloudy?

The proverbial shoebox full of pictures in the attic no longer exists. Today, photographs are produced digitally and are often never printed. While for a period the shoebox full of printed photographs was replaced by a shoebox full of CDs and then DVDs, more and more these pictures are being stored in the cloud. Once in the cloud, it is anybody's guess where they are really being stored; it is clear, though, they are not in a shoebox.

Personal photographs are just the tip of the iceberg. Wherever we look, an increasing fraction of the world's data is born digital, and much of this data must be preserved and kept usable for decades or longer. Consumers are keeping a wide range of data – movies and photographs, letters, and personal records – in electronic-only form. Organizational data is also being created and maintained electronically. Health care, government, media and entertainment, and engineering and scientific data all must be preserved for periods of time that can range from decades to "forever." Just as personal photographs have moved from analog formats to digital formats stored in the home and then to digital formats stored in the cloud, this data controlled by organizations is taking the same migratory path from analog records to digital records stored in a traditional data center and now into the cloud.

Thus, these two trends – the need to preserve ever increasing amounts of data and the movement of data to clouds – are intertwined. Preservation requires that digital data continue to be useful and available when it is needed, in spite of obsolescence at any layer of the technology stack. Clouds inherently abstract away a great deal of information that is required to manage information over time. While neither preservation nor cloud computing is a solved problem, addressing issues of long-term management and obsolescence in a fully abstracted environment presents a large number of novel research questions; beyond just raising new issues to address, the elasticity and flexibility of clouds also potentially enables new approaches to preservation.

This seminar will bring together leaders from the library, preservation, provenance, and cloud communities to examine these research questions and opportunities. Specifically, we envision discussing topics such as the following:

Context Setting Issues: What types of information will be preserved in the cloud, for how long, for whom, by whom and in what type of cloud?

Technical Issues: What is the right approach to preservation in the cloud, how do we choose what information to preserve, how do we find the information, and how do we address obsolescence of a cloud provider? It will also include issues related to trust, security, privacy, provenance, and reliability over time.

New Approaches: Can the elasticity provided by clouds enable better automation of preservation actions or extracting greater value

from preserved data and can we leverage multiple clouds to increase resiliency over time?

By bringing together experts from diverse fields to discuss this intersection of preservation and cloud, we aim to:

Ensure a common understanding of preservation in the cloud.

Begin building a community focused on issues that arise in digital preservation in the cloud.

Define and document the issues and opportunities of preservation in the cloud.

Identify new research problems.

Classification

- Data Bases / Information Retrieval
- Society / Human-computer Interaction

Keywords:

- Long-term preservation,
- Cloud storage,
- Provenance,
- Obsolescence,
- Data access,
- Storage systems

Organizers:

Erik Elmroth (University of Umeå, SE)

Michael Factor (IBM - Haifa, IL)

Ethan Miller (University of California - Santa Cruz, US)

Margo Seltzer (Harvard University, US)

For support, please contact

Schloss Dagstuhl - Leibniz-Zentrum für Informatik GmbH

Oktavie-Allee

66687 Wadern

Germany

ph +49 (0) 68 71/905 0

Fax +49 (0) 68 71/905 133

lzi(at)dagstuhl.de

Website: <http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=12472>