

Tutorial proposal for iPRES 2013

TUTORIAL: Getting Started in Web Archiving and Web Archives Preservation

Description: This tutorial is proposed by the Preservation Working Group of the International Internet Preservation Consortium (IIPC).

It will present a broad overview of the issues that an institution needs to face when it starts its web archiving activity, running through all decisions: clarify objectives of web archiving activity (heritage, scientific and legal purposes); clarify legal possibilities and constraints; define collection approach (domain or selective harvest); identify curatorial, technical and organizational issues. It will especially focus on long term preservation issues (metadata, formats, ingest procedures, preservation strategies).

This tutorial will present the main standards (ISO WARC format for web archives storage and preservation; ISO standard on statistics and quality indicators for web archives), and the main tools in use (notably open source tools such as the Heritrix crawler, NetarchiveSuite and WCT curator tools, the Wayback Machine for access...). It will propose best practices and will be built on the experience of some of the main IIPC member institutions: national libraries of *Australia*, France, Germany, and Netherlands.

Intended audience / profile of participants: Librarians, archivists or other professionals who are considering doing web archiving in their organisation. Researchers who want to learn about web archiving and discover new areas for research.

Duration: One full day (perhaps possibility to distinguish between attendance to the morning session, for beginners, and the afternoon session, more dedicated on web archives long term preservation).

Program overview:

Morning (3 hours)

- web archiving objectives and (legal, ethical, technical) issues

- curatorial and organisational aspects: how to involve your curators and your management in this activity

Tutorial proposal for iPRES 2013

- presentation of (open source) tools commonly used by IIPC member institutions

- discussion on cost issues

Afternoon (3 hours)

- challenges and issues of web archives preservation: amount of resources, diversity and lack of knowledge about formats

- standards for web archives preservation (presentation of WARC format, how to use METS and PREMIS for web archives)

- ingesting web archives in digital repositories: goals, difficulties, available tools

Presenters:

Presenters come from institutions with different approaches towards web archiving: small or large scale collections, in-house or outsourced activity...; with different ways of describing, storing and preserving web archives. They are all members of the PWG and long-time practitioners in the field of digital preservation:

- **Clément Oury**, former co-leader of the IIPC Preservation Working Group, is Head of Digital Legal Deposit at the National Library of France (BnF), the service responsible for harvesting, giving access and preserving French web archives. He is Treasurer of the IIPC and serves also as convenor of the ISO working group on WARC format. He is involved in digital preservation activities since several years; he has notably been in charge of the specifications of the web archives ingest track in SPAR, BnF digital repository. Clément Oury holds a PhD in early modern history.

- **David Pearson** is currently Manager of the Digital Preservation section at the National Library of Australia. He has worked in the digital preservation field for the last 10 years and in cultural institutions for over 18 years. During this time he has project managed or worked on many aspects of digital preservation, including policy, software and tools development, system and metadata requirements, data recovery from obsolete media, and building and fitting out computer rooms. David holds an honours degree in archaeology from the Australian National University, Canberra. He has written a number of articles in academic journals on both conflict archaeological and digital preservation issues.

- **Barbara Sierman MA** is digital preservation manager at the KB National Library of the Netherlands. In this role she advises the involved KB people on matters related to digital preservation and the consequences for the day to day activities. She participates in several European projects like Driver, Planets, SCAPE and APARSEN and is involved in the creation of the ISO standard 16363 Repositories Audit and Certification. She also participates in the Preservation Working Group of the IIPC, of which the KB is a member.

Tutorial proposal for iPRES 2013

- **Barbara Signori** is Head of the Department e-Helvetica at the Schweizerische Nationalbibliothek (National Library of Switzerland).

- **Tobias Steinke**, co-leader of the Preservation Working Group, is responsible for digital preservation in the ICT department of DNB since 2003. He was project leader in the German project kopal to build up a co-operative used long-term archival system for digital documents. Tobias represents DNB in several national and European projects and is member of national committees for standards for technical approaches. Born in 1971 in Berlin, Germany, Tobias studied Computer Science at the University of Technology Darmstadt and worked for three years for a software development company.

Contact:

For any question, please contact clement.oury@bnf.fr