**JISC, the DPC and the UK Web Archiving Consortium Workshop**
Date: July, 21st, 2009
Place British Library Conference Centre in London

# From Web page storages to Living Web Archive

Thomas Risse (L3S Research Center), Julien Masanès (European Archive)

Web content plays an increasingly important role in the knowledge-based society, and the preservation and long-term accessibility of Web history has high value (e.g., for scholarly studies, market analyses, intellectual property disputes, etc.). There is strongly growing interest in its preservation by libraries and archival organizations as well as emerging industrial services. Web content characteristics (high dynamics, volatility, contributor and format variety) make adequate Web archiving a challenge.

LiWA will look beyond the pure "freezing" of Web content snapshots for a long time, transforming pure snapshot storage into a "Living" Web Archive. "Living" refers to a) long term interpretability as archives evolve, b) improved archive fidelity by filtering out irrelevant noise and c) considering a wide variety of content. LiWA will extend the current state of the art and develop the next generation of Web content capture, preservation, analysis, and enrichment services to improve fidelity, coherence, and interpretability of web archives.

By developing advanced methods to capture all types of Web content including the Hidden and Social Web content, for detecting capturing traps as well as for filtering out Web spam, the project will contribute to adequate preservation of complete and high-quality content. For ensuring the long-term usability of Web archives, new methods for dealing with issues of temporal archive construction will be developed that enable the identification and repair of temporal gaps. Furthermore LiWA will develop new methods for ensuring the accessibility, and long-term usability of Web archives by especially taking into account the evolution in terminology and conceptualization of a domain.

Two exemplary LiWA applications - focusing on audiovisual streams and social web content, respectively - will showcase the benefits of advanced Web archiving to interested stakeholders.

**Biographies**

Thomas Risse is the deputy managing director of the L3S Research Center in Hannover. He received a PhD in Computer Science from the Darmstadt University of Technology, Germany in 2006. Before he joined the L3S Research Center in 2007 he lead a research group about intelligent information environments at Fraunhofer IPSI, Darmstadt. He worked in several European and industrial projects. He was the technical director of the European funded integrated project BRICKS, which aim was to build a decentralized infrastructure for distributed digital libraries. Currently he is the coordinator of the FP7 Living Web Archive (LiWA) project.

Julien Masanès is the cofounder and Director of the European Archive, a non-profit foundation for Web preservation and digital cultural access. Before this he directed the Web Archiving Project at the Bibliothèque Nationale de France. He also actively participated in the creation of the International Internet Preservation Consortium (IIPC), which he has coordinated during the first two years. Julien Masanès is a curator and received a degree in Librarianship at enssib (Lyon) in 1999. He was a digital preservation adviser in various national and international initiatives. He also launched and presently chairs the International Web Archiving Workshop (IWAW) series.