ABSTRACT: Planning for the Lifecycle Management and Long-Term Preservation of Research Data: A Federated Approach

The "data deluge" is a recent but increasingly well-understood phenomenon of scientific and social inquiry.[1] Large-scale research instruments extend our observational power by many orders of magnitude but at the same time generate massive amounts of data. Researchers work feverishly to document and preserve changing or disappearing habitats, cultures, languages, and artifacts resulting in volumes of media in various formats. New software tools mine a growing universe of historical and modern texts and connect the dots in our semantic environment. Libraries, archives, and museums undertake digitization programs creating broad access to unique cultural heritage resources for research. Global-scale research collaborations with hundreds or thousands of participants, drive the creation of massive amounts of data, most of which cannot be recreated if lost. A recent watershed report, *Harnessing the Power of Digital Data for Science and Society* produced for the National Science and Technology Council summed up the promise and the problems we now face:

> *In principle, a digital data deluge can result in rapid progress in science through wider access and the ability to use sophisticated computational and analytical methods and technologies. In practice, the current landscape lacks a comprehensive framework for reliable digital preservation, access, and interoperability, so data are at risk.*[2]

The University of Kansas (KU) Libraries in collaboration with two partners, the Greater Western Library Alliance (GWLA) and the Great Plains Network (GPN), seek a one year (Oct 2012-Sep 2013) IMLS National Leadership Grant designed to leverage collective strengths and create a proposal for a scalable and federated approach to the lifecycle management of research data based on the needs of GPN and GWLA member institutions. Our proposal meets the IMLS strategic goal to "Practice exemplary stewardship of collections and use the power of technology to facilitate discovery of knowledge and cultural heritage." KU is a public university engaged in very high research activity and an active member of both GPN and GWLA. GPN is a consortium of universities in the Midwest that partners to facilitate the use of advanced cyberinfrastructure (network, storage, computation) for research computing. GWLA is a consortium of 31 research libraries. In building on the strength of the partners, the planning process we will focus on these three goals in service to the current and future generations of researchers and scholars:

Goal #1: Undertake an in-depth environmental scan focused on current national and international data management initiatives and on the needs of our member universities for research data management services and infrastructure.

Goal #2: Bring together a GPN and GWLA member forum and two-day workshop for the university research, library, and technology communities focused on understanding challenges and solutions in managing, sharing, and preserving research data.

Goal #3: Create and disseminate a plan for a scalable multi-institutional approach to research data management to support the university members of GPN and GWLA and advance this plan for funding.

---

[1]Lord, P., A. Macdonald, L. Lyon and D. Giarretta (2004): "From Data Deluge to Data Curation. "In *Proceedings of the UK e-science All Hands meeting 2004*, pp. 371–375

[2]"Harnessing the Power of Digital Data for Science and Society." Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council. January 2009. www.nitrd.gov/about/harnessing_power_web.pdf