

Collaborative Establishment of a Long-Term Archive for Stewardship of Interdisciplinary Scientific Data

Robert R. Downs (1), Robert S. Chen (2), Robert T. Cartolano (3), Rajendra Bose (4)

(1) rdowns@ciesin.columbia.edu, (2) bchen@ciesin.columbia.edu, Center for International Earth Science Information Network (CIESIN), Columbia University

(3) rtc@columbia.edu, (4) rbose@columbia.edu, Columbia University Libraries/Information Services

AGU 2008 Fall Meeting. Paper Number: U23A-0047 Abstract Reference Number: 6768

SEDAC LTA Web Site: <http://sedac.ciesin.columbia.edu/lta/>

Much of the scientific data that are being collected today cannot be recreated if they are not properly preserved and documented. Establishment of reliable long-term digital archives is essential to preserving these data and associated documentation beyond the working lifetimes of current scientists. Numerous challenges, both technical and institutional, need to be addressed before these data or their documentation become lost or inaccessible. Direct collaboration between university research libraries and active scientific data centers is one approach to addressing these challenges.

We report here on the collaboration between the Columbia Libraries and the Center for International Earth Science Information Network (CIESIN) to establish an interdisciplinary long-term archive (LTA) for data from the NASA Socioeconomic Data and Applications Center (SEDAC). SEDAC has been categorized as a "reference data collection" by the National Science Board (NSB, 2005). The SEDAC LTA serves as a trustworthy digital repository to support preparation, submission, appraisal, ingest, discovery, integration, and interoperability of scientific data that are expected to be of long-term interest to both natural and social scientists.

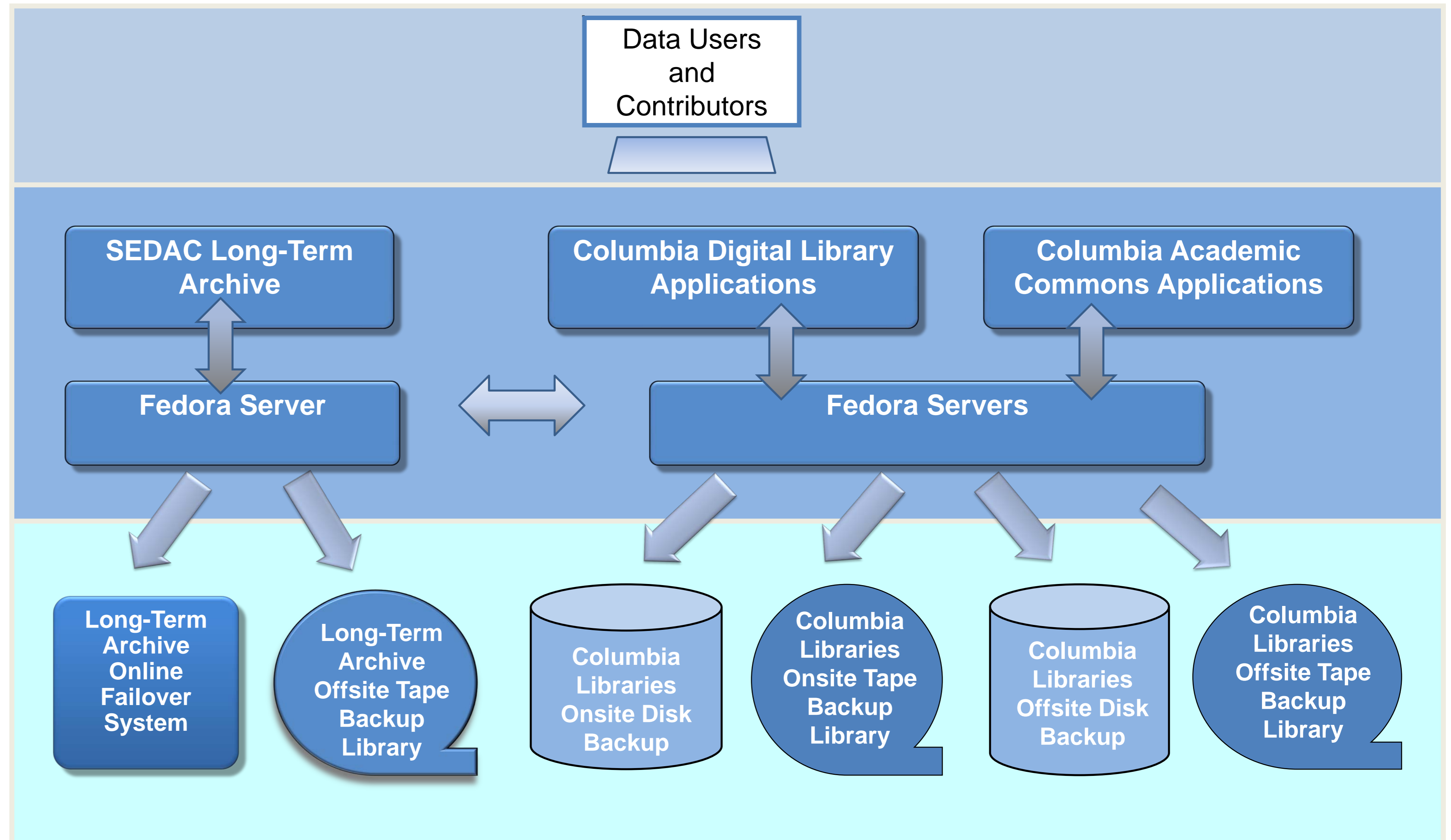
Significant progress has been made in establishing the necessary governance structure and associated policies and procedures, implementing needed standards and technologies, and assessing strengths and possible weaknesses in the long-term sustainability and usability of the digital archive (OCLC & CRL, 2007). Benefits have included sharing approaches and best practices for information technology solutions and data stewardship.

A key issue is the expected future integration of this specialized archive into the long-term digital repository currently being developed by the University. The University has more than 250 years of experience in knowledge preservation and dissemination and has the institutional longevity and commitment to ensure the long-term accessibility and usability of selected SEDAC digital data resources. Upcoming activities include testing the migration of selected data from the SEDAC long-term archive to the Libraries' repository and the development of interfaces between the digital object management systems being implemented by SEDAC and the Libraries, which are both based on the Flexible Extensible Digital Object Repository Architecture (Fedora).

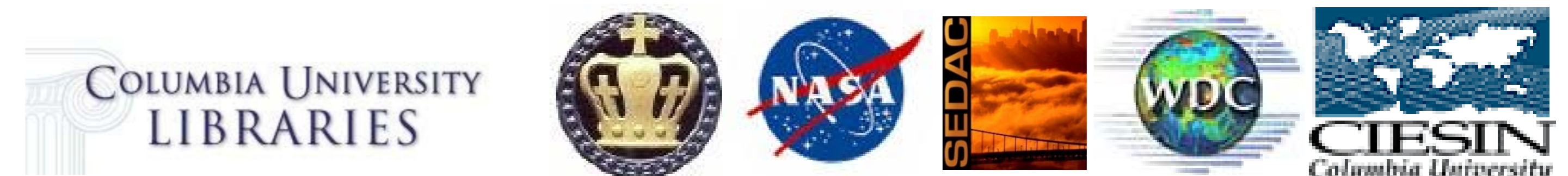
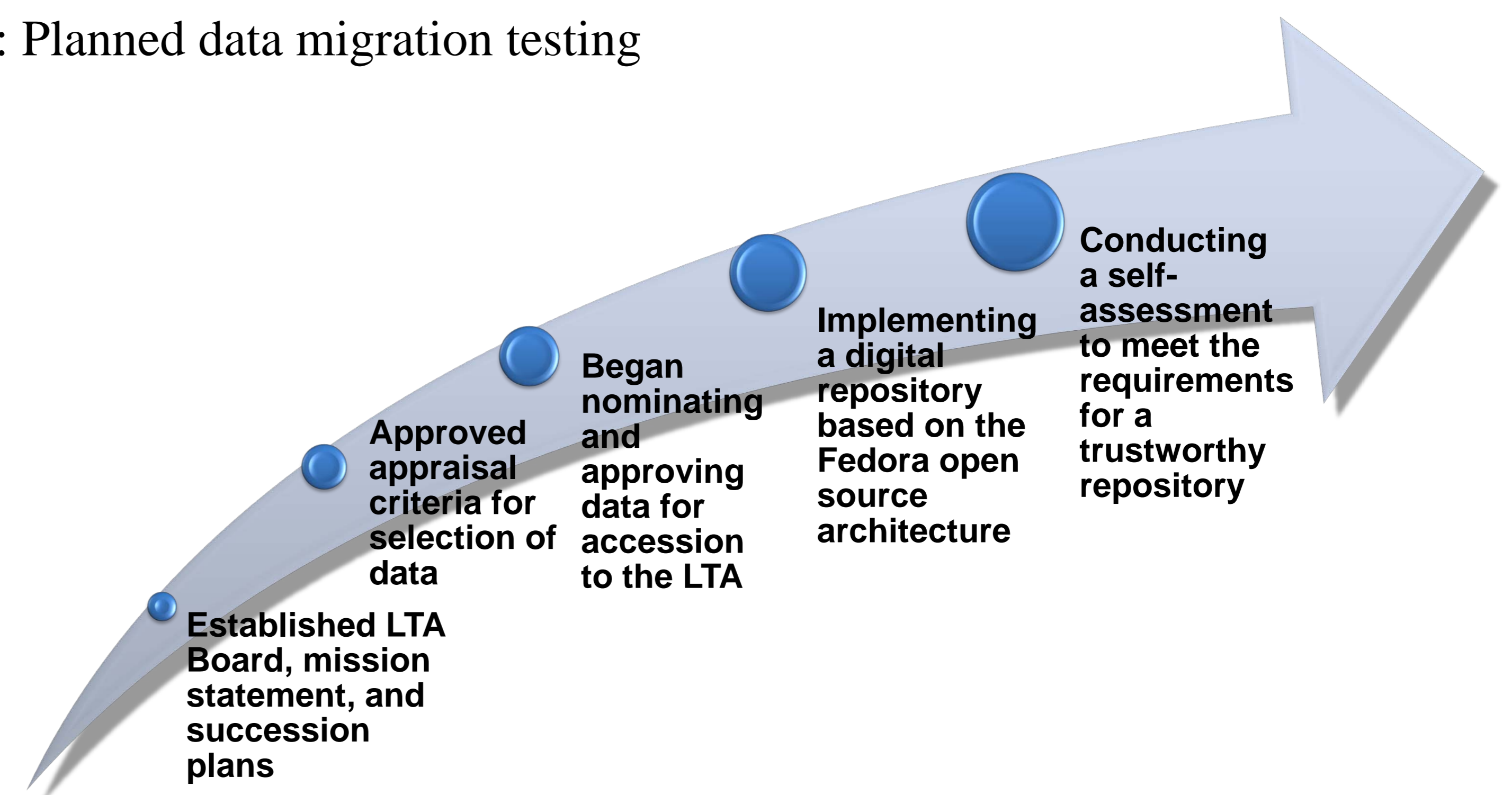
References

National Science Board. 2005. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. Washington DC: National Science Foundation.

OCLC & CRL. 2007. *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC), Version 1.0*. <http://www.crl.edu/PDF/trac.pdf>



LTA Architecture: Planned data migration testing



SEDAC is one of the Earth Science Data Centers in the NASA Earth Observing System Data and Information System (EOSDIS) and is operated by CIESIN of Columbia University under Contract NNG08HZ11C.