SEDAC Long-Term Archive Board Meeting

Friday, October 19, 2007
1:00 p.m. – 2:40 p.m.
523 Butler Library
Columbia University

Minutes

Attendance: Robert S. Chen (Proxy for Kerstin A. Lehnert), Stephen P. Davis, Robert R. Downs (Chair), Arthur Lerner-Lam, Marc Levy, David Millman, Sri Vinayagamoorthy.

Guest: Robert T. Cartolano, Director, Library Information Technology Office, Columbia University Libraries.

Agenda
The SEDAC Long-Term Archive (LTA) Board members agreed that the proposed agenda for the SEDAC LTA Board Meeting did not require any changes. The minutes of the November 29, 2006 SEDAC LTA Board Meeting were distributed and the Board acknowledged that they had been approved previously by the participants via e-mail.

Appraisal of Data Nominated for Accession to the Long-Term Archive
The Chair presented to the Board two SEDAC data sets that had been nominated for accession to the Long-Term Archive during the meeting of November 29, 2006. Each data set was presented along with a recommended level of preservation service and a recommended level of dissemination service to be provided by the Long-Term Archive. The Board members also were presented with descriptions of each data set, the Selection Criteria for Submission of SEDAC Data to the LTA, descriptions of how each candidate data set met each criterion for selection, descriptions of the levels of dissemination and preservation services currently offered by the LTA, and a list of citations to published scientific articles that reported on the use of the nominated data sets. Each of the nominated data sets was discussed by the Board in terms of the relationship between the data sets, the appraisal criteria, and the recommended levels of service.

Following these discussions, the Board approved each of the nominated data sets and the recommended levels of services. The titles of the approved data sets appear in Table 1, along with the level of preservation service and the level of dissemination service that was recommended and approved for each data set.

Titles and Service Levels of Data Sets Recommended and Approved for Accession

Title	Preservation Service	Dissemination Service
Gridded Population of the World (GPW) Version 2	Preserve and Maintain Content in Supported Formats	Public Dissemination
Gridded Population of the World, Version 2: Ancillary Data	Preserve and Maintain Content in Supported Formats	Public Dissemination

The Board also recognized that these two data sets would have to be treated as complementary, due to the relationship between the objects, and that multiple linkages would need to be established within the structure of the repository to facilitate the discovery and use of any objects that are related to any other objects in the digital repository. After discussing issues about relationships between objects in the repository, the Board recommended establishing linkages within the repository between the two archival objects that were approved during this meeting as well as between any of the objects that either are part of the same series or have been derived from the same series.

Digital Repository Status and Demonstration of LTA Collection
The Board reviewed activities that had been conducted to implement and enhance the digital repository as well as the current status of the digital repository. Several screen shots of the repository were reviewed by the Board, including the SEDAC Long-Term Archive Community web page, the SEDAC Long-Term Archive Collection search interface, the list of titles currently ingested into the SEDAC Long-Term Archive Collection, the display of a SEDAC Long-Term Archive object, and an example of a data stream from a SEDAC Long-Term Archive object.

The Board discussed the handle global persistent identifiers that are assigned automatically to each object and to each data stream upon ingest into the digital repository and that can be used in a citation to afford linking to a resource. The Board also discussed the multiple handle services that are currently being operated independently by various groups within the university and whether a plan exists for university-wide coordination of handles services and the assignment of prefixes. David Millman volunteered to inquire whether there is any current university-wide coordination of handle services and to report any findings to the Board.

Plan to Request SEDAC UWG to Deaccession Disseminated LTA Data
The Board reviewed the plan to request the SEDAC UWG to approve the deaccession of SEDAC LTA data from the SEDAC Active Archive after the SEDAC Long-Term Archive begins disseminating the data. The Board discussed whether it might be valuable to have objects stored in an additional archive that has different policies and procedures. The Board recommended internal and external testing of objects being disseminated before they are deaccessioned from the SEDAC Active Archive. In addition, the Board recommended documentation of the test plan and a review of the test plan by the LTA Board prior to the deaccession of LTA disseminated data from the SEDAC Active Archive. The Board asked for the presentation of the test plan to be included on the agenda for the next Board meeting.

Next Steps
Plans for enhancing the digital repository were discussed by the Board, including plans to ingest additional collections of information, similar to the kinds of objects that would be managed in a document management system, as collections managed by the LTA. The Board recognized that the LTA will be ingesting digital objects into the digital repository

that represent other domains of information, in addition to data and acknowledged that such collections could eventually be managed by the Columbia University Libraries.

The Board also discussed efforts to become certified as a Trusted Digital Repository and requested references for relevant models. The Chair volunteered to provide the Board with references for models of Trusted Digital Repositories. The Board recommended conducting a self-assessment of the LTA to meet the requirements for certification as a Trusted Digital Repository and agreed that a report on the status of the self-assessment should be included on the agenda for the next Board meeting.

The Board also discussed possible enhancements to the digital repository to facilitate the use of objects and data streams and requested the identification of obvious protocols needed to support the use of LTA data by the user community.

Related Presentations
The Board discussed recent presentations relevant to the activities of the Long-Term Archive that have been collaboratively co-authored by Board members. These included presentations at the International Association for Social Science Information Services & Technology (IASSIST 2007) conference, Building Global Knowledge Communities With Open Data; the Science Archives in the 21st Century conference; the International Conference on Digital Preservation (iPRES 2007); and the conference, Ensuring the Long-Term Preservation and Value Adding to Scientific and Technical Data (PV 2007).

Next Meeting
The Board recommended inviting a speaker for the next meeting and holding the meeting on the Lamont campus. The Chair volunteered to identify potential dates for the meeting and to propose the dates to the Board members, requesting their availability.

The meeting adjourned at 2:40 p.m.

SEDAC Long-Term Archive Board Meeting

Friday, October 19, 2007
1:00 p.m. – 2:30 p.m.
Butler Library 523
Columbia University

Agenda

1:00 – 1:05 p.m.   Review agenda

1:05 – 1:30 p.m.   Appraise data nominated for accession to the LTA
- Gridded Population of the World (GPW) Version 2
- Gridded Population of the World, Version 2: Ancillary Data

1:30 – 1:50 p.m.   Digital repository status and demonstration of LTA Collection

1:50 – 2:05 p.m.   Plan to request SEDAC UWG to deaccession disseminated LTA data from Active Archive

2:05 – 2:25 p.m.   Report on related presentations (IASSIST 2007, Science Archives in 21st Century, PV2007, iPRES2007)

2:25 – 2:30 p.m.   Next Meeting