

SEDAC Long-Term Archive Board Meeting

Wednesday, May 3 2006

10:00 a.m. – 12:00 p.m.

522 Butler

Columbia University

Minutes

Attendance: Robert S. Chen (Proxy for Arthur L. Lerner-Lam), Robert R. Downs, Kerstin A. Lehnert, W. Christopher Lenhardt (Proxy for Deborah Balk), David Millman, Sri Vinayagamorthy.

Review of Meeting Minutes

The Board reviewed the draft Minutes from the SEDAC Long-Term Archive Board Meeting of September, 8, 2005, and agreed to approve the minutes without amendment. The Chair will post the approved minutes on the LTA website to be publicly accessible. The Board recommended distributing the minutes soon after the meeting to those members who attended and requesting virtual approval. The Board discussed the need to meet more frequently and to have more general meetings to facilitate discussions of digital preservation issues, which have relevance to strategies and other initiatives.

Demonstration of Long-Term Archive Website

The Board briefly reviewed the Long-Term Archive web site, which has been moved to production.

Selection of SEDAC Data Nominated for Accession to the LTA

The Board discussed the seven SEDAC data sets (listed in Table 1) that had been nominated for accession during the Board Meeting of September 8, 2005, and have been accessible by the Board on the LTA website since December. The Board viewed archival information packages, including data, metadata (both DC and FGDC CSDGM compliant metadata), and permissions documents, for several nominated data sets. These data and associated documentation were displayed using the Fedora digital repository system web-based interface implemented at CIESIN.

Table 1. Titles of SEDAC Datasets Nominated for Accession to the LTA

Environmental Treaties and Resource Indicators (ENTRI) - The Update of the Treaty Status Data
HALOPH: A Data Base of Salt Tolerant Plants of the World
Environmental Subset of Collection of Multilateral Conventions at the Fletcher School of Law and Diplomacy
Freedom in the World (1995-1996)
World Resources 1996-97
World Resources 1998-99: A Guide to the Global Environment: Environmental Change and Human Health (Data Tables)
Gridded Population of the World (GPW) Version 1

The Board recommended correcting descriptions of coverage for some of the nominated data sets by replacing “world” which is an obsolete term, with “global”, the currently used term. Issues included limiting dissemination to only one instance of each data set, disseminating some metadata, and not disseminating permissions documents, and converting documents to a long term archival format, establishing workflow for automated conversion of all files matching a specific file type, and periodically reviewing supported file-types.

After reviewing the definitions of the LTA levels of service for preservation and dissemination, the Board discussed issues of preserving data in supported formats versus original formats. The Board suggested standardizing ‘item’ naming of datastreams in Fedora, and asked whether standards, such as MPEG21 DID or METS, could be used for organization and classification. Data preparation and dissemination also were discussed.

The Board discussed the procedure to evaluate the nominated data and to select a level of preservation and a level of dissemination for each data set. The Board agreed that it would be helpful to have a summary of each data set and a proposed rationale and recommendation by the LTA Manager. The Board also recognized the value of preparing notes to summarize preservation issues for each data set, along with information about usage/recovery issues which could make the data useful in the future. Such notes could be reviewed by the Board in two to five years to improve its processes.

Recommendations (listed in Table 2) for each of the nominated data sets and levels of service were moved by W. Christopher Lenhardt, seconded by Sri Vinayagamoorthy, and approved unanimously by the Board (no abstentions).

Table 2. Titles and Service Levels of Datasets Approved for Accession to the Long-Term Archive

Title	Preservation Service	Dissemination Service
Environmental Treaties and Resource Indicators (ENTRI) - The Update of the Treaty Status Data	Preserve Content in Original Formats	Public Dissemination
HALOPH: A Data Base of Salt Tolerant Plants of the World	Preserve Content in Original Formats	Public Dissemination
Environmental Subset of Collection of Multilateral Conventions at the Fletcher School of Law and Diplomacy	Preserve and Maintain Content in Supported Formats	Public Dissemination
Freedom in the World (1995-1996)	Preserve Content in Original Formats	Public Dissemination
World Resources 1996-97	Preserve Content in Original Formats	Public Dissemination
World Resources 1998-99: A Guide to the Global Environment: Environmental Change and Human Health (Data Tables)	Preserve Content in Original Formats	Public Dissemination
Gridded Population of the World (GPW) Version 1	Preserve and Maintain Content in Supported Formats	Public Dissemination

Policies and Procedures

The Board recommended replacing “media” with “system” in the description of levels of service for preservation, which the Chair will revise. The Board also suggested that the Chair should ask Stephen Davis to review the draft SEDAC Long-Term Archive Collection Management Policy before it is reviewed by the entire Board. The Board also recognized the need to document risk assessment and procedures and infrastructure for risk mitigation, such as offsite archiving, suggesting that these could be topics for discussion during the next meeting. In addition, the Board recognized the need to record usage statistics on the LTA web site and items accessed. The Board suggested conducting an annual review to determine formats to be supported.

Plans for next meeting

The Board began planning a Board Meeting to be held at CIESIN, on the Lamont Campus, that would facilitate discussions of digital preservation issues. The Board also discussed the possibility of inviting a speaker and other guests. After reviewing schedules, July 6, 2006 at 10:00 a.m. was identified as a possible meeting date.

The Board adjourned at 12:00 noon.

Acronyms and Abbreviations

CIESIN	Center for International Earth Science Information Network
DC	Dublin Core
Fedora	Federated and Extensible Digital Object Repository and Architecture
FGDC CSDGM	Federal Geographic Data Committee Content Standard for Digital Geospatial Metadata
LTA	Long-Term Archive
MPEG-21DID	Moving Pictures Experts Group Digital Item Declaration
METS	Metadata Encoding and Transmission Standard
SEDAC	Socioeconomic Data and Applications Center

SEDAC Long-Term Archive Board Meeting

Wednesday, May 3, 2006

10:00 a.m. – 12:00 noon

522 Butler Library

Columbia University

Agenda

1. Introductions and Discussion of Agenda
2. Review of Minutes from Previous Meeting
3. Demonstration of LTA Website and Revised Selection Criteria
4. Selection of SEDAC Data Nominated for Accession to the LTA
5. Status of LTA policies and procedures
6. Discussion of broader EI LTA needs
7. Discussion of relevant library digital archiving activities
8. Plans for next meeting