

Geospatial Data Provider Workshop

The University of California, Santa Barbara (UCSB) and Stanford University invite you to participate in a day-and-a-half workshop on the subject of long-term preservation of geospatial data. The workshop is tentatively scheduled to take place March 7-8 in Santa Barbara.

Background

In 2000 the Library of Congress created the National Digital Information Infrastructure and Preservation Program (NDIIPP) with the goal of finding strategies and developing technologies for preserving at-risk digital information. From the NDIIPP website (<http://www.digitalpreservation.gov/>):

Information is being produced in greater quantities and with greater frequency than at any time in history ... How will society preserve this information and make it available to future generations? How will libraries and other repositories classify this information so that their patrons can find it with the same ease that they can locate a book on a shelf?

UCSB and Stanford are one of eight teams of awardees tasked by NDIIPP to look at specific types of at-risk content. Our project, the National Geospatial Digital Archive (NGDA; <http://www.ngda.org/>), is focusing on preservation of geospatial data on a national scale and over the long term, which we define as a period of time beyond our lifetimes and beyond the lifetimes of current applications and computing platforms. Additional information on the NGDA project can be found in a preliminary project roadmap, available at <http://www.ngda.org/~gjance/ngda/roadmap.html>.

This Workshop

Given the national scale of our efforts, we believe that such preservation will necessarily be a collaborative and distributed enterprise between data providers and archive providers.

As a creator, provider, and/or aggregator of geospatial data, we believe that your content is exemplary of the kind that the Library of Congress envisions preserving. We are inviting you to this workshop, along with other such geospatial data providers, to help us gain a greater understanding of the issues involved in archiving geospatial data. We intend to use your input to begin formulating a distributed archive architecture that fulfills the requirements of long-term preservation while recognizing the realities of how geospatial data is produced and managed.

On the next page is a series of specific questions to which we seek answers. Workshop participants will be asked to deliver a 20-minute presentation addressing these questions; following the individual presentations we will facilitate a group discussion to identify common themes and issues. The results of the workshop will be written up as a project technical report and made available to the participants and to the Library of Congress.

Questions

1. Characterize the value of your content. Does the value change (increase or decrease) over time?
2. Data and metadata:
 - a. How much data do you currently have? What are the current and anticipated production rates?
 - b. How is your content organized? How are individual items identified?
 - c. Is your content subject to updates and/or revisions? If so, how are these currently handled?
 - d. How is your content typically used? Can individual data items be used when taken out of the context of your organization and out of the context of any collections or other organizational structures to which they may belong?
 - e. What knowledge is necessary to use your content? Consider both low-level knowledge (e.g., file formats) and high-level knowledge (e.g., sensor characteristics in the case of remotely-sensed imagery).
 - f. Are any applications (e.g., specialized visualization tools) or Web services (e.g., data access services) intimately associated with your content, without which the content's usability would be impaired?
3. What is your organization's workflow for producing and managing geospatial data?
4. Is your content being archived now?
 - a. If yes, how is it archived, and with what redundancy? How do you envision your archival system working cooperatively with a larger archive system?
 - b. If no, how could our archive system be most smoothly integrated into your existing workflow?
5. Are there any intellectual property rights issues regarding your content? If your content is restricted in any way, is there any mechanism or date by which it can and/or will enter the public domain?
6. Does your organization have the funding and staff resources required to archive your content in our national archive? If not, what would be required to participate?
7. Is a transition plan in place for your content in case your organization ceases to exist?