April 2009

People Profile: Julie Sweetkind-Singer

Editor

digital data. While we were not able to allocate money towards digitizing paper collections, we could archive previously scanned materials. The geospatial data and imagery we chose to collect spanned a wide array of content types and formats including scanned historical maps from the **David Rumsey Collection** and the **United States Geological Survey**, to satellite imagery such as **LANDSAT**, digital aerial photography, and data layers created to provide information about the earth's surfaces and features including elevation, ocean depths, land use, transportation, and weather, to name a few. Increasingly geospatial content is being used to inform decisions both in the private and public sector in areas ranging from population studies and census construction to land use policy and government aid determinations, and as such, it is valuable data to retain for future generations.

Data Unlike Any Other

Digital geospatial data are different from other types of data in significant ways, which affected the way we thought about and dealt with the content. First, the amount of data being created is massive. A single satellite may send down a terabyte of raw data per day. Second, the data are often released in time slices requiring decisions to be made early on as to the frequency of capture. For example, **MODIS** satellite data are constantly collected and then aggregated into 16- and 32-day composites. **MODIS** satellites capture data in 36 spectral bands, which can then be used to study large-scale changes in climate and land, ocean, and atmospheric processes. Third, proprietary software makers, such as **ESRI**, dominate the marketplace resulting in file formats that are ubiquitous and, at times, less well understood than their open source counterparts. Fourth, there are a large number of file formats, many of which require contextual information in order to be understood in the future. Finally, the data structures are often quite complex with multiple files creating a single "layer" of information, meaning they always need to travel together in order for the file to be read.

Rules of Engagement

The issues regarding massive amounts of content immediately made us realize that we would need to write Collection Development Policies (CDPs) detailing what would and would not be collected by each **NGDA** member, called a node. Choices would have to be made about what to collect and we wanted to elucidate why we were deciding one way or another. While both subject specialists, **Mary Larsgaard** at **UCSB** and I, had CDPs governing our paper map collections (with a nod toward digital materials), neither of us had written any specifically for our digital collections. With the help of **Tracey Erwin** from **Stanford**, we ended up writing three policies: an overarching policy that would apply to any node that joined in our collecting effort, and then one for each campus that was specific to that university's research needs. The CDPs

people profile

against the grain

Head Librarian, Branner Earth Sciences Library & Map Collections
397 Panama Mall, Stanford University, Stanford, CA 94305
Phone: (650) 725-1102 • Fax: (650) 725-2534
<sweetkind@stanford.edu>
http://library.stanford.edu/depts/branner/index.html

J. Sweetkind-Singer

PROFESSIONAL CAREER AND ACTIVITIES: I've been at **Stanford** since May 2000. Before that I worked for two private map collectors in the San Francisco Bay Area. Head of the Continuing Education Committee for the **Western Association of Map Libraries** (**WAML**).

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: I see libraries playing a vital role in a broad array of information and data types. Many libraries will be deeply involved in working jointly with faculty and students to manage their digital information. Libraries will continue to straddle both the paper and the digital worlds, working to redefine our roles as trusted information stewards.

include the typical topics such as collection purpose, selection criteria, and scope. They then continue with additional sections on date/chronology, formats, copyright, metadata recommendations, sources for collecting data, and a glossary.

Once we knew what we wanted to collect, we needed to ensure that if the collections were not in the public domain there was an agreement with the content provider as to the rights and responsibilities of each entity detailing how the information would be stored, used, and distributed. A Content Provider Agreement (CPA) was drafted by the relevant working group with the help from the legal staff at **Stanford** and **UCSB**. The agreement is structured in three parts. First, the main section of the agreement describes the nature of the **NGDA**, the grant of license allowing the university to hold the data/imagery, the distribution and use of the materials, and how the contract may be terminated. This section may be amended as a node sees fit to meet the needs of its specific institution. Exhibit A provides space to describe the content and any procedural matters relating to that content. Finally, Exhibit B lists in detail the authorized users and uses of the licensed materials as well as the management of the materials by the "custodians" of the content. This section of the contract is required to be a part of any agreement signed by the content owner regardless of the node in which the content is deposited. Having all of the universities (or other archiving entities) agree to the terms of Exhibit B allows us to share the data and the metadata as needed for preservation purposes. This provision also makes it clear that no matter which node originally receives the content, it will be treated in the same way.

The next step was to create a contract between the collecting institutions who agreed to participate in the **NGDA**. We worked to create a contract that does not violate any provisions of the Content Provider Agreement, allows the participating institutions to adapt to new circumstances and technologies over time, and gives the content providers a say if there were to be large-scale sweeping changes in the way we decide to do business. The decision was made to create a highly structured and yet general contract that clearly laid out the expectations and obligations for participation. We set up a governance structure, noted each member's responsibilities, laid out how to remove content from a node no longer able to host it, and specified how a node would leave the organization. The specifics for how processes would be handled are filled out in the procedure manual. This two-part structure allows us to change the procedure manual as necessary without the need to get the main agreement between the partners re-signed. For example, the main contract states that the nodes will convene "as provided in the Procedure Manual," to discuss topics such as the acquisition of new content, adding new nodes, and operating procedures. What the contract does not do is state how often this will happen, who will pay for it, who will host it, and if the meeting must be in person. All of these particulars reside in the manual, which is much easier to change. It is hoped that this structure will lessen bureaucracy and allow us to adapt quickly to changes over time.

Collaborative Collecting

Content collection began in earnest from the start of the award period. Both univer-