



Hiding in Plain Sight: Lessons Learned from Klamath Network Data Mining

Elizabeth E. Perry and Sean R. Mohren, National Park Service Klamath Inventory and Monitoring Network

Background

Most of us have done it: looking through our files and finding something interesting that we've completely forgotten about – an image, a scientific paper we meant to read, a map we could have used in other presentations, etc. Now, think of all the people who've held your position since the park you work in became a park, and imagine all the intriguing bits of information and data tucked away and forgotten about, yearning to be rediscovered and put to use...

Recognizing information's role in sound management, the NPS' Inventory and Monitoring (I&M) Program led a nationwide effort to compile data deemed crucial to parks' natural resources management. This was accomplished through 12 inventories, one of which was data mining. From 2001-2007, the Klamath Network (KLMN) undertook an extensive data mining project to document park information, in many cases helping parks discover information that had been lost to obscurity. For most parks, this was the first standardized park-specific bibliography and species list open to resource managers outside of the archives collections.

For park staff to efficiently access natural resource information over time, it needs to be accessible, in standardized digital formats, and comprehensive. This allows staff and researchers to review and build upon past studies and adapt future management needs without reinventing the wheel. The KLMN's data mining team gleaned natural resource-associated information from each park, consolidating it for park management needs and supporting I&M inventory and monitoring efforts.



Results and Discussion

As a result of our efforts, thousands of references and hundreds of datasets were documented for the parks. Data mining successfully brought each park's records current to the time of the data mining activity, bringing the parks up-to-date in their knowledge and access to critical natural resource data.

One key reason for the success of the Klamath Network's data mining effort was its flexibility to work with each park. This increased communication and tailoring to the park-specific needs helped make the project's products more useful and transparent.

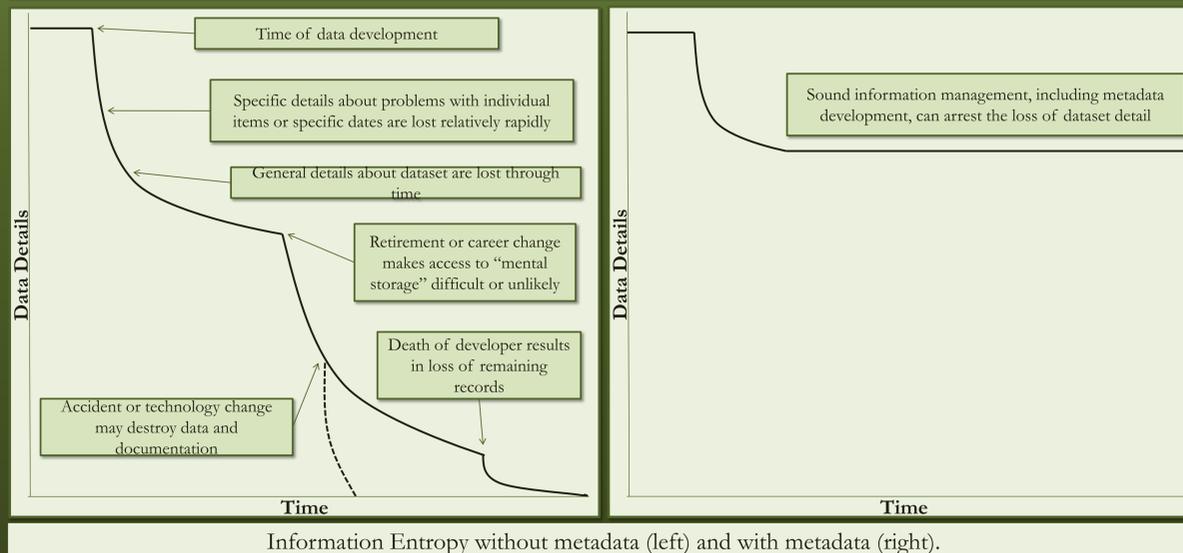
Some key "lessons learned" and recommendations to further strengthen and streamline the data mining process:

- ◆ Sufficient lead time and ongoing communication with park staff is needed before data mining starts, to organize the park's files, determine the park's data entry priorities, and schedule the effort so it does not coincide with the field season.
- ◆ Systematic consolidation, archiving, and uploading of the park's important files from all physical and digital locations would be ideal. While a complete reorganization is usually not practical/feasible, having initial rules established for efficient file recognition and access is helpful.
- ◆ Once the backlog information is entered, park personnel should implement a system for updating the databases as new information develops, ensuring the databases maintain park usefulness and relevancy.
- ◆ In-park presentations of how to utilize and update the databases is a valuable way to increase park staff's familiarity and use of the databases.
- ◆ Data mining should encompass outside facilities (universities, museums, etc.) that likely hold park information gathered by non-park personnel and researchers.

The KLMN data mining effort was hugely successful in cataloging a wealth of the information available at the six parks and in making it readily available through the I&M databases. This effort not only provided a comprehensive natural resource bibliography to support I&M and park activities but also provided a means for familiarizing park staff with pertinent park-focused natural resources information and how to practically access it.

This intensive effort was indispensable in helping the parks gain access to decades of accumulated information. With continued park engagement and partnership with the I&M Program and other research institutions, these data mining efforts will provide an invaluable base of knowledge and a data management system to support park management for future generations.

For more details on the KLMN Data Mining project and links to the reports produced, please visit: http://science.nature.nps.gov/im/units/klmn/Inventories/Basic_Inventories/INV_Bibliography.cfm



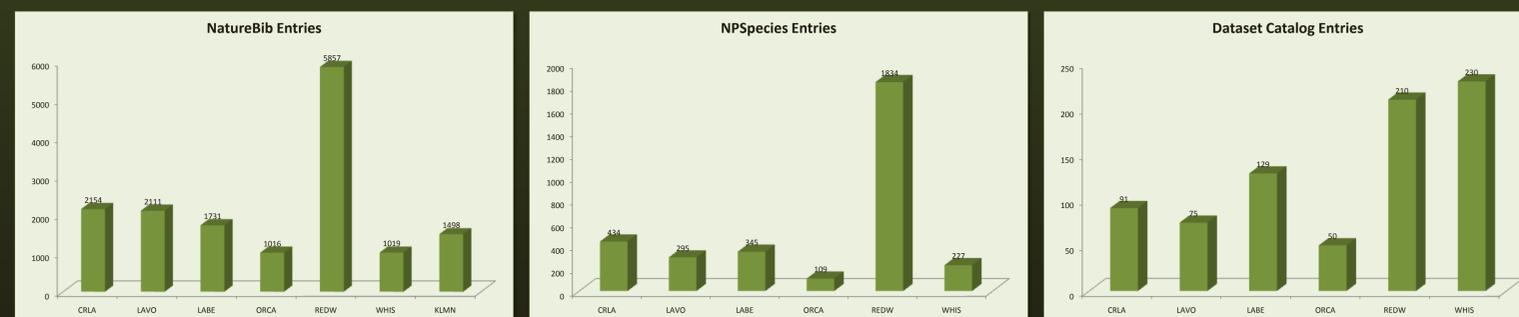
Methods

The KLMN populated NPS databases with standardized park-specific bibliographic and species list metadata for vouchers, references, and datasets in three phases, which matched the parks' priorities for data documentation:

- Phase 1. Vertebrate and vascular plant vouchers at parks, universities, and museums.
- Phase 2. Hardcopy references and datasets on vertebrates and vascular plants.
- Phase 3. Digital references and datasets on vertebrates and vascular plants, plus all digital and hardcopy references and datasets on the remaining 12 inventories.

Protocols were developed for each phase, to institute a consistent, standardized data mining effort that allowed some flexibility to meet park needs. Any park-specific deviations from the protocols were noted in the final reports. Additionally, data miners received training on the NPS databases (NPSpecies, NatureBib, and Dataset Catalog) that were used to document the information.

After discussing the park's priorities with park staff, the data miners scoured all accessible and non-sensitive files for pertinent vouchers, references, species lists, datasets, and maps. They then entered the information into the appropriate databases and kept lists of the entries. The data mining team worked extensively with each park's staff to ensure concerns were addressed and that the resulting databases would be populated with relevant and comprehensive information for the park.



Number of entries from the Klamath Network data mining effort into NatureBib, NPSpecies, and Dataset Catalog. CRLA = Crater Lake National Park, LAVO = Lassen Volcanic National Park, LAVE = Lava Beds National Monument, ORCA = Oregon Caves National Monument, REDW = Redwood National and State Parks, WHIS = Whiskeytown National Recreation Area, KLMN = Klamath Inventory and Monitoring Network.