Illinois State Geological Survey Approach to Data Preservation

Just to clarify terms…

- *A "collection" refers to both physical collections and digital collections or databases*
- *As does "data" in "data preservation"*
- *More specific:*
  - **"physical collection"**
  - **"digital dataset"**

First some overarching issues

- *Almost everybody is collecting / generating data*
- *Significant amount of existing paper and other physical collections*
  - **Many not digitally indexed or not digitally available**
- *Staff have widely varying capabilities to manage digital datasets*
- *Limited resources and infrastructure*

Components of the ISGS Approach to Data Preservation

- *…with some issues*
  (in other words, it's not perfect…yet.)
- *1. Collection relevance*
- *2. Information Management reference*
- *3. Data Management Committee*
- *4. Information Resources Database*
- *5. Library*

1. Collection relevance

- *“Universal” relevance*
  - Relevant to much or all of the Survey or deserving of centralized management
- *”Programmatic” relevance*
  - Relevant to specific Survey programs or administrative sections
- *”Individual” relevance*
  - Relevant only to a specific individual or small group of individuals

Collection relevance

- *Collection relevance level periodically reviewed by custodian and managers*
  - Issue: not fully implemented across all collections
  - Issue: not always enforced

“Universal” relevance collections

- *Dedicated staff manage these collections*
- *Example physical collections*
  - General Records Unit (GRU)
    - *State repository for borehole and well records*
    - *Holds paper copies of individual well data such as permits, drillers logs, geophysical logs, completion reports, historical records, etc.*
    - *Records on over 500,000 boreholes*

"Universal" relevance collections

- *Example physical collections*
  - Core Storage Facility
    - *Facility built in the early 1970's*
    - *Contains over 12,000 cores and over 60,000 cuttings samples*
    - *Issue: core storage has been near capacity for years*
      - Cores are being rejected or stored in inadequate temporary locations

“Universal” relevance collections

- *Physical collections*
  - Miscellaneous collections
    - *Example: Type and Figured Paleo collections*
      - Some go back over 100 years predating modern survey
      - Issue: Part-time retired staffer does bulk of curation
      - Issue: Most not digitally indexed
    - *Issue: other collections less well curated*
      - Some past attempts incomplete
      - Some collections in multiple locations under differing management (example: thin sections)

“Universal” relevance collections

- *Digital datasets*
  - Centralized and well-maintained
  - Example: Wells and Borings Database
    - *Digital database contains over 500,000 boreholes/wells*
      - Much of the paper data housed in the GRU repository has been entered
      - All incoming (new) data is entered
      - Issue: some historical paper data not entered

“Universal” relevance collections

- *Additional Digital datasets*
  - GIS Database and file repository
    - *Contains our most widely relevant GIS layers /orthoimagery*
  - Geologic Map Data Model
    - *Assists in standardizing and archiving map products*
  - Geospatial Data Clearinghouse
    - *Public gateway to much of our GIS data /orthoimagery*

“Programmatic” relevance collections

- *Physical Collections*
  - Much historical/legacy paper data
  - Issue: many not cataloged, few digitally cataloged
  - Issue: some paper data missing key indexes

“Programmatic” relevance collections

- *Digital datasets*
  - Sections and programs may devise their own databases
    - *Example: ILES interpretive log database*
      - Provides customized functionality for Quaternary and Hydrogeology sections
- *Issue: some collections need better organization and/or discoverability*

“Individual” relevance collections

- *Physical Collections*
  - Includes collections stored in individual offices and other storage areas
- *Digital datasets*
  - Includes data stored on staff computers or personal network areas
- *Reviewed upon leaving survey*
  - Issue: past collections have not been reviewed

2. Information Management Reference

- *Inform staff on data management procedures and conventions*
- *Standardizes our approach*
- *Issues:*
  - Gaining full user acceptance
  - Addressing a staff with wide range of capabilities

3. Data Management Committee

- *Handles Data Management issues and policy*
  - **Example: different policy may apply to digital versions of physical items**
    - *Distribution, licensing*
    - *Infrastructure, storage and preservation*

4. Information Resources Database

- *Contains information on our important collections*
  - For both discovery and preservation purposes
- *Each collection must have a curator*
- *Issues:*
  - Trying to get better coverage / staff buy-in
  - As curators change, details can be lost

5. Library

- *Archives specific collections*
  - Publications
  - Maps
  - Photographs
  - Field Notes
- *Issue: some procedures have changed*

(6. Work with limited resources)

- *Staffing decreases*
  - Some have learned to develop using database tools
- *Infrastructure issue/example:*
  - Can we learn from past history?
  - During the last core space crisis:
    - *core was thrown away or reduced*
    - *some of this core is now sorely missed and irreplaceable*

The role of the NGGDPP project

- *Provides some resources and particularly impetus to research and inventory collections*
- *Requires us to develop a more formal data preservation strategy*
  - **Formalize our Long Range Data Preservation Plan**

Some current issues:

- *What constitutes a "collection" for NGGDPP?*
  - How to organize "collections"
    - *Many partially overlapping "collections"*
    - *Lump or split, and to what extent?*
- *Prioritizing collections for metadata creation*
  - We have very limited resources relative to the scope of the full task

Some current issues:

- *Managing our data with NGGDPP data*
  - Internal spatially-referenced metadata system
    - *Includes data not in NGGDPP such as confidential data*
  - Metadata reporting
    - *Open format for most data: easy to supply*
    - *But likely difficult to search*

Conclusions / Summary

- *ISGS established approach to data preservation is working reasonably well*
  - (given our resource limitations)
- *A number of issues remain to be resolved*
- *NGGDPP is helping us examine our data preservation issues more closely*

Copyright Board of Trustees of the University of Illinois

SECTION S13
SECTION S14