DATA CURATION – Retaining Data over Time

While all disciplines work and share information differently, all have the same issues with long-term retention as well as ensuring the data is useful and used. There is an increase in institutional, organizational, and even governmental policy development vis-à-vis data curation, however, this will not result in sharing or even higher use of data.

Policies can help develop a structure for those who want to maintain and/or use the data, but the data has to be known, useful, and even provide added value (raw data versus processed data). Much of the present funding is for technical solutions that do not address the long-term sustainability and maintenance of the data which requires a long-term fiscal and business structure.

While there are many ways to save and share data, no single model fits all exigencies.

Instead of single repositories, multiple (different objects, things, data requiring storage) repositories are necessary; it is too complicated to accommodate differences. It is better to link different repositories than to try to develop a single repository.

- Questions to address to develop the best data curation method:
    - Metadata
        - Simple and descriptive?
        - Controlled?
        - Both?
    - # of 'items/things'
        - Purging versus need over time
            - Who decides?
            - Who does this?
            - How?
    - Record Types
        - Minimal data and very structured (NRAOPapers) versus
        - Huge data files such as those from an observing run (the very technologies that allow the manipulation and use of huge data files are also the same that put the future of this data at risk)
    - Updates
        - Continually or static?
        - Who does this?
        - Under what conditions?
    - Data Manager
        - End-user or group or organization?
        - IT versus Library versus specially organized group (such as e2e)?
    - Preservation required (more or less)
        - Is the data useful after the paper is published?
        - To whom?
        - Who decides?
    - ACCESS - Open or Closed or Limited
        - Benefit for the organization?
        - The wider audience?
    - Exposure (Level of Use?)
        - Are researchers aware of the tool?
        - Is it used? Regularly, occasionally?

- Who uses it it?
  - Ongoing curation relies on metadata that bridges the past to the present and continues into the future
    - How likely this will occur if different individuals are using uncontrolled vocabulary?
    - Scientists primary concern is the data, not the curation of the data, especially after publication of findings – devolves to the organization to retain for future needs –
    - A link between publisher (ADS as agent?), organization, and data producer (scientist)
- Data Changes over time – why?
  - New forms of data
  - Usefulness
  - New forms of publication (large data files) / Open Access

Policies can help create a framework for data curation but only in a broad sense and will only work as well as the individuals who produce or own the data want to give the policies priority.

Data curation is dependent upon those who are passionate about the retention and furnace of knowledge – who better than Librarians to be the Data Curators of the future?

+++++++++++++

BRAVE NEW WORLD – Strategies for Survival

If we agree that Librarians are the natural choice as data curators, then it follows that Data Curation can be one of the methods used to ensure the survival of the Library.

The NRAO Library has been involved in data curation over 30 years. In that time the NRAO Librarians have continued to collect, expand, add to, and update the radio astronomy bibliographic database which is of Organizational Importance as it is used not only to identify papers using NRAO data, but also track NRAO authors' publications and provide accurate statistics to NRAO's funding body.

The present version of the database – previously RAPs, now known as NRAOPapers

- Contains over 15,000 (15,450 as of this date) Bibliographic Records
- Incorporates data from the first NRAO publication in1957
- Manages the data
- Adds Value to the records (controlled metadata)
- Allows Data Sharing (use and reuse) (is available to anyone)
- Has ensured Data Preservation
- Benefits NRAO by producing statistics for NSF
- In the future will benefit the wider radio astronomy community by linking to the raw data used in the papers through the Proposal Numbers

But what else could the database do?
- What communities could use the information?

- With its basis in 'telbib' (the ESO bibliographic software developed by C. Erdmann), the NRAOPapers software could be used by other organizations to collect data
- Working relationship with ADS to provide the links from papers to data
- Other uses (what would astronomers like to have?)

But data curation only relates to current and future issues: the NRAO Library also
- Responded to 7472 internal inquiries for reference, books, and other information in 2008
- Had over 16,000 visits to the NRAO Library Web page
- Responded to almost 2000 requests from outside NRAO from scientists, students, and other Libraries were processed
- Processed 372 page charge requests

How can Libraries survive in this time of consolidation, down-sizing, and the thought (erroneous) that 'everything' is 'on the 'net'?
- Look at what the Library does
- Look at areas of demand
- Think outside the box
- Ask your customers (everyone will have a different opinion, but there will be gold among the dross)

While the NRAO Library is struggling through a tough transition, we more than doubled the level of service the Library provides in 2 years as well as streamlining, updating, and looking at what we can do, not what we are precluded from doing. We are now serving our wide-spread clientele by overnighting (over several days to Chile) books on request, filling all reference inquiries from Charlottesville, using any and all new technology. The transition continues to be rough, but our main concern is to continue the level of service we have established and maintain the unique world-class radio astronomy collection NRAO has acquired over the past 50 years.