Appraisal and Selection

By Ross Harvey, Charles Sturt University

1. Introduction

Selection and appraisal are key to ensuring that scientific data and records are usable and re-usable over time. Appraisal (a term originating in archival science) is "the process of evaluating records to determine which are to be retained as archives, which are to be kept for specified periods and which are to be destroyed".[1] Selection is a more general term, usually applied when deciding what will be added to a repository.

A popular description of appraisal as 'an evil necessity' acknowledges that bias cannot be avoided in its application. This bias, combined with our increasing ability to store and access large quantities of digital information and with its high cost, might suggest that appraisal is unnecessary. However, factors such as the exponential growth in digital data, and the current costs and limited effectiveness of solutions such as digital archaeology or reliance on information retrieval, mean that some appraisal is highly preferable.

Appraisal involves measuring the drivers for retaining a dataset or record against the costs of doing so, and determining the point at which the costs outweigh the drivers. It requires assessing the data against criteria such as:

- Does the data or record fit into a repository's selection policy? (Is there a selection policy in place at all?)
- Who will or might use the data or record in the future? (Is there a defined 'designated community'?)
- Is it economically feasible to keep the data or record? (Can we afford to do so?)
- Can acceptable legal and intellectual property rights, to keep and re-use the data, be negotiated?
- Is there a legal requirement to keep the data (and make it accessible) for a certain period of time?
- Does the data constitute the 'vital records' of a project, organisation or consortium and therefore need to be retained indefinitely?
- Is it both technically feasible and worthwhile in cost/benefit terms to preserve the data or record? (What file formats are used, for example? Is their maintenance viable?)
- Does sufficient documentation and metadata exist to explain the character, and enable the discovery of the data or record?

Different disciplines require different approaches to appraisal. Collaboratively curated databases, such as those in biomedicine and chemistry, contain source experimental data, annotations, metadata, and data extracted from other curated databases and have great potential for re-use; they are less likely to require appraisal. In other discipline-based contexts, however, appraisal is highly desirable, and these communities must determine which data or records should be maintained for use in the future, as well as any additional information that must be integrated in this process.

Appraisal criteria for specific research datasets indicate the kinds of considerations that are taken into account. The Data Preservation Alliance for the Social Sciences (DataPASS) provides appraisal guidelines for social science data.2 Key questions addressed are:

- How significant are the data for research?
- How significant is the source and scientific progress and society?
- Is the information unique?
- How usable are the data?
- What is the timeframe covered by the information?
- Are the data related to other data in the archives?
- What are the cost considerations for long-term maintenance of the data?
- What is the volume of data?

Possible retention criteria for epidemiological datasets include the nature of the questions being asked by the study; whether the question has been asked before; the richness of the data set; if it is a longitudinal study; the stability of the measures used; whether it is possible to go back to the population (e.g. for consent, ethical committee access); and its value for possible future comparisons.3

Few online tools to assist appraisal exist. One, the *Records Appraisal Tool*4 developed for use by the U.S. Geological Survey to assist in appraising collections offered to them, provides an indication of the questions that the appraisal process poses. Projects such as ECHO, PLANETS and PRESERV5 are funding the development of automated tools, but no implementable products are available yet. More work is also required to develop and test different models of appraisal that take better account of domain differences, technical issues, and cost-benefit consequences.

2. Short-term Benefits and Long-term Value

The benefits of appraisal revolve around the quality of long-term management of scientific data and records, which is directly related to the quantities managed. It is as important to determine what we want to *exclude* from our repositories as it is to decide what to include.

Short-term benefits of appraisal include:

- Better management of resource limitations (e.g. funding, skills) by reducing the quantity of data and records maintained
- Increased assurances that the collection's focus is maintained
- Better curation; for example, creation of adequate metadata for discovery and preservation is expensive
- Increased reputation, by limiting the quantity of data and records maintained and, thus, the costs of verification and other routines to demonstrate trustworthiness

Long-term value includes:

- An increased likelihood of economical long-term viability of data and records by reducing the cost of maintaining large quantities of data and records (note that costs of digital preservation are still unclear)

3. HE/FE Perspective

In the higher and further education context, appraisal, recognised as essential in the pre-digital environment, is just as relevant in the digital context, as this quote from JISC makes clear:

"Appraisal decisions are based on a number of criteria including the historical, legal, administrative, and financial value of the records. …Identifying permanently valuable records through appraisal is one of the basic aims of records management. The management and appraisal of electronic records therefore contributes to digital preservation."

— JISC Digital Preservation and Records Management Programme

4. e-Science Perspective

It is increasingly recognised in the context of e-Science that appraisal is required. John Faundeen of the U.S. Geological Survey sums it up:

"We should be expending our resources on the data we most value. Determining that value requires us to make judgments, but utilizing a repeatable and comprehensive scheme can allow us to judge data responsibly. Documenting those judgments is essential, because future generations will depend on the current scientists and records managers to preserve the data that will 'advance knowledge'"

— **Faundeen, J. L. and Oleson, L. R.** (2007). "Scientific Data Appraisals: The Value Driver for Preservation Efforts" , p.5.

5. Roles and Responsibilities

Different user groups are involved in appraisal at different levels. Data creators should ensure that the datasets they create have sufficient metadata and documentation, and use 'curation-ready' or 'preservable' formats (usually open-source) to ensure preservation and re-usability. Data curators should develop selection policies, guidelines for appraisal, and liaise with depositors to ensure datasets are in the best shape to ensure preservability when they reach the repository and with creators to ensure data is conceived in a form which facilitates its preservation. Repository managers should ensure that selection and appraisal criteria are clearly defined and publicly available, and that resources (funding, staff, technical infrastructure) are available to ensure effective implementation.

6. Additional Resources

- ERPANET/CODATA Workshop (2003). "The selection, appraisal and retention of digital scientific data: final report" ERPANET/CODATA Workshop, Biblioteca Nacional, Lisbon, 15-17 December 2003.
- Faundeen, J. L. and Oleson, L. R. (2007). "Scientific Data Appraisals: The Value Driver for Preservation Efforts" PV 2007 International Conference, 9-11 October 2007, DLR, Oberpfaffenhofen/Munich, Germany.
- Harvey, R. (2007). "Appraisal and Selection", DCC Digital Curation Manual, S. Ross, M. Day (eds).
- PARADIGM Workbook on Digital Private Papers, "04 Appraisal and disposal" .

1 Ellis, J. (1993) (ed.). "Keeping Archives" 2nd edn (Melbourne: Australian Society of Archivists) p.461. 2 http://www.icpsr.umich.edu/DATAPASS/pdf/appraisal.pdf. 3 Lord, P. and Macdonald, A. (2003). "E-science Curation Report: Data Curation for E-science in the UK" (London: Digital Archival Consultancy) p.46. 4 http://eros.usgs.gov/government/RAT/tool.php. 5 http://www.ndiipp.uiuc.edu/, http://www.planets-project.eu/, http://preserv.eprints.org/.