Digital curation: from ingest to trusted storage

Gloucestershire Archives recently (10 March) hosted a digital curation event. This was to mark the conclusion of a six month project funded by the Society of Archivists research fund to develop the SCAT (Scat is Curation And Trust) digital curation tool. It was well attended by archivists and some librarians having a specialist interest in digital archives. The SCAT development project extends an earlier digital curation project by Gloucestershire Archives funded by CyMAL: Museums Archives and Libraries Wales. The event took the form of a workshop/conference and included two guided hands-on practical sessions during which delegates undertook a variety of digital curation exercises. The conference component comprised presentations by Viv Cothey, the project archivist, to explain the SCAT approach to digital curation and especially issues related to ingest and trusted storage.

SCAT, (see figure 1) is primarily a learning and advocacy tool for archivists in local government archive services. Its aim is to facilitate development of digital curation skills and to support advocacy by offering a practical demonstration of the concepts and notions of digital curation. In particular, at Gloucestershire Archives, SCAT is used to pursue the mission of “ensuring access to information beyond its operational environment” especially when considering a 100 year or post-Microsoft point of view.

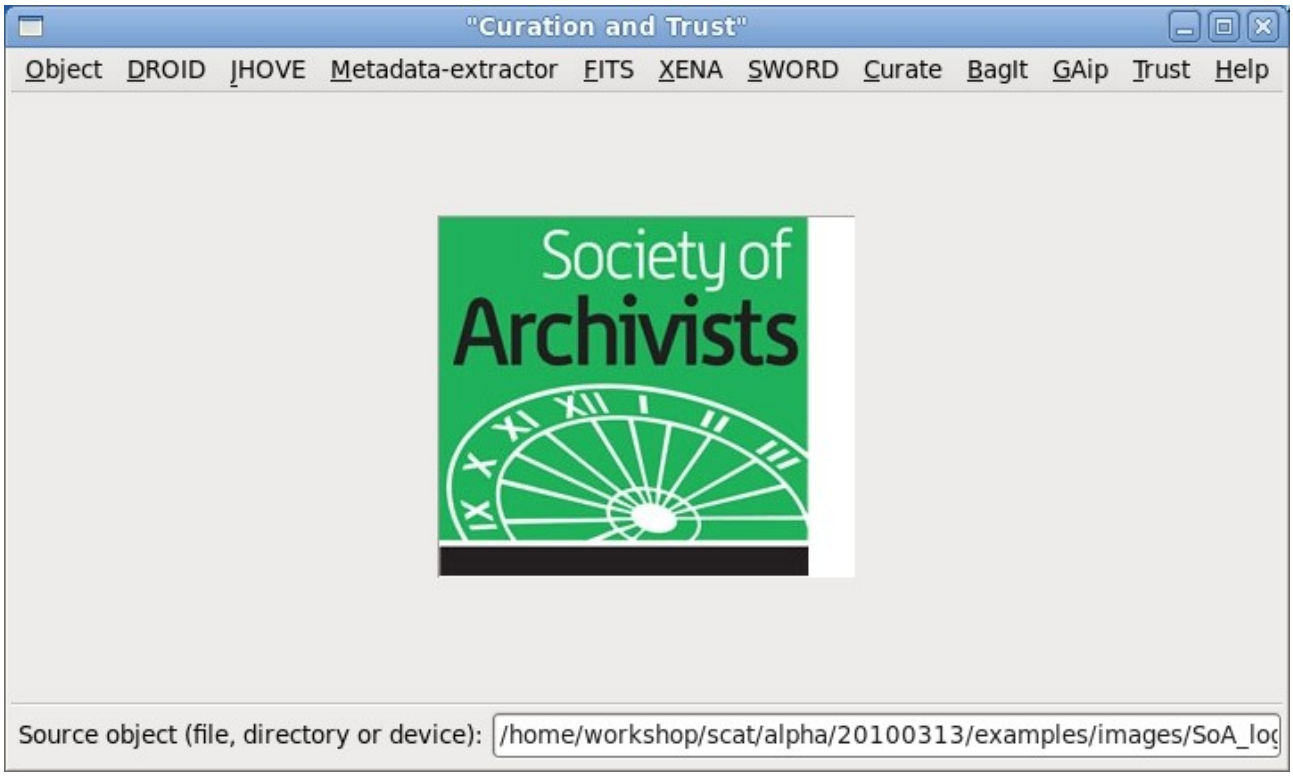

Figure 1: “Curation and Trust – principal dialog box”

Just as a goal of SCAT is *practice* based advocacy, so a goal for the event was *demonstration* based advocacy. The first of the workshop practical sessions was a series of exercises in constructing an archival information package (AIP). This made use of the “Gloucestershire Archives ingest packager” to produce a “gaip” file. A gaip file is

analogous to a physical archival storage box and contains a faithful copy of the original digital object that is being curated. This object could be a single file, a collection of files or, for example, an "image" of a an entire CD/DVD. The gaip file also contains a separate file containing curation metadata (called a "sidecar" file) and a package manifest file providing a list of gaip file contents. Delegates were able to compare the gaip file with a Library of Congress BagIt[1] file and use SCAT tools to verify that the gaip file was a valid BagIt file. The BagIt digital bag is designed to reliably convey its digital contents from place to place. The gaip file is intended to reliably convey its digital contents from one decade to another.

Local government archive services, such as Gloucestershire Archives, lack the skills and resources to operate specialist IT facilities such as those required to store digital objects. Hence we will need to enter into partnerships with competent providers of digital storage services. But how do we know that the services provided can be trusted especially when contemplating a 100 year period? The Planning Tool for Trusted Electronic Repositories[2] emphasises that compliance (to some agreed set of criteria) must be demonstrable and it is this that warrants trust. In order to establish and sustain the *trusted* storage of AIPs, SCAT provides a mechanism to regularly verify the current *message digest* fixity values of stored the AIPs against their reference value. Fixity is the quality of remaining unchanged; this is measured using message digests. A message digest, which can be represented by a string of characters, is uniquely derived from the content of a digital file. It has the property of being highly sensitive to any change in content. If there is any change in content then its value, that is the string of characters, will be different.

The second practical session included exercises to upload AIPs, that is gaip files, to a test digital repository/library which was standing in place of the facility that would be provided by a trusted storage partner. Further exercises demonstrated how the ongoing fixity of stored AIPs could be verified. An important point is that this verification is an *audit* of the operational practices of the storage partner. In all probability the trusted storage partner will have sophisticated technology that continuously monitors and manages the well-being of all the stored AIPs.

SCAT provides an archivist with a a practical set of tools that span a digital curation workflow. The software is open source and is available under the General Public Licence. The current version runs under Linux (Fedora, although other distributions should work) but it is planned to produce a version for WindowsXP. A website (<http://www.gloucestershire.gov.uk/digitalcuration>) is also being produced which will include the software.

In concluding the event, Heather Needham hoped that the participative nature of the workshop could be a catalyst to forming a participative community that would further develop the practice of digital curation.

Viv Cothey

Gloucestershire Archives

1 See <http://www.digitalpreservation.gov/partners/resources/tools/>

2 DigitalPreservationEurope, (2008), "DPE repository planning checklist and guidance DPE-D3.2". Retrieved form <http://www.digitalpreservationeurope.eu/publications/reports/Repository_Planning_Checklist_and_Guidance.pdf>