Digital Curation 101

COMMUNITY WATCH & PARTICIPATION

About Community Watch & Participation

Topics:
- Community watch and participation
- Collaboration: intrinsic to data curation
- Standards: essential for data curation
- Tools and toolkits
- The next action in the curation lifecycle

Community watch and participation

Community Watch and Participation is a full lifecycle action in the data curation lifecycle. It encompasses:

- Maintaining a watch on appropriate community activities
- Participating in the development of shared standards, tools and suitable software.

In practice this means that data curators participate in collaborative activities. Indicative activities include:

- Keeping up to date with data curation activities and developments in related areas
- Sharing data sharing and participating in other activities which form the basis of data reuse
- Participating in the development of standards for data curation
- Participating in the development of tools and toolkits for data curation.

Collaboration: intrinsic to data curation

Collaboration is one of the keys to effective curation. All communities involved in curation—data creators, users, and all the intermediaries in between—should participate in the discussions and in creating the solution.
Digital Curation 101

Collaboration is firmly embedded in data curation. Actively managing data for current and future use relies on effective data sharing, which in turn relies on agreement about and adoption of standards. Partnerships have been important in digital preservation from the start, because it was quickly realised that no single organisation could adequately archive, preserve and continue to provide access to digital material. The reasons for this include the scale of data curation issues, uncertainty about how to address them, and the high cost of data curation in relation to the resources available. Collaboration allows best use of resources by building shared resources.

The UNESCO Guidelines for the Preservation of Digital Heritage identify benefits of collaboration as:

- Access to a wider range of expertise
- Shared development costs
- Access to tools and systems that might otherwise be unavailable
- Shared learning opportunities
- Increased coverage of preserved materials
- Better planning to reduce wasted effort
- Encouragement for other influential stakeholders to take preservation seriously
- Shared influence on agreements with producers
- Shared influence on research and development of standards and practices
- Attraction of resources and other support for well-coordinated programmes at a regional, national or sectoral level

Effective collaboration requires participants to be fully aware of other activities in the data curation community. This requires that a watch on appropriate community activities is maintained on an ongoing basis. Ways to do this include regularly scanning of relevant web sites (the DCC, for example) and subscribing to relevant RSS feeds.

**Standards: essential for data curation**

Data curation is based on data being shared and reused. This implies interoperability. Standards are essential for reliable interoperability, and as the basis for building functional and safe digital curation systems. Applying standards can also provide greater economies, such as economies associated with limiting the number of file

---

1 UNESCO Guidelines for the Preservation of Digital Heritage (2003), pp.64-65
http://unesdoc.unesco.org/images/0013/001300/130071e.pdf
Digital Curation 101

formats handled. However, standards require consensus so that confusions and misunderstandings are reduced, which in turn requires awareness of relevant curation activities. This can be achieved by maintaining a community watch.

Standards should be identified (or developed if they don’t already exist) for data capture, citation, annotation, classification, technology interoperability, software integration, representation information … the list is long.

Two examples illustrate the significance of standards and of participating in standards development.

1) Much of the instrumentation used in creating experimental data is commercially manufactured. However, many instruments use manufacturer-specific data capture tools based on manufacturer data standards developed by that manufacturer. There is little uniformity, even for similar types of instrumentation. The scientific community is increasing pressure on manufacturers to conform to standards at least in the way the data collected can be exported from their software (for example, XML), so that data from a wide range of instrumentation can be combined.

2) Using standard data formats that are going to remain accessible over time is a commonly-applied digital preservation strategy. Standards that are stable and have been widely adopted are much more likely to be supported over a long period. If standards are open source (that is, not proprietary), this reduces the risk of obsolescence if there is a large user base willing to participate to ensure that the standards are maintained. XML is often adopted: it is a stable standard with a very good track record, and in widespread and increasing use.

Tools and toolkits

Current data curation processes are labour-intensive – they are often referred to as ‘artisan’ or ‘hand-crafted’. It is acknowledged that we must develop and implement ways of automating curation workflows and processes. Automated procedures need tools.

Tools are required that can be applied to many areas of digital curation – for:

- Identification of data (for example, where it is located, what formats it is in)
- Describing data (for example, automated metadata creation)
- Manipulating data (for example, data management, data storage, repositories)
Digital Curation 101

- Preserving data (for example, migration)
- Rights management and access control (for example, restricting access to bona fide users).

These tools should be *usable* and *useful* in order that they will be widely adopted.

Tool development is being actively addressed. For example, recent EU-funded digital preservation projects such as CASPAR and PLANETS aim to develop automated tools:

- CASPAR\(^2\) aims to ‘produce tools and techniques to support digital preservation and make it easier to share the cost’. These tools must be useful and usable, with characteristics including ease of use, requiring little effort to adopt, sustainable after the CASPAR project ends, and open source.

- PLANETS\(^3\) has developed prototype tools and services for preservation planning, preservation action and preservation characterisation, and will make publicly available PLATO, a toolkit of preservation planning tools, PLANETS-compliant migration tools for digital objects, emulation tools for specific environments, and characterisation tools which extract significant properties from digital objects.

These and similar projects should be monitored on an ongoing basis to see if the tools developed are applicable to local data curation activities. In addition, input into their development, where feasible, as part of collaborative approaches to data curation will ensure that the tools developed are usable and useful.

The next action in the curation lifecycle

The first sequential action in the lifecycle model is *Conceptualise* which explores best practice when conceiving and planning the creation of data.

---

\(^2\) [http://www.casparpreserves.eu/](http://www.casparpreserves.eu/)

\(^3\) [http://www.planets-project.eu/](http://www.planets-project.eu/)