Digital Curation 101

TRANSFORM

About *Transform*

Topics:
- Transform
- Migration
- Creating new results
- The lifecycle continues

Transform

Transform is the eighth and final stage of the data curation lifecycle. Its activities are:

- Create new data from the original.

Transformation of data occurs most commonly as a result of migration into a different format, or by creating a subset (by selection or query) to create newly derived results, perhaps for publication.

Migration

Migration refers to the process of transferring digital materials from one technology (hardware or software generation) to another, or from one format to another. It is further differentiated as version migration, migration to newer or standard file formats, and migration on request.

The process of migration changes data. These changes can, for example, transform the digital encoding into one suitable for preservation, or transform the original into a form that can be processed with the tools current at the time of access.

Repeated migrations, each making small changes, may cumulate into major alterations to the data – in other words, the data are transformed and become new data, with different characteristics and behaviour from the original data.

Digital Curation 101

Creating new results

New data are also produced from existing data by creating a subset of that existing data and transforming this subset to create newly derived results, perhaps for publication. The subset is created by selection of data or by querying a dataset.

In this process curated data is used in several ways:

- for verification of post-analysis results
- as the basis of further experiments
- for cumulative analysis.

This curated data, combined with data from a wide variety of new and existing sources and with the programmes used to transform the data, can be reproduced for use in the future. It becomes a new foundation for new research, science, knowledge and discovery. In this way research is reliably extended.

An example comes from a DCC case study on Integrative Biology (IB):

> Current IB users are predominantly data creators, running numerical codes to model physiology. As the IB infrastructure develops, it is envisaged that new numerical physiologists will become re-users, accessing the stored results of previous calculations for further analysis (Martin Donnelly, Victoria Boyd, and Jill Spellman, *DCC Case Study: Integrative Biology* (2008) pp.10-11[1])

The lifecycle continues

And so the curation lifecycle begins again.

---

[1] http://www.dcc.ac.uk/resource/case-studies/case_study_integrative_biology.pdf