Digital Repositories from the Perspective of a Web-based Consumer of Repository Services

Eric Kansa
UC Berkeley / OpenContext.org

Unless otherwise indicated, this work is licensed under a Creative Commons Attribution 3.0 License <http://creativecommons.org/licenses/by/3.0/>
- Not a repository, a consumer of repository services
- Presents some special issues with regard to data preservation
The Roman Cemetery at Brougham, Cumbria
Excavations 1966-67

Mark Bell and Hilary Cool, 2004

Downloads

(Data Copyright Mark Bell and Hilary Cool)

The archived database is available as a series of comma-delimited text files. To save these files, use file --> save as from your browser menu.

- Cremated_animal_bone_catalogue.csv
- Cremation_catalogue.csv
- Grave_data.csv
- Pathology_codes.csv
- Pot_catalogue.csv
- Pottery_fabric_code.csv
- Small_find.csv
- Veneer_type.csv

Documentation for the database is available from your browser menu.

- brougham_documentation.pdf

Data archived in the ADS
- HTTP URIs to identify resources at a meaningful level of granularity ("a URL per potsherd")

- Use HTTP URIs published by others

- URIs act as "primary keys" allow data to be related

- Identifiers need to last
- Citation provided for each item in Open Context.
- CDL / EZID issues ARKs and DOIs, related to Open Context Web URIs.
Use with Zotero make data citation convenient!
Data Sharing Incentives

Positive:
Recognition

Negative:
Failure to share results in loss of funding

Data Citation:
Reliable citation key incentive issue and essential for linking datasets with other publications
Who archives others on the “Web of Data”? Must consider sustainability in linking to others.
Data portability with Web standards (Atom, JSON, KML) and more specialized standards (ArchaeoML, RDF, OAI-PMH)
CDL accessing materials by reading Open Context Atom feeds
• Versioning, curation, format migration, metadata documentation (MARC, Atom, ArchaeoML)
• Institutionally to CDL, Open Context is a UC Berkeley library
• Not hugely efficient right now. Need to synch back to reconcile CDL's archival identifiers with our identifiers.
• Crawling (“pull” approach) versus a “push” approach
Future possibilities of integrating Git version control with trusted repositories?

Advantages: distributed, tools and services, community.
Some additional considerations in data curation for Websites

- Need to think about Web services (Atom, AtomPub / SWORD) to maintain data.