Key Directions for Research Libraries
Balancing Quality with Community Participation for the eHeritage Community

Jane Hunter, University of Queensland

Over the past ten years, metadata-based approaches for managing digital libraries have evolved from traditional standardized metadata schemas, to semantic annotation systems, to machine learning techniques that automatically extract metadata and classify resources. In recent years, there has also been an explosion of social tagging sites such as Flickr, Del.icio.us, Connotea and LibraryThing that provide a community-driven, "organic" approach to classifying information and resources on the Web.

Concurrently there has also been an explosion in the number of projects that involve volunteers and the wider community collecting data for scientific analysis and uploading it to a shared online database – so-called citizen science (e.g., GalaxyZoo, ReefCheck, WaterWatch)

Community Challenges

Many research libraries are keen to explore how they might leverage the current enthusiasm for community participation in the generation of tags, metadata and data – in order to enrich their collections without compromising the quality of the content and metadata.

The challenges include:

- **Interoperability of annotation and tagging systems** – standardized models, schemas and ontologies for defining and exposing tags and annotations are required to enable the aggregation of tags or annotations across applications, communities or web sites.
- **Improving the quality of community-generated data and metadata** without destroying communities' enthusiasm for tagging or compromising the ease and simplicity by which the data/metadata is generated. Approaches include: suggestive tagging; post-processing; registered users; and editorial curation; will be discussed.
- **Managing and adapting to changing terminologies**, folksononomies and ontologies that evolve over time, to ensure maximum relevance to communities.
- **Semantic annotation** of dynamically generated and constantly changing documents (e.g., wikis, blogs, podcasts, sensor data)
- **Hybrid classification and search systems** that combine community tagging approaches with traditional library cataloguing and machine learning approaches. One method is to merge and correlate the different types of metadata through a

common ontology, and to also apply weightings that reflect the reliability and accuracy of the source.

About the Author

Jane Hunter is Professor of eResearch at the School of Information Technology and Electrical engineering at the [University of Queensland](). She leads a research group specializing in the development and application of innovative semantic web technologies to the analysis and management of mixed-media scientific data collections. She is currently a member of the National eResearch Architecture Taskforce (NEAT) and the Academy of Sciences Committee for Data in Science. She has published over 80 peer-reviewed papers on the semantic web, digital libraries and data management and is currently a CI on projects associated with water information management, marine sciences, protein crystallography and cultural heritage preservation.

Jane is also a member of the DL.org [Liaison Group]().