Digital Curation Centre research agenda

Michael Day
Digital Curation Centre
UKOLN, University of Bath
http://www.ukoln.ac.uk/

**The Digital Curation Centre: a NIEeS community awareness day,
Centre for Mathematical Sciences, Cambridge, UK, 16 June 2005**

DCC research

– The DCC research team

  • Led by Professor Peter Buneman (School of Informatics, University of Edinburgh)

  • Distributed throughout all four DCC partner organisations

  • Strong links with other DCC components, through multi-team working, etc.

– Links with other research groups

  • Visitors programme

Research goals (1)

– To draw together the various functions of curation, from the traditional archival functions to the maintenance and publication of evolving knowledge as seen in scientific databases

– To conduct research in areas already identified by the partners as crucial to digital curation

UKOLN

D|C|C

Research goals (2)

- To identify through direct research collaboration, and through interaction with the service arm of DCC, the key projects in which research is needed

- To institute two-way conduits between research and service in which practical issues can be drawn to the attention of researchers and the products of research can be tested in practice

Current priorities (1)

– Data integration and publication
  - Review of techniques
  - Publishing data that conforms to a given format or schema

– Performance and optimisation
  - Safe data analysis environments within data centres
    – Initial testbed based on sky survey databases (in collaboration with the Wide Field Astronomy Unit and AstroGrid)

Current priorities (2)

– Performance and optimisation (continued)
  - Automated metadata extraction and generation
    – Essential for testing the scalability of metadata-based preservation strategies
    – Review of tools, assessment of text mining techniques

– Annotation
  - Survey of the forms of annotation
  - Annotation and provenance
    – A model for data transformations that maintains annotation and provenance

Current priorities (3)

– Appraisal and long-term preservation

- Appraisal techniques
  - Investigating the applicability and scalability of traditional appraisal techniques in 'data-intensive' contexts
  - Dynamic databases
  - Preservation techniques for evolving metadata and databases

Current priorities (4)

- Socio-economic and legal contexts
  - Networks of trusted repositories
    - Varying preservation role for repositories
    - Roles for co-operation, exchange formats, replication, etc.
  - Economic cost-benefit analysis of curation processes
    - Quantifying costs and benefits
    - Testing economic viability of curation processes

Current priorities (5)

– Socio-economic and legal contexts (continued)

- Rights and responsibilities
    - The legal contexts of curation, e.g. impacts of the Database Directive
    - Complexity of rights held in databases, impacts on aggregation and reuse of data

UKOLN

D|C|C

Further information

- – Digital Curation Centre Web site:
  http://www.dcc.ac.uk/
- – Contact: research@dcc.ac.uk