dart
dataset acquisition
accessibility & annotation
e-research technologies

CR3
Science Commons

- Shan Chen
- DART-UQ

Agenda

- Introduction to CR3
- Introduction to Science Commons
- Issues and Challenges
- What I will focus on
- Expected Outcomes
- Future work

DART Workpackage CR3

"Reduce barriers to content acquisition by providing more rights options for science researchers."

- Apply Science Commons[1] work to scientific research data and results
- Develop software that enables the attachment of standardized licenses

Science Commons I

- Sciences critically depend access to and use of scientific data
- The growing problem of protectionism – inflexibility of copyright and patents
- *'some rights reserved' – legal innovations with technical mechanisms*
  - allow knowledge sharing
  - encourage scholarly communication
  - facilitate research collaboration
- Science Commons extends Creative Commons[2] into Scientific Output

Science Commons II

- Publishing
  - scientific literature, publication and archiving
  - Open Access Business Model (OABM)
    - open access publishing
    - retain sufficient rights for researchers
- Data
  - explores ways to provide broader access to scientific data
  - whilst protecting creator's rights
- Licensing
  - extension of CC Licensing model
  - two key extensions:
    - peer-reviewed publication domain
    - patents and intellectual property

Issues/Challenges - Publishing

- Open Access
  - self-archiving e.g. ccPublisher/IA[3,4], ePrintsUQ[5]
  - searching, reading, copying, printing, metadata harvesting (OAI) etc
  - what kinds of access are available to the public?
  - articles could contain text, diagrams, photographs, video clips, as well as links to research data/software prototypes and to cited works
- Legal implications of OABM
  - lack of verification in the existing open archival systems, e.g.
    - check author is allowed to upload to repository
    - check action against publisher's copyright agreement

Issues/Challenges - Data

- Different scientific data formats
  - e.g., focusing on scientific data formats e.g., CIF (Crystallographic Information File), PDB (Protein data bank), FITS (Astronomical)
  - Database columns
  - Composite objects (SMIL)
- Explore ways to access data efficiently, without relying on a centralized database
  - e.g., using Web services

Issues/Challenges - Licensing

- SC license metadata issues
  - How do SC licenses differ from CC licenses?
  - Different file formats, e.g.,
    - SMIL file – different licenses to different objects in the file
    - Database – different licenses to different columns
- Legal issues
  - which existing agreements are good targets for standardization
    - e.g. MTA[7]
  - licensing for multiple copyright holders

Additional Issues/Challenges

- Workflows
  - steps to guide the creation of licenses
- Search Engine
  - basic SC licensed works searching
  - advanced searching
    - search for SC licensed works
      - e.g., "give me all the audio files with licenses like this"
    - search for works correctly/incorrectly use SC licensed works
    - search multiple licenses in a composite object

What I will focus on

- Extending CC license software for scientific data types e.g., real-time streams, visualization of data, database tables
- Use SMIL files from DMQ4 – composite objects that combine video, images and real-time data streams
  - enable different licenses to be attached to different components
- Tool that checks when author uploads object to a repository that copyright doesn't conflict with publisher's rights (SHERPA/ROMEO)

Expected Outcomes

- Software tools
  - Extensions to CC tools for scientific publications
  - Copyright checking tools
- Documentation
  - SC specifications
  - SC workflows
- Publication(s)

Next 6 months

- Identify user groups
- Collect experimental data (SMIL files, database)
- Ongoing literature review (SC, SMIL, existing tools/techniques)
- Identify software development issues
- SC Tools development and documentations
- Reports/Research papers

dart dataset acquisition accessibility & annotation e-research technologies

References

- [1] http://sciencecommons.org/
- [2] http://creativecommons.org/
- [3] http://wiki.creativecommons.org/CcPublisher_2_Beta_2
- [4] http://www.archive.org/
- [5] http://eprint.uq.edu.au/
- [6] http://www.sherpa.ac.uk/romeo.php
- [7] http://www.spo.berkeley.edu/guide/mtaquick.html

dart dataset acquisition accessibility & annotation e-research technologies

Thank You

Any questions?

schen@itee.uq.edu.au