INTEGRATED DATA AND RECOVERY:
THE UC CEISMIC FEDERATED ARCHIVE

James Smithies
Senior Lecturer in Digital Humanities
University of Canterbury
New Zealand
Strategic Principles

- Leverage existing national infrastructure.
- Federate.
- Iterate.
- Build an ecosystem.
- Engage with the community.
- Value and enable research.
- Build an on-ramp to HPC and Data Fabric.
- Embrace heterogeneity:
  - of ontologies, schemas, systems, data;
  - of copyright;
  - of communities.
- View problems as research problems.
- Be in it for the long-haul.
Ecosystem
The Numbers

Public or in development: 84, 593.

UC CEISMIC:

- Total items available to public: 54,943
- Images: approximately 39,000
- Video: approximately 1,000
- Audio: approximately 200
- Newspapers: approximately 10,000 pages (156 newspapers)

UC QuakeStudies:

- Total parts (public and non-public): ~ 74,000
  - Public: 34,914
  - Researcher only: 4974
  - Embargoed: 590
  - In development: 24,015
  - Creative Commons Licence: 24,237
- Formats:
  - Image: 46,999
  - Text: 14,755
  - Video: 938
  - Audio: 615
  - Other: 1186
Constraints

• The problem of scale:
  • Governance;
  • Political will;
  • Ownership.

• The problem of complexity:
  • Forget the systems; good engineers can deal with that;
  • Governance / project definition / funding / initiation.

• The problem of time:
  • We want to be here forever; key agencies aren’t;
  • Knowledge about the data, and where it is, is as important as the data itself.

• Funding:
  • UC CEISMIC in 2015, 2016, 2020?