

data from an archival perspective

- creates data about holdings
- data held on excel spreadsheets and standard doc
- now being put into ica format
- australian dictionary f biography deposited --- NLA????
- ocr of typed lists
- links to images in digital collections
- sharing with trove to provide search nationally
- images held in digital collections
- hold researchers lab books and field notebooks all sorts of material
- also processed data, computer printouts, card indexes etc

data management in big science

- oceanographic data citation
- good quality observational data
- no credit for citing and re use
- justification of major expense on observational projects
- score - scientific committee for ocean research

- automated workflows for publishing data and associated meta data
- assert provenance
- need to also show provenance of theories and how each is built up on others
- ??relation to dublin core working party on provenance??

- metadata can later be expanded about provenance
- woods hole looked at building provenance bundle and relating papers and research data and providing a single citation for item

- internet is the only scale free environment

- found needed to extend dublin core to woods hole core
- metadata needs change to use in different ways
- tension between flexibility and interoperability

- versioning of data sets

- need to provide point in time sets as longitudinal studies subject to change and update

- important to ease author burden - unrealistic to get researchers to do this
- author concerns on reuse and reuse of data
- data held in proprietary instrument specific file types

- data has a life cycle from information through publication to knowledge

data ethics

- data licensing should be simple but is made complex by institutional complexity
- human data has ethical and cultural constraints

- dissemination versus publication - publication make data explicitly available, dissemination more peer to peer sharing

- need to involve ethics committees with question of sharing

- data sharing myths

- data cannot be destroyed
- data must be destroyed later time
- ethics clearance cannot be amended once issued

in practice most ethics committee have no expertise on data sharing and tend to tie themselves in knots

- Victoria University have turned data deposit question around to ask if there is NO reason to share data with assumption of sharing
- important researchers discuss sharing with ethics and research committee
- issue of informed consent
- excessive licensing is a barrier to rescue
- need a set of standard licences
- Ausgoal / creative commons/cc0

information literacy program

- workshop for research students to cover data management
- publish handbook on rdm
- general outreach activities
- teaches how to produce data management and curation plan
- course and handbook originally funded by apsr
- courses tend to small <12 participants

Australian data archive

- started as social science data archive
- can recover 9 track and punch cards
- nodes in QLD, UWA, UniMelb and UTS
- focus on curation
- nci provide infrastructure
- approx. 2500 data sets - social sciences
- idea to have lenses - i.e. access to comprehensive sets of related data to provide a structured search and access
- e.g. ATSIDA at UTS hold aboriginal data
- provide single point of access
- provide set of special visualisation tools etc
- use OAIS framework and have described workflow
- have online deposit tool that solicits information to build metadata
- creates provenance statement
- creates a preservation copy
- labour intensive as have curators to QA
- do normalisation to preservation formats
- do some data cleaning as part of QA
- use nstar to hold data and export in a range of formats such as sass stata etc
- end result is catalogue records with added features -description and provenance and also on the variables covered
- attempts to cover related material
- provides restricted access to data if requested by researcher
- trying to push metadata collection as far back in process as possible

NCEPH

- 80-100 researchers in epidemiology and population health
- obsession with data analysis software
- literate programming approach
- idea is to document script and functions to manage process
- big data not just about size but complexity

- principles: sustainable, secure, flexible

- researchers care about publication not data ingest or curation
- data sharing not on agenda due to emphasis on publication
- publications means points and points mean prizes
- data sharing has issues including reuse for poor or unethical research and imputation of poor provision to original data provider

- nceph did not employ an archivist - support of research key driver

- need to build a culture round reuse

digital humanities hub

- cross discipline intersection between humanities and computing
- archaeology and anthropology produce data
- data can be in a range of file formats including sound and moving image with transcription
- multimedia and set of related files
- whole can be greater than the sum of its parts

- occams can create a metadata record for discovery
- has workspace for file analysis and annotation
- has a range of predefined record types
- each record type has specific metadata definition
- definitions expressed in dublin core

- austlang collaboration work with aiatsis
- focused on textual data to do with linguistics
- has gis data associated with recordings and language use

- can demonstrate gis connectivity in occams as well
- also can use annotation documents
- complex architecture/graph of representation of knowledge
- can use annotation to record discussion

- ipad based field tool for use to add information - e.g. from discussion
- can then upload files and data straight into a project in occams
- automatically geotagged images of value
- useful to show location of findspots in archaeology and distance to other points of interests

- also to overlay historical maps on current maps

College of Asia pacific Photographic collection

- basically cartographers moving into data management
- now wish to put historic map data online for reuse
 - cf digital humanities

- moving from cartography to digital bureau

- progress to digitise old paper maps not previously cataloged
- supporting new online delivery methods
- building a set of gis metadata
- have a scanning project to get casuals to scan documents with no formal project plan
- originally started with filemaker to build database
- has advantage of xml export
- basically a simple repository/digital asset management system
- plan to enhance to make a jpeg and eps file available to university and the world
- online resource maintained as a college system
-
- question of copyright for digital versions of printed map sources
- consideration of a youtube style takedown system
- need to control access to reduce workload and make service self administering
- could eventually provide a higher res in house alternative to google map
- need to consider geocoding standards to ensure open and reusable

Data Citation

- citation cannot happen unless curation findability and discoverability is in place
- same issue as used to be with journals
- same citation problem faces wikipedia
- open data becoming important
- RCUK to require data citation from 01/04/2013
- commercial data citation metrics tools coming
- google scholar investigating as do Thompson reuters
- open and citable dat adds values to links to papers
- data being seen increasingly as a scholarly output
- citation allows tracking of reuse
- need to build a culture of data citation
- 40M doi s produced for datasets
- RDA provides a portal to institutional repositories
- data needs to be persistent to be cited
- not yet there - ROI lag due to changes in practice
- ands doi datcite service