



Archival principles and data sustainability

Dr Andrew Wilson
Queensland State Archives

1



Session Topics

- Drivers for sustaining data
- What is sustainability
- Archival theory
- Archival principles
 - appraisal
 - arrangement & description
 - long-term storage
- Context
- Authenticity



Drivers

- Research is increasingly data driven and data intensive
- Making research data more widely accessible is a ‘Good Thing’®
- Governments around the world want data made available (usually at no cost)
- Data is no use to anyone unless it is actively managed over time



Why active management?

- Media wear out
- Software becomes obsolete
- Researchers move on and/or retire and are no longer available
- Long-term accessibility is not about technology (although it's a necessary component)



What is Sustainability

- Another term for Data Curation
- Management and preservation of digital data over the long-term
- Something archivists have been thinking about & doing for decades



The study of archives

- A scholarly discipline as well as a practice (profession)
- Archival theory + methodology + practice = archival science



The discipline

- Archives have existed for thousands of years – almost as long as humans have been writing
- Archival theory codified between about 1830 and 1950s
- Influence of diplomatics
 - study of documents and their forms to establish authenticity



Core tenets of archival theory

- *Respect des fonds*
 - Records should be grouped according to the nature of the body that accumulated them
- The Principle of Provenance
 - Records of the same creator should not be mixed with records of a different creator
- The Sanctity of Original Order
 - Records should be maintained in the original order in which they were created and kept



The archival mission

To manage business/jurisdictional regimes to ensure the capture of essential recorded evidence in the course of personal and work activity, and to ensure **its authenticity and accessibility over time**

[adapted from McKemmish, S. et al., *Archives: Recordkeeping in Society*, Wagga Wagga, 2005, p. 63]

9



Archival Practice

- Appraisal
- Accessioning
- Arrangement & Description
- Storage
- Access



Archival Practice

- Appraisal
- Accessioning
- **Arrangement & Description**
- **Long-term Storage**
- Access



Appraisal

- *“Appraisal is the noblest function, the central core of contemporary archival practice.”*
- Generally referred to as selection outside the archival community

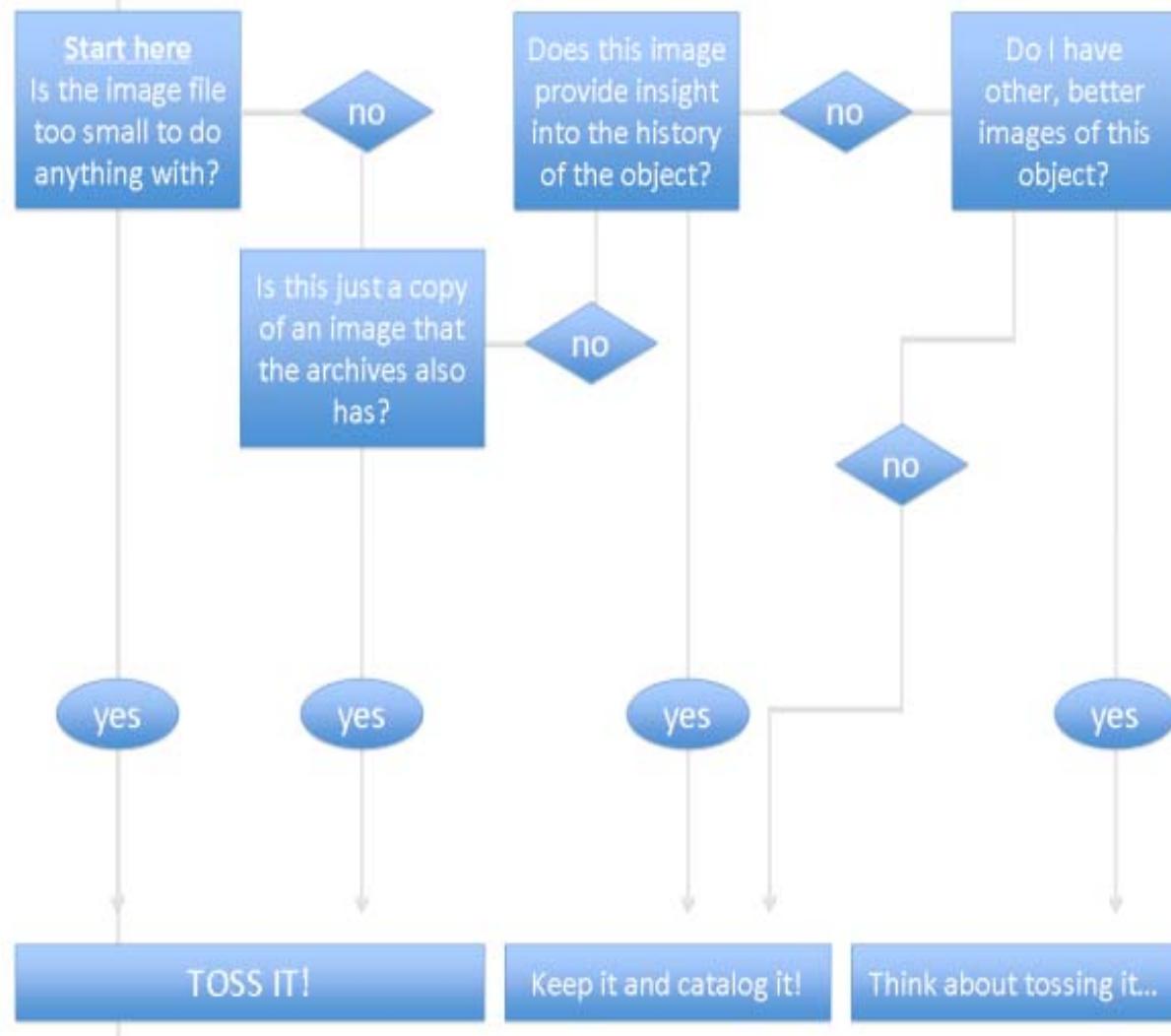


Selection

- Can we keep everything?
 - It is technologically possible at the moment
 - Other problems if we do?
- Should we?
 - Why?
 - Why not?
Digital Information created in 2010 = 72 stacks of books.
 - If not, how do we choose?

Data Selection

- Who?
- Criteria?
- When?



14



Arrangement & Description

- The process of organising materials with respect to their provenance and original order, to protect their context and to achieve physical or intellectual control over the materials



Conceptual drivers for A&D

- Inter-relatedness of records
- Central importance of context
- Function of records as evidence



Conceptual drivers for A&D

- Inter-relatedness of records
- **Central importance of context**
- Function of records as evidence



Context

- The knowledge necessary to sustain a record's meaning over time. Describes 'who', 'what', 'when', 'where', and 'why' of records creation and management



Long-term Storage

- Since archives exist to preserve records forever, archivists have been focussed on the issues of preserving digital objects
- Archives community has been involved in digital preservation projects since the early 1990s



Digital Preservation

- fundamental challenge is to preserve the accessibility and authenticity of digital objects over time and domains, and across changing technical environments
- we must accept the
 - inevitability of change
 - necessity of separation of logical information object from its physical environment

20



Essence/significant properties

- everything that is needed to ensure the object is usable and understandable over time
- not everything that the original object is constructed with
- different rules for different genres and contexts
- work is needed to attempt to establish the ‘significant properties’ of files in a wide range of digital data formats



Critical Characteristics

- Reliability) Identity
- Usability)
- Integrity
- Authenticity
 - [AS ISO 15489 *Information and documentation – Records management*]
- Protecting the authenticity of a record thus implies the need to preserve its identity and integrity over time

22



Reliability (part of Identity)

- A resource is reliable if:
 - it provides the same information over time
 - its operational context is visible
 - it is persistent over time



Usability (part of Identity)

- A resource is usable if it can be:
 - located
 - retrieved
 - rendered
 - understood



Integrity

- A resource has *integrity* when it is complete and uncorrupted **in all its essential respects**
- In practice this means that there should be no undocumented changes to the resource
 - malicious alteration (deliberate)
 - stray gamma rays (non-deliberate)



Authenticity

- A resource is authentic if it:
 - is what it says it is
 - was created or sent by the person who said they or sent it
 - was created or sent at the time it was said to have been created or sent

[ISO 15489 *Information and documentation — Records management*]



Coping with authenticity

- Involves establishing identity and integrity
- Identity: attributes such as context and provenance which uniquely distinguish an object from other objects
- Integrity: wholeness and soundness, ie. It remains complete and uncorrupted
- Any environment involved in the preservation process has to cope with authenticity issues



The Issue

- “People need to be trained to save data, funded to do it, and rewarded for it”
 - Michael Lesk “Public and Permanent Scientific Data”, Sun Preservation and Archiving Special Interest Group (PASIG) Conference 2009



Thanks for listening



Any questions?

29