UNIVERSITY
of
OTAGO

SAPERE AUDE

Te Whare Wānanga o Otāgo

NEW ZEALAND

Enriching books and printed theses with linked digital artefacts

David Eyers, John Egenes, Howard Amos, Allison Brown, Thelma Fisher

Motivation

- Some doctoral theses & research documents need to link to multimedia data
  - e.g., research artefacts and other complex digital objects
  - Links should be usable even when the document is in paper form

- Sharing of linked artefacts will be increasingly important
  - Increase value of research and its impact
  - University of Otago Library see cases of higher download statistics for crucial supporting artefacts than written documents themselves

- We would like to best support this process, and to share experiences, problems and solutions with others

Developing a general solution

- Technically, straightforward to support a given example
  - However, this risks building ad hoc solutions

- Long-term support requires careful planning/management
- Track (inter)national initiatives and services for RDM

- Overall process from the content authors' perspective:
  - upload digital artefact somewhere
  - gain Internet-accessible identifier for it (e.g., URI)
  - potentially encode this identifier
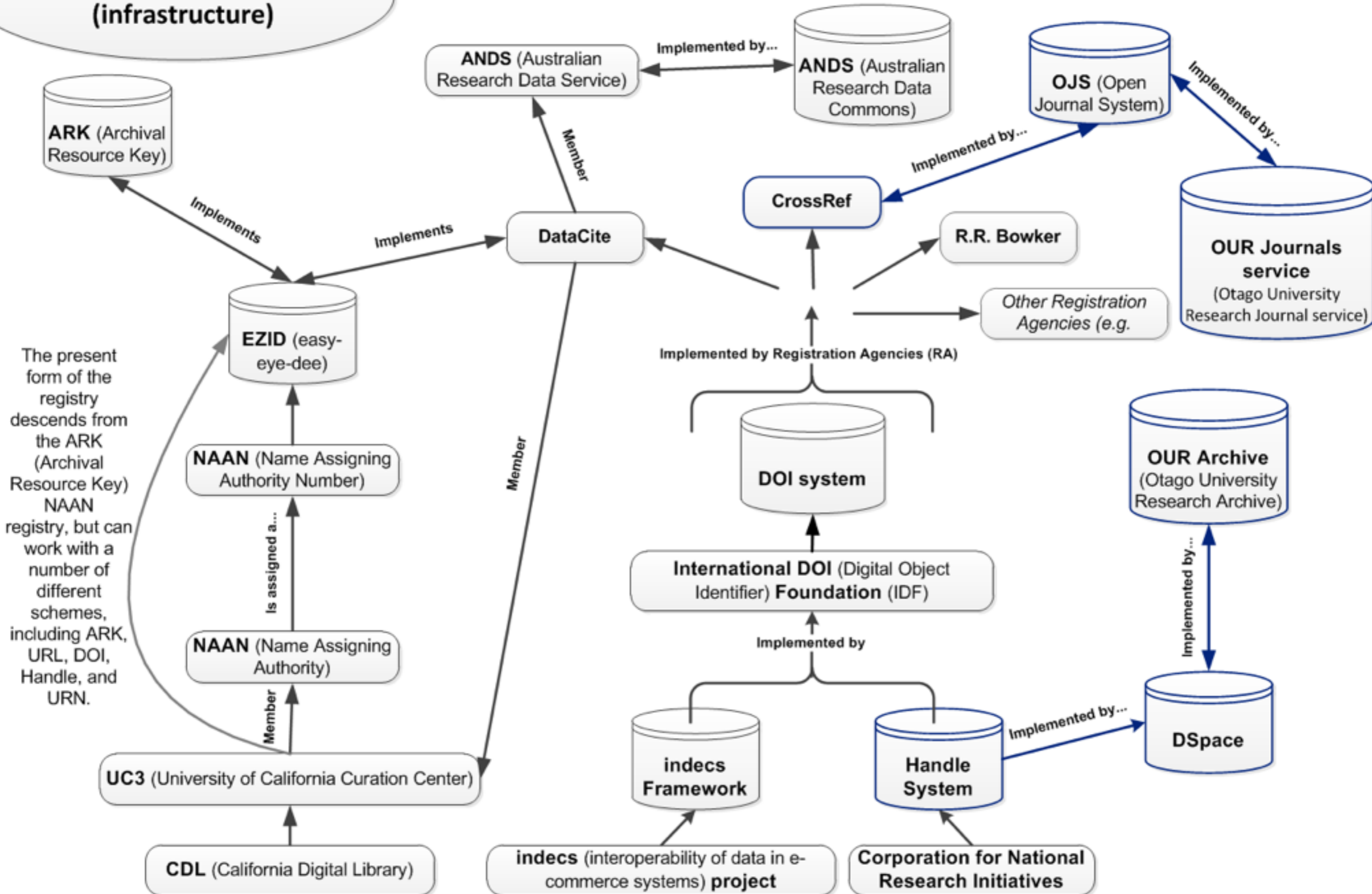 - include encoded or literal identifier in document to be printed

1. Upload digital artefact

- Researchers can easily acquire sharable digital storage
  - … but again researcher self-service risks ad hoc solutions
- Publication via their host organisation is a different matter
  - More arduous process … since "storage" has a different meaning
  - Host has obligations regarding quality of service of storage
- Infrastructure side: how to provision reliable, long-term storage?
- Distinction between "living" and "finished" artefacts
  - Storage systems for live data are frequently different from archival services
  - What if updates need to be made to the content? Self-service?
  - Will the storage system be tightly coupled to the identifier and metadata?

2. Gain Internet-accessible identifier for artefact

- How (some) researchers see an identifier:
  - simple: a URL
  - e.g., "share" link from Google, Dropbox, etc., …

- However University Libraries and other organisations aware of the management and collation needs are likely to see something more like…

Persistent Identifier schemes (infrastructure)

ANDS (Australian Research Data Service) — Implemented by... → ANDS (Australian Research Data Commons)

OJS (Open Journal System)

ARK (Archival Resource Key)

Implements

DataCite

Implements

CrossRef

Implemented by...

R.R. Bowker

Other Registration Agencies (e.g.

OUR Journals service (Otago University Research Journal service)

Member

EZID (easy-eye-dee)

The present form of the registry descends from the ARK (Archival Resource Key) NAAN registry, but can work with a number of different schemes, including ARK, URL, DOI, Handle, and URN.

NAAN (Name Assigning Authority Number)

Is assigned a...

Implemented by Registration Agencies (RA)

DOI system

Member

NAAN (Name Assigning Authority)

International DOI (Digital Object Identifier) Foundation (IDF)

Implemented by

OUR Archive (Otago University Research Archive)

Member

Implemented by...

UC3 (University of California Curation Center)

indecs Framework

Handle System

Implemented by...

DSpace

CDL (California Digital Library)

indecs (interoperability of data in e-commerce systems) project

Corporation for National Research Initiatives

2. Gain Internet-accessible identifier for artefact

- **Problem**: ideal research artefact ID may not be an ideal Internet-accessible ID
  - Existing and legacy systems cannot simply be moved to URL space
  - Likely that there are many different types of IDs already in use
  - Maintaining mappings between different ID spaces is risky
- **Again**: what about live data versus archival content?
  - Must researchers use different systems until content is "complete"?
- Ideally want **self-service** with **delegated authorisation** to researcher, however:
  - Organisations want control over **quality** of what is in the ID space
  - Is the researcher able to manage all of their **different ID versions**?

3. Potentially encode the identifier

- Assuming a URI is provided to researcher, do they know what should and shouldn't be done to it prior to its use?
  - Ideally a canonical form should be agreed and maintained …
  - … but if URIs are too long, researchers may be tempted to use URL shorteners, adding another processing step to the chain

- Self-service works in terms of producing printed codes: doesn't change the ID (e.g., URI)
  - … but has author thought about potential problems the document's readers may face?
  - Ideally avoid reinventing the wheel
  - e.g., QR-codes have choices regarding density and error correction

4. Include (encoded) identifier in target document
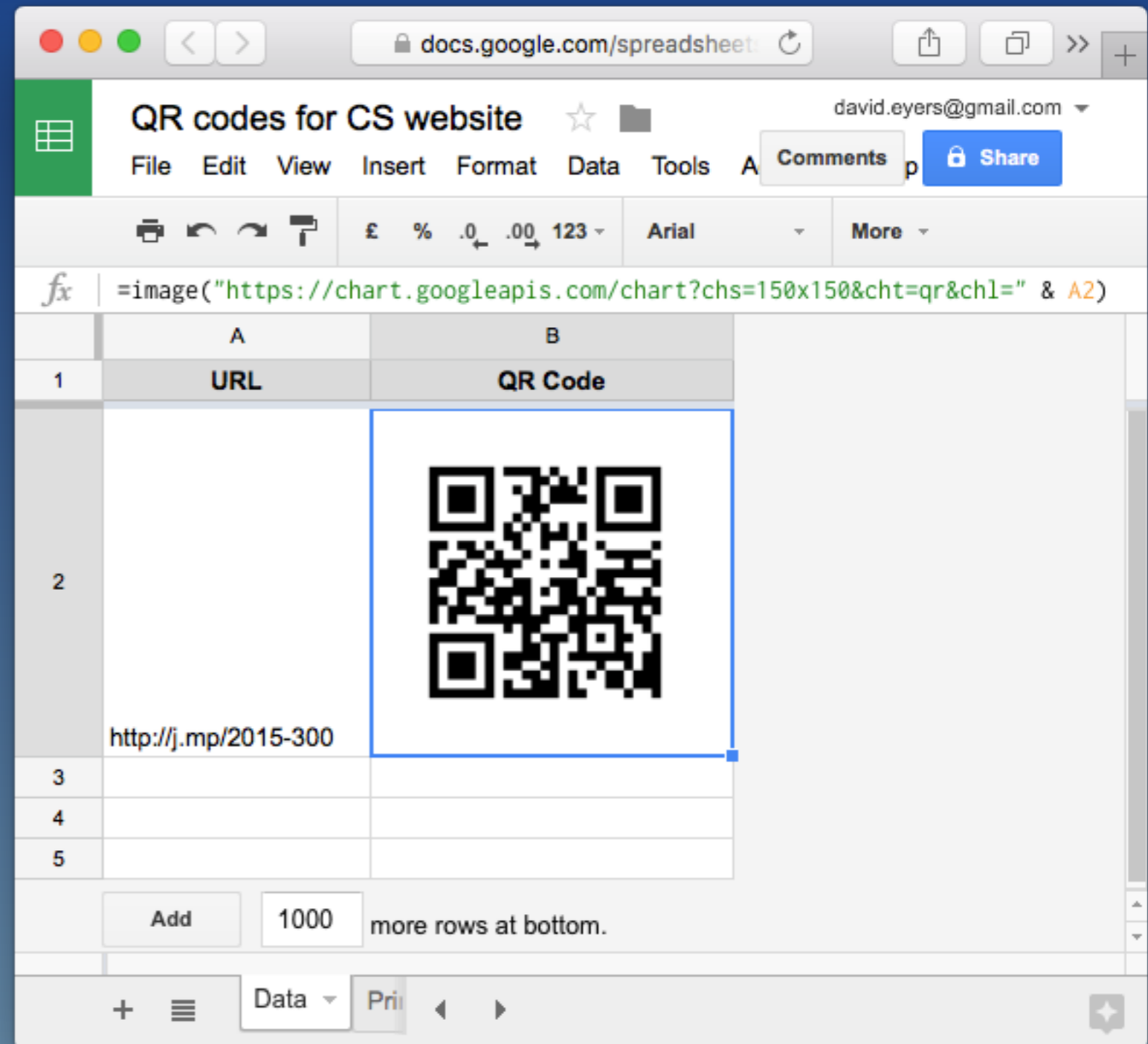
- Researcher must manage artefact links within document
  - Most document preparation tools should provide reasonable support
  - What about managing large sets of URIs?
  - Should the media ever change if the referencing ID doesn't change?

- Physical printing considerations
  - Barcode distortion: likely not a problem with 2D codes
  - What size is appropriate? Balance between obtrusiveness and scanning convenience
  - What can best support readers who may be using different types of browsing devices?
  - Has a convenient fallback mechanism been provided?

Identifiers: URIs, DOIs, …

- The DIY ID can be a URI to an online resource
  - However citation is likely to work poorly: lack of clear metadata
  - For example, no clear organisation may be provided
  - The URI may well not be unique

- DOIs can be "minted" by University Libraries; are citable
  - However they may involve a more heavyweight setup process
  - DOI URIs are not directly oriented to playback
  - They are not easy for humans to type, so fallback is difficult

- Figshare provides free storage and DOIs for artefacts…
  - … provided a Creative Commons licence is used for their release

Encoding identifiers: QR codes

- QR code generation using Google tools is easy
  - Image formula uses Google's charting API
  - Your data is seen by Google… but that's probably good, here

- (Note that I am not expecting the j.mp service to persist)

11

More detail about encodings of identifiers

- QR codes are de facto standard for targeting smartphones
  - Key point was the infringement-free intellectual property release
  - Creative things done with QR-codes: e.g., explicit distortion
  - Downside is that they do need two-dimensional space

- Other "1D" barcode forms available… but often numeric

- OCR may be sufficiently good to read short URLs directly
  - … particularly if they are guided by some formatting like a box

Case study

- Our specific case study: John Egenes' doctoral thesis

- John is a Lecturer in Contemporary Music at the University of Otago, and a long-term contributor to eResearch here

- He has used QR codes to link music samples from his thesis, supported by the University Library
  - In addition to linking to John's multimedia content itself, there is great potential to enrich the media to indicate interrelationships:
  - e.g., where multiple independent artefacts are "takes" of the same excerpt of music, or where a set of artefacts were parts that were mixed together to form a different artefact

John's musical examples (step 1)

- Audio recordings:
  - Some whole tracks
  - Some excerpts

- Available on Sound Cloud
  - … not yet available through Otago

UoO Library URIs (step 2)

- University of Otago DAMS system: still under development
  - Marsden Archive was a technical pilot
  - John's music collection is a subsequent pilot
  - https://dams.otago.ac.nz/OUA_0001_004

- Library-side screenshot shown—project was harder to launch than expected

John Egenes thesis music album 1

View | Manage

Grid view | List view

1 2 3 next › last »

04 Angels Flow By Anthony Ritchie

05 Intro and Scherzo By Graham Downes

06 The Sleeping Keeper By Claire Cowan

07 Valse Oubilee By Ross Carey

John's QR codes (step 3)

- The QR code on the right is for the track "Valse Oubilee"
- Specifically, it encodes:
  - http://goo.gl/zBB3gK
- … which redirects to the longer:
  - https://dams.otago.ac.nz/OUA_0001_004

- Already have a range of pure and impure names (in the computer science sense)

- Need to be careful about image scaling: high-frequency transitions may get interpolated in a non-ideal way

16

The printed dissertation (step 4)

- A subset of thesis content is shown on the right
  - Printed QR-codes have correct padding
  - Ideally URLs would provide a backup
  - (… although that would expose the Google shortening)

5.3.1. *Stone Soup:* Basic Tracks & Scratch Templates

Production on *Stone Soup* began with the creation of guide sessions, produced within the DAW (*Sonar Producer)*, for use by the participating musicians.[242] Once satisfied with the collection of songs I had chosen, I set about mapping templates for rough recordings of each of them. These templates were demo recordings that I refer to as "scratch templates" (see glossary for explanation) and shared certain similarities with the rough mixes given to the distance musicians on the *Tyre Tracks* project. Their primary purpose was to provide background music over which the musicians could play their particular parts. In doing so, the scratch templates locked every musician into the same tempo, key, and song arrangement, thereby automatically aligning their performances to each other and easing the job of the mixing engineer. The relative volumes and stereo placement of instruments and vocals

Scratch Templates

Scratch template succession for "One Hundred Dollars"

Scratch Template (with count-in)

Guitar solos: Tony [left], Rob [right]

Ongoing questions…

- Different storage services are available at the UoO
  - Can existing systems be incorporated, and data migrated between services over time, avoiding deploying yet another storage service?

- What ontologies and ontology mapping tools can be provided to researchers easily so that they can enrich the metadata associated with their multimedia content?

- Can the Open Archives Initiative's Object Reuse and Exchange specs. provide resource mapping and reporting?
  - If relationships between media clips are discoverable via semantic web technologies, is it possible to measure an increase in impact?

Future: context-sensitive apps

- QR Code reading is freely available but not a core feature on all smartphones (i.e. iOS does not include it)
  - Other reading support standards may become available

- Appealing potential: apps that track your context
  - Read the context using camera recognising the shape of the page
- Not just linking from document to artefacts in isolation
  - Can also encode relationship between those artefacts
  - different takes of a particular piece, or
  - tracks that were mixed together to form a given recording,
  - Allow navigation through a graph of multifaceted interconnections

Summary

- Progress is being made to provide an Otago service to support multimedia assets linked from paper documents
  - Still early days: plenty of time to refine our plans

- Questions?