

Digital information life-cycles and policies

David Ryan
DPC Forum
3rd July 2006

The challenges of digital information lifecycles

- What are they?
 - Determining whether long-term data retention is in line with your business needs
 - Establishing the business model
 - Ensuring information is captured and stored in a fixed form
 - Responding to technological advances to increase chances of longevity
 - Retaining the authenticity of records during the migration process
 - Auditing compliance

The challenges of digital information lifecycles

- Determining whether long-term data retention is in line with your business needs

Don't worry

- One thing to bear in mind is that the majority of existing files are stored on hard drives as if those drives were filing cabinets. Most documents, especially those in Microsoft Word, are created and sent and then rarely (or, more commonly, never) used again. Therefore, you don't have to worry about file format issues for every file on every file server in your enterprise
 - Kathy Ivens (Microsoft web site 12 July 2002)

The lifespan of electronic records

- The 'Microsoft view' - I think this reflects most people's view, if they have a view!
- Most processes take little time
 - minutes, hours, days, few take weeks, months, years and once task is completed...
- People move job every 3-4 years
- Software releases – every 2-3 years

The challenges of digital information lifecycles

- Establishing the business model

Policies

- Creator or receiver or both?
- Creator
 - Retention schedule
- Receiver
 - Selection policies
- Both
 - See above and integrate them

Policies

- Internal
 - Organisation specific
 - Based on business need
- External
 - Sector specific compliance
 - Some requirements generic e.g. H&S
 - For public sector only
 - To meet customer or funding expectations

Policies

- Creator
 - Retention schedule as a policy repository
 - JISC template for HE
 - TNA and Records Management Society advice for all
 - Legal
 - Business need

Policies

- Receiver
 - DPC template Decision Tree as a policy repository
 - Business and technical logic for long term preservation

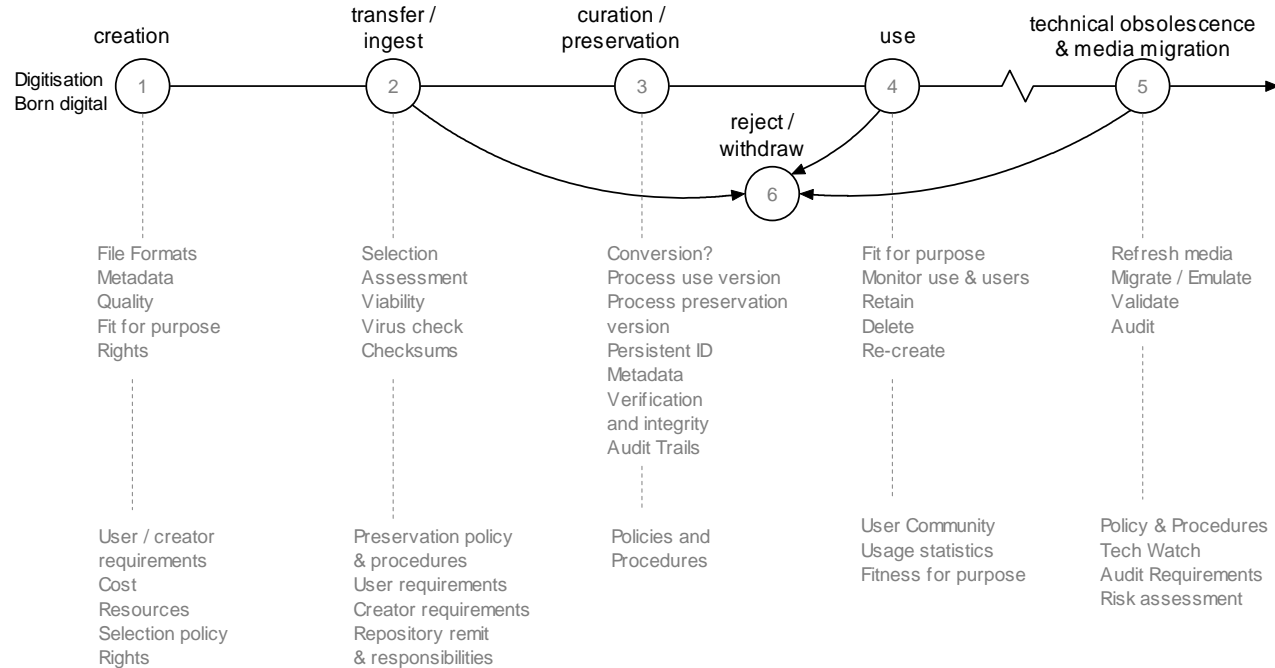
<http://www.dpconline.org/docs/handbook/DecTree.pdf>

Policies

- Receiver
 - The Institutional Repository
 - Advocacy
 - Intellectual property

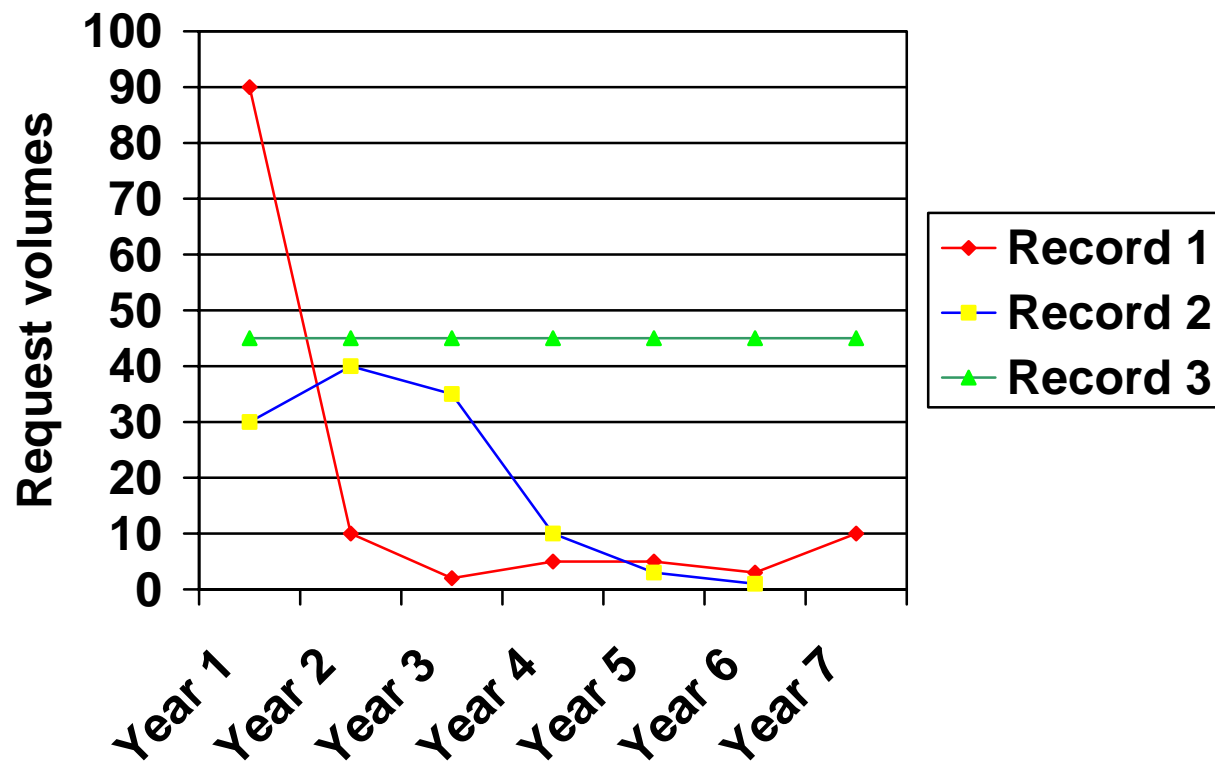
Jones, Andrew & MacColl, Chandos Publishing Ltd.,
Oxford, 2006

The life-cycle model for moving images and sound

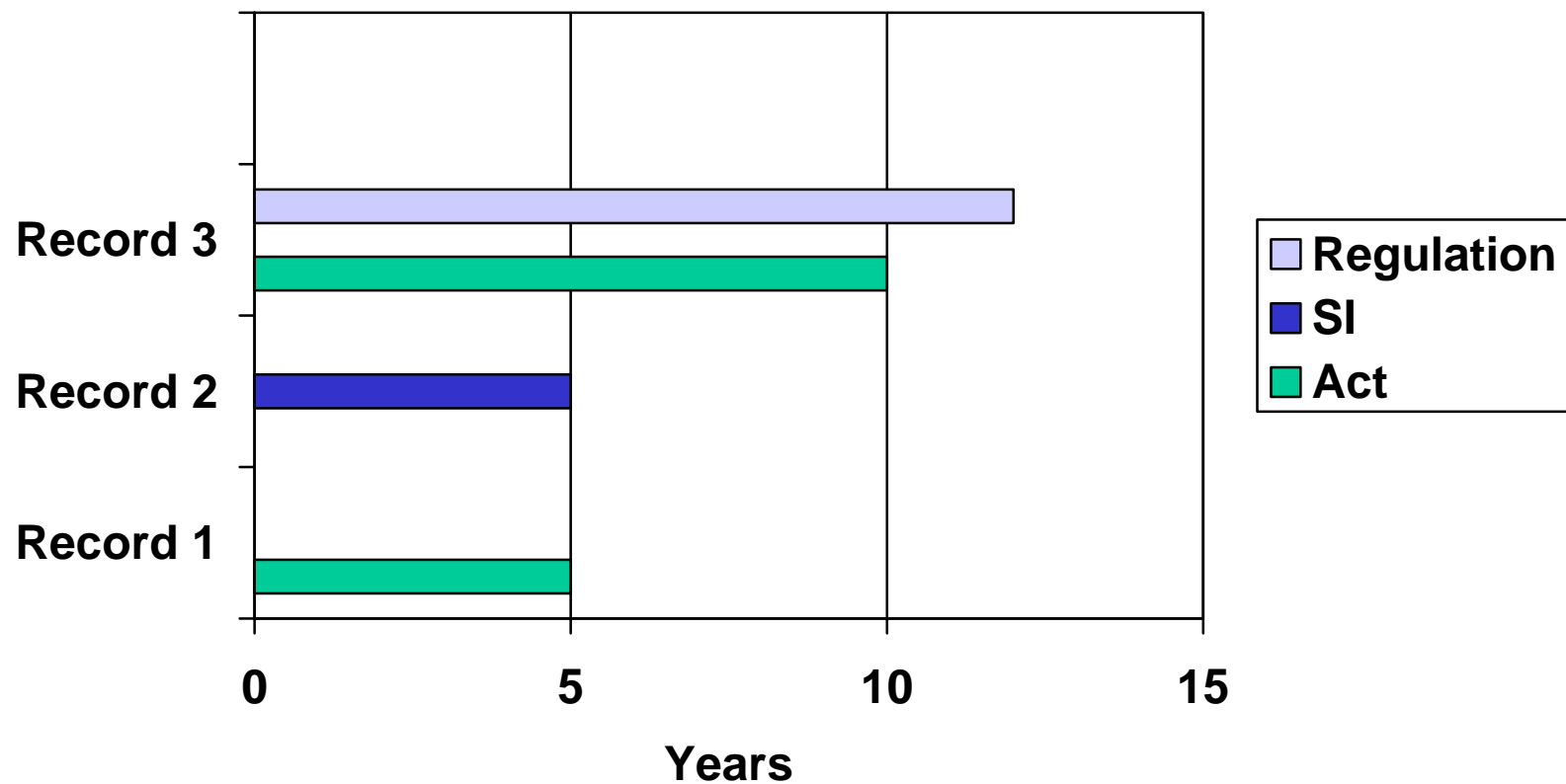


AHDS study for JISC, 2006

Lifecycle – business eye view



Lifecycle – legal view



The challenges of digital information lifecycles

- Ensuring information is captured and stored in a fixed form

Practical actions

- Identify regulatory and internal time expiry drivers
 - Review policies periodically
- Discover records wherever found e.g. Dbases, GIS, CMS, File system
 - Metadata for file format captured
- Regularise deletion or transfer to archive
 - E.g. ePrints, VDC, DSpace

Organisational issues

- Organisational permanence
 - Maintenance of policies
- Secure funding
- IT infrastructure
 - Maintain systems to ensure non-repudiation
- Available expertise
- Cross-community influence
- Data acquisition and validation

The challenges of digital information lifecycles

- Responding to technological advances to increase chances of longevity

Market watch

- New, cheaper, more reliable hardware
- New software utilities to preserve data
- Better cost models
- Better business models
- Opportunities to share

Availability of information

- Publicly available
 - As long as provider wants to
- Public domain
 - As long as provider exists
- Proprietary:
 - May be available for a fee
 - Reverse engineering:
 - » Provenance
 - » Copyright

The challenges of digital information lifecycles

- Retaining the authenticity of records during the migration process

Registry attributes

- Persistence
- Unique identifiers
- Trustworthy content
 - Identifies source of information
- Coverage
- Publicly discoverable
- Machine processable (ideally)

PRONOM

More practical actions

- Establish if any migration is required before deletion
- Establish business criteria to select migration approach
 - most effective; least expensive; etc.
- Use tools
- Remember, this is a technical issue
 - A record is a record no matter which stage of the lifecycle it is in

The challenges of digital information lifecycles

- **Audit compliance**
 - How do I know what I need to do?
 - See establishing the business model
 - How do I know I have done it?
 - Policies agreed
 - Practical actions undertaken
 - How do I know I've been successful?
 - Self assessment
 - Internal or external audit

The future

- Centralised yet distributed expertise
 - DPC, DCC
- Improved technical and cost modelling
- More co-operation and cost sharing
 - UK Webarchiving Consortium