

Implementing Metadata that Guides Digital Preservation Services

Angela Dappert, Adam Farquhar

The British Library

06 October 2009

Goals for preservation metadata



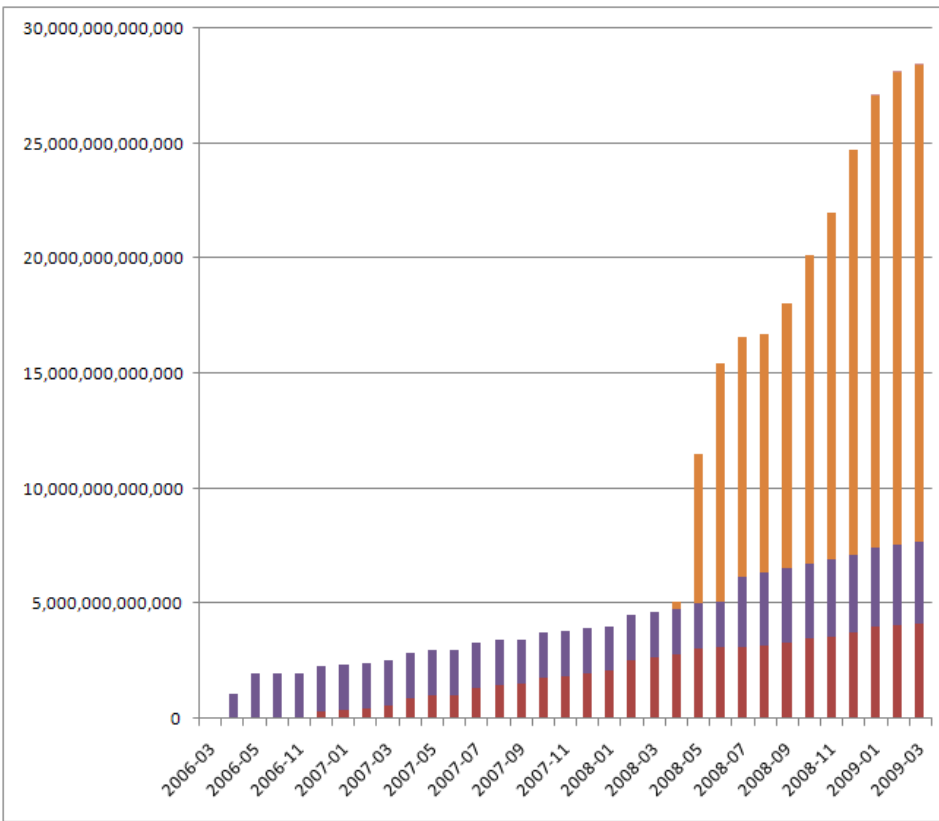
Keep it simple



But not too simple!

- Design and build
- Procure
- Implement

Does your digital library look like this?



Things are getting serious

- Investment in digital content
- Operational impact
- Moving into the steep portion of the curve

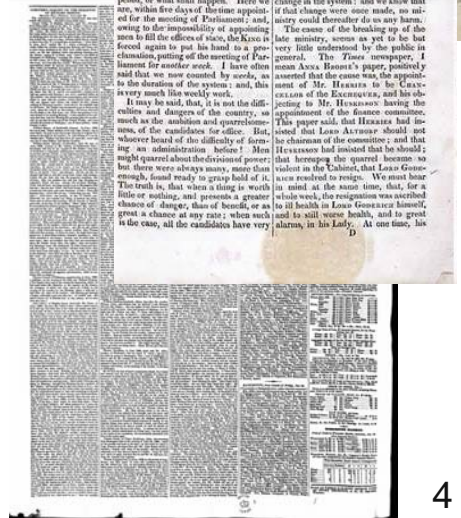
We are learning from experience

Example – TIFF image preservation



A simple, but realistic case study

- 2,000,000 compressed TIFF images taking up 100TB
- Preservation risk: decaying media
- Planning options:
 - Refresh storage only
 - Refresh storage and convert to JP2K (lossless)
 - Refresh storage and convert to JP2K (lossy)
- Organisational requirements:
 - Result must be quality assured
 - Fixed storage budget



Characteristics belong to objects...



Objects do have characteristics

- File size, page count, pixel depth, ...

Environments have characteristics

- Has a TIFF viewer
- Has a JP2K viewer
- Processor type
- Memory or CPU speed

Actions have characteristics

- Cost
- Speed
- Memory or CPU consumption
- Single item or batch
- Metadata retention

A value is a value is a value...



Examples:

- Page count
- Image count
- Image size
- Colour
- Reliability

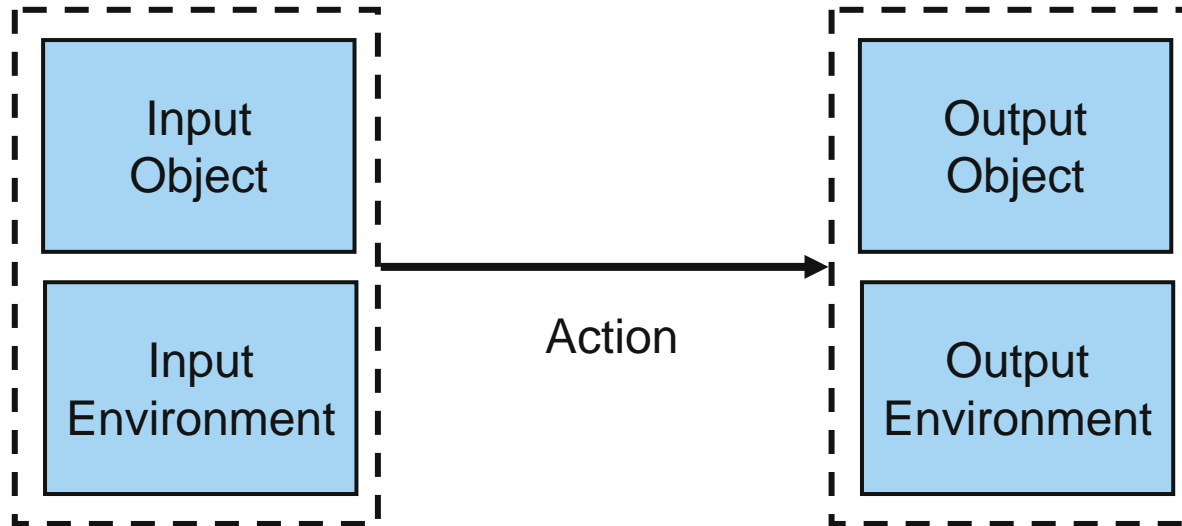
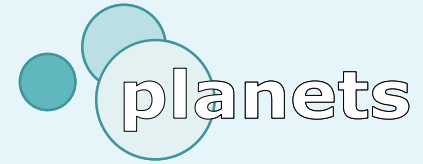
Do multiple values conflict?

Units

Origins

- Source
 - Eg Pronom, UDFR, inventory, controlled vocab
- Agent
 - Eg JHove, imageMagick, Droid, curator, ingest engine
- Technique
 - Print and count
 - Interpret XCEL
 - Test on 1000 sample files
 - RDA Cataloguing rules
 - Fast Fourier Transform

Migration is about the object...



Migration may change both the object and the environment!

TIFF & TIFF-Viewer

JP2K & JP2K-Viewer

What's good for the goose ...



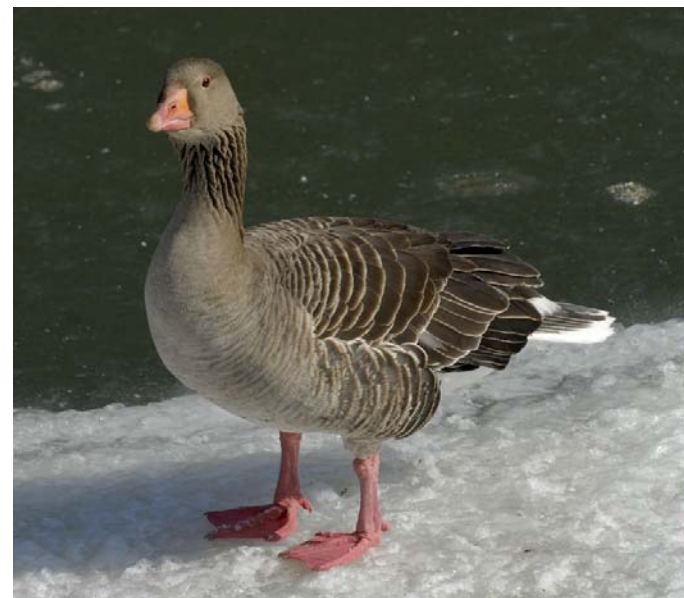
TIFF to JP2K can be an excellent plan

- **British Library**
- **Welcome Trust**

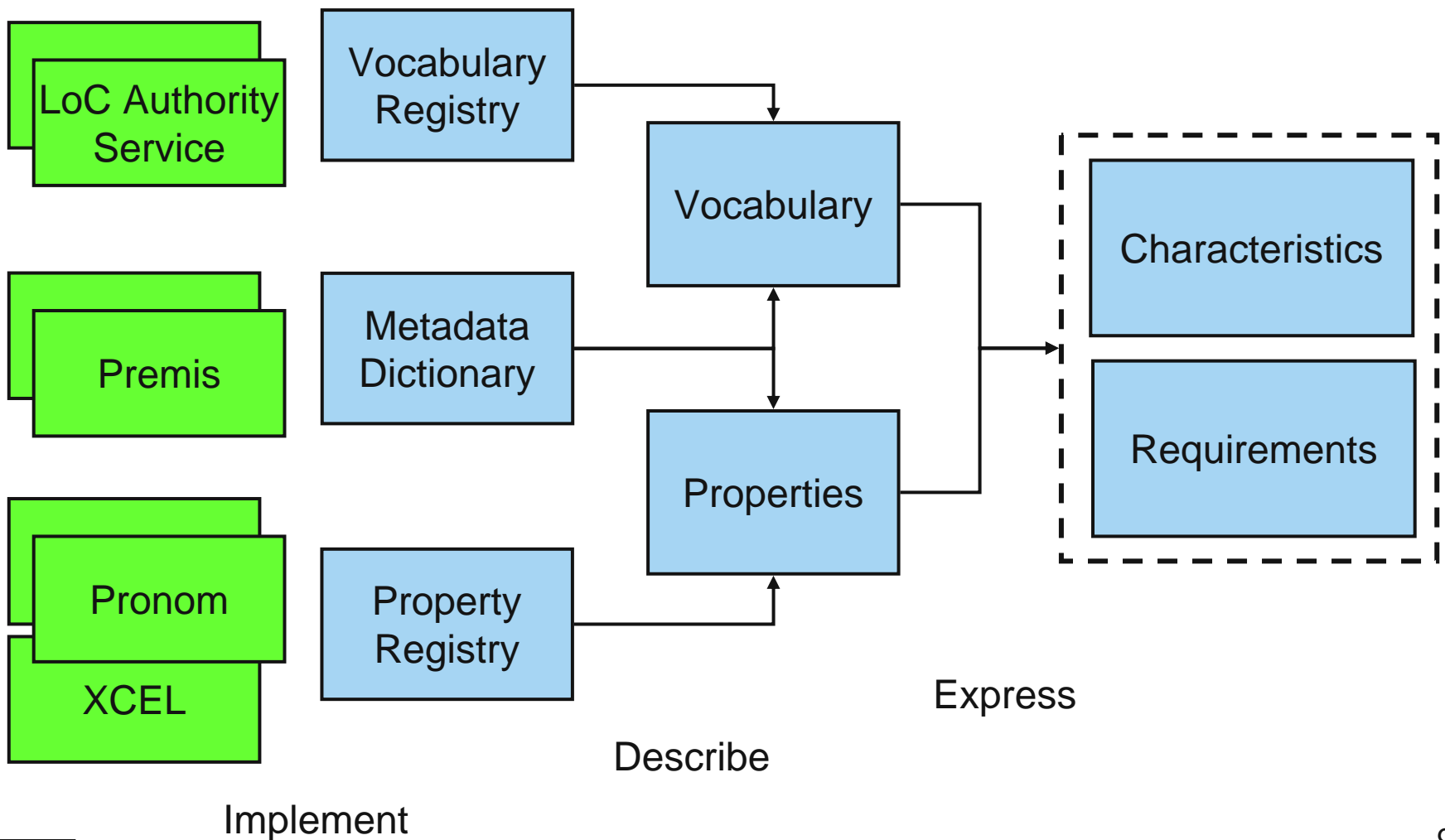
But

- **Lossy compression saves more space**
- **Some users don't have viewers**
- **JP2K can take longer to produce, ...**

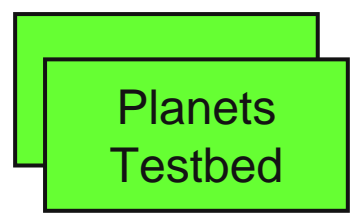
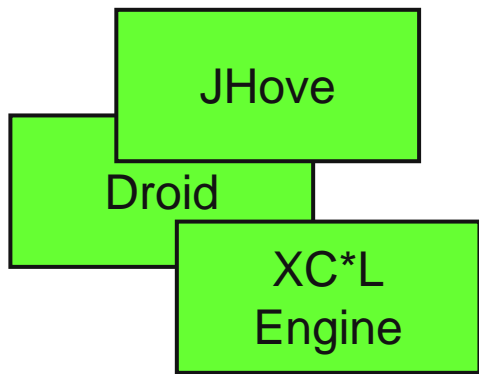
Your requirements are the key



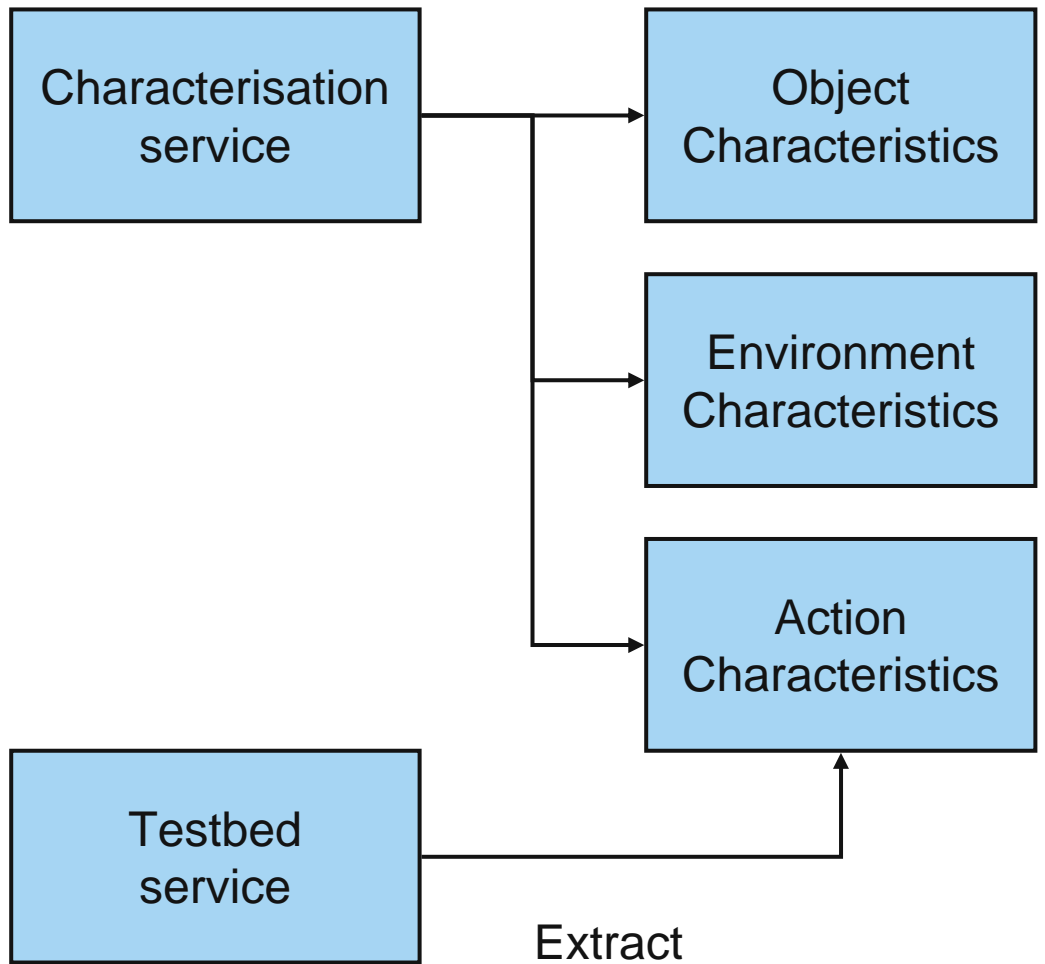
Expressing characteristics



Describing objects, environments, actions



Implement



Extract
Infer
Measure

Summary



Preservation services need several types of metadata

- Yes, information about objects, but
- Information about who measured what and how
- Information about environments
- Information about actions
- Information about organisational requirements

Keep it simple, but not too simple!

Contact

[angela.dappert {@} BL.uk](mailto:angela.dappert@bl.uk)

<http://www.planets-project.eu/>

<http://www.bl.uk/dp>