

Establishing the significant properties of digital research

Gareth Knight
Centre for e-Research
King's College London

iPRES 2008
29 September 2008





Overview

- Definitions
- Potential risks to significant properties
- Criteria for deciding significance
- Recording and comparing SPs
- General observations



InSPECT Project

- **Project:** Investigating the Significant Properties of Electronic Content over Time
- **Development Partners:** Centre for e-Research, KCL; The National Archives; The British Library (advisory)
- **Objectives**
 - Expand and articulate the concept of ‘significant properties’
 - Determine the properties that are significant to the long-term accessibility of different types of digital object (email, presentation structured text, audio, raster images)
 - To develop methods for expressing and measuring properties to:
 - validate the results of preservation actions
 - support the needs of user communities



Many definitions...

“The characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability and meaning of the objects”

Wilson, 2007

“Significant properties are those properties of digital objects that affect their quality, usability, rendering, and behaviour.”

Hedstrom & Lee, 2002

“Those characteristics (both technical, intellectual, and aesthetic) agreed by the archive or by the collection manager to be the most important features to preserve over time.”

The Cedars Project Report, 2001

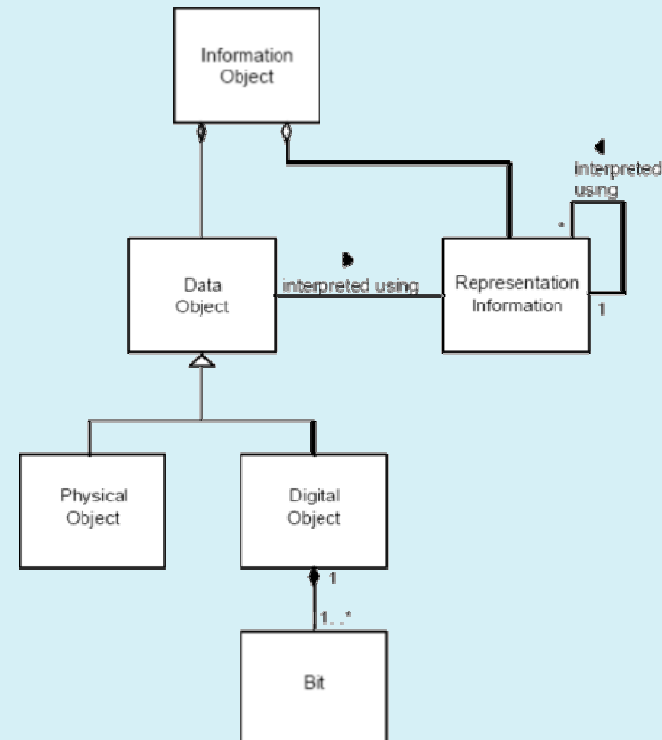
Closely tied to notions of ***authenticity*** and ***integrity***



Representation Information

RI consists of:

- *Structure information* that describes the encoding scheme in which data is stored, e.g. format, encoding algorithm
- *Semantic information* that indicate how the values are to be interpreted. E.g. documentation that indicates how numeric values in a CSV or tab-delimited format must be interpreted.



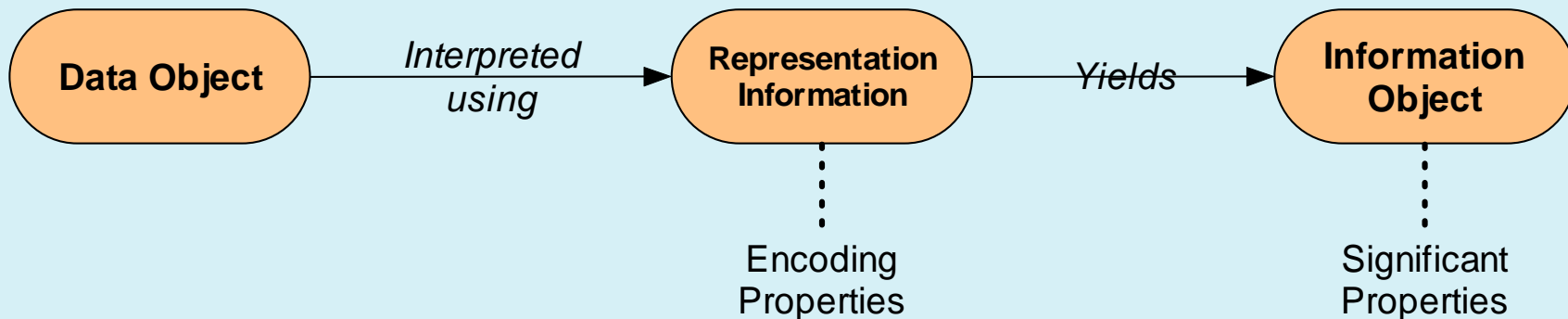


Interpreting SPs in abstract

NAA Performance Model



OAIS Reference Model





Risk scenarios

“traditionally, preserving things meant keeping them unchanged; however ... if we hold on to digital information without modifications, accessing the information will become increasingly more difficult, if not impossible.”

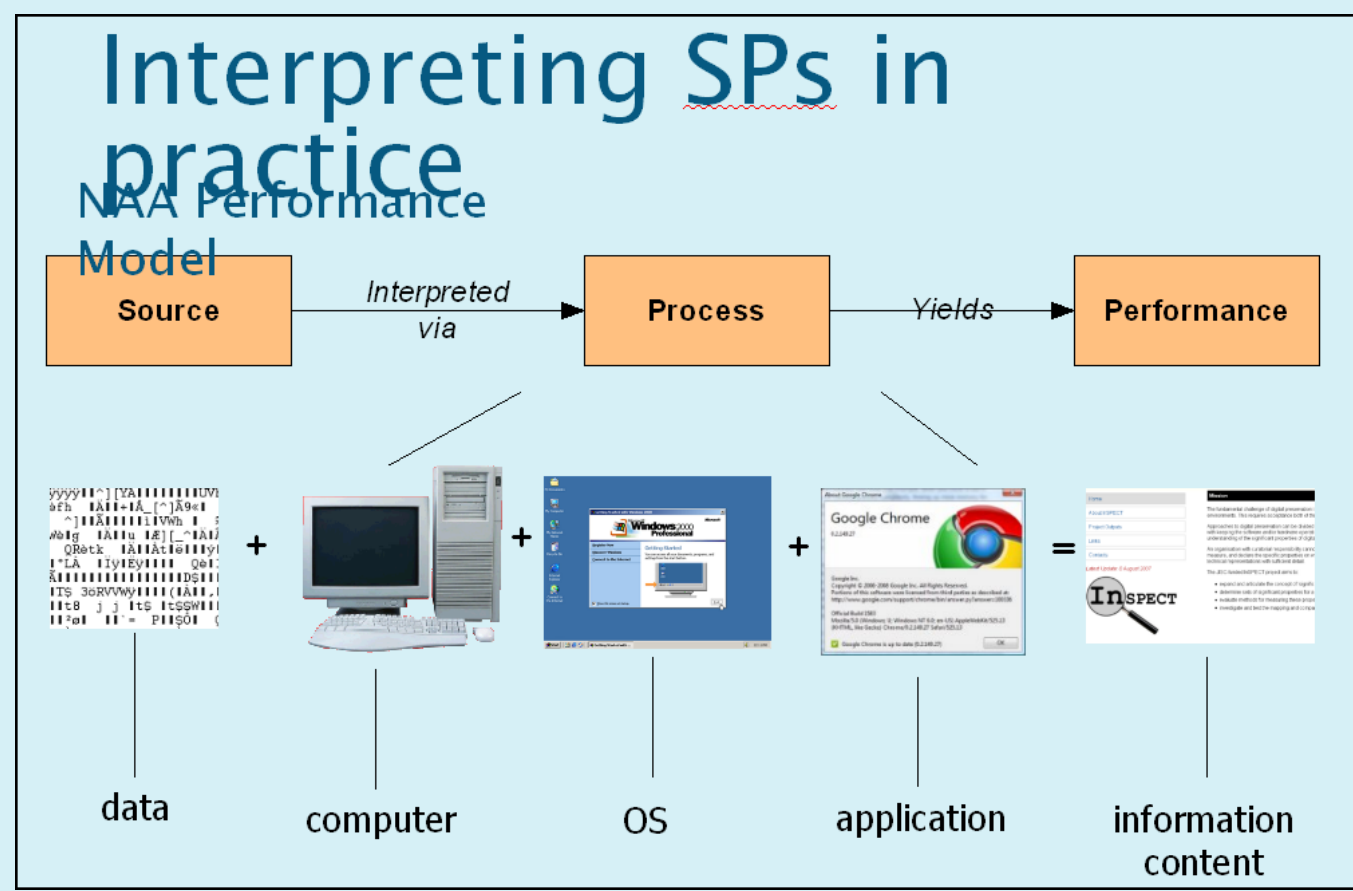
Su-Shing Chen, “The Paradox of Preservation”, *Computer*, March, 2001, pp. 2-6.

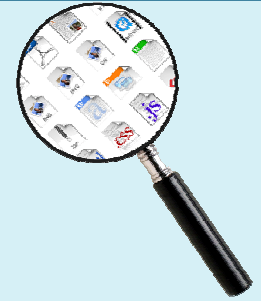
1. Recreation of source data
 - Hardware (e.g. upgrades, virtual machines, emulators)
 - Operating system
 - Software application
2. Conversion – format normalisation/migration
 - File Format
 - Encoding format



Differences in rendering...

Previous slide in OpenOffice Impress 2.0





What is significant?

“A digital object’s Significant Properties are not empirical; archives will make judgments at levels appropriate to fulfil their preservation responsibilities and meet the needs of the archive’s user communities”

The Cedars Project Report, 2001

“Definitions of Significant Properties that affect the aesthetics, implied meaning, and affordances of digital objects tend to be much more subjective and tied to the context of creation and use.”

Hedstrom & Lee, 2002

Fundamental Questions of digital preservation:

1. What must you retain to ensure the integrity and authenticity of the digital object?
2. What can you lose without potential implications?



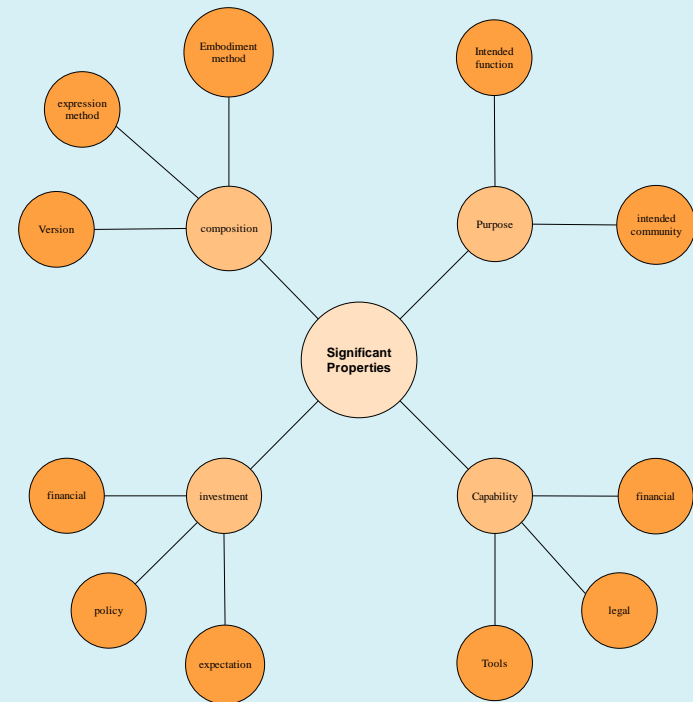
Frameworks

- Rothenberg & Bikson (1999)
 - Encouraged analysis of business functions, followed by technological capabilities
- Digital Diplomatics (2001)
 - Created by InterPARES project, based on archival diplomatics
 - Analysis of Records rather than Objects
 - Examines Documentary form, Annotations, Context of creation and use
- Utility Analysis (2004)
 - Developed during DELOS project and used in PLANETS
 - File characteristics, process characteristics

Criteria for deciding significance



- **Composition of the digital object**
 - Form in which the idea is expressed
 - Expression method in a digital environment
- **Purpose**
 - Intended function (e.g. diplomatic analysis)
 - Type of user
- **Organisational investment**
 - Strategic
 - Financial
 - Expectation
- **Capability**
 - Tools
 - Legal
 - Financial





Significant property types

- Characteristics of the intellectual content itself
 - Length (duration of audio recording, no. of characters)
 - Placement (e.g. audio playback through left/right speaker, position and size of shape, sequential order of several paragraphs)
- Properties that indicate the environment in which the intellectual content may be reproduced
 - Quality level (no. of colours, audio quality)
 - Access status (viewing, editing)



Cataloguing SPs (1)

Data Dictionary for Significant Properties:

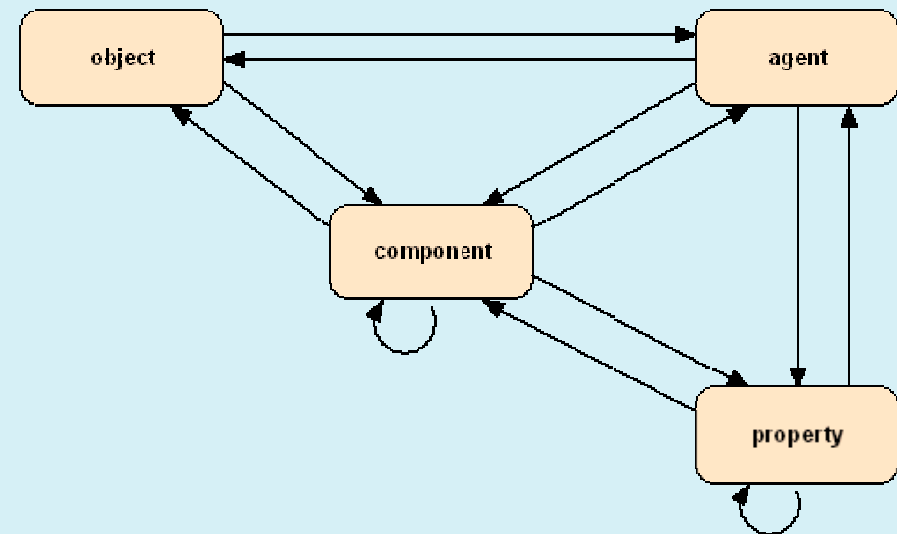
- Catalogue significant properties of digital object that must be maintained
- May be applied to range of resource types, file formats and subject disciplines
- Validate that information content is authentic, in regards to its original meaning
- Note any property constraints and their application to specific functions and designated communities

XML schema in the near future...



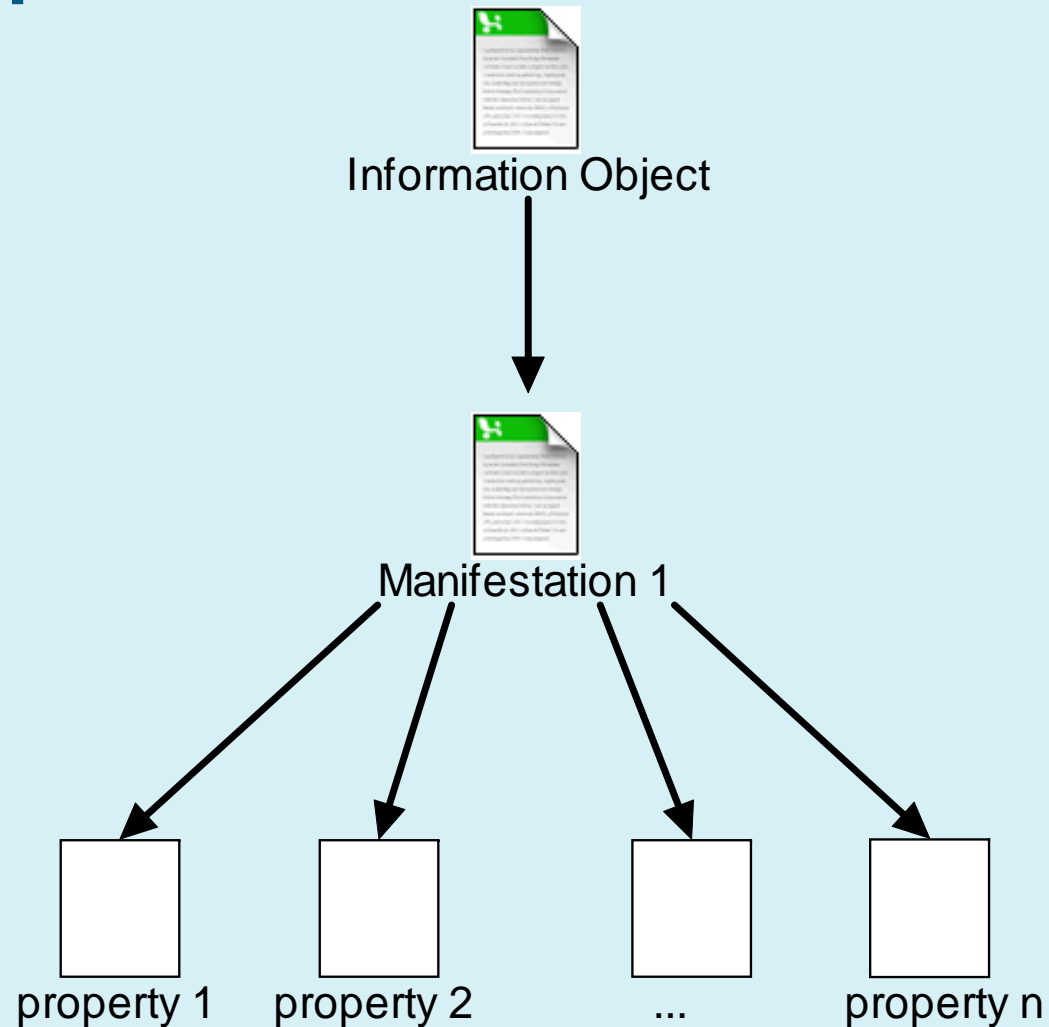
Cataloguing SPs (2)

- identifier
- title
- description
- function
- genre
- preservation Level
- specification registry
- measurement
- relationships



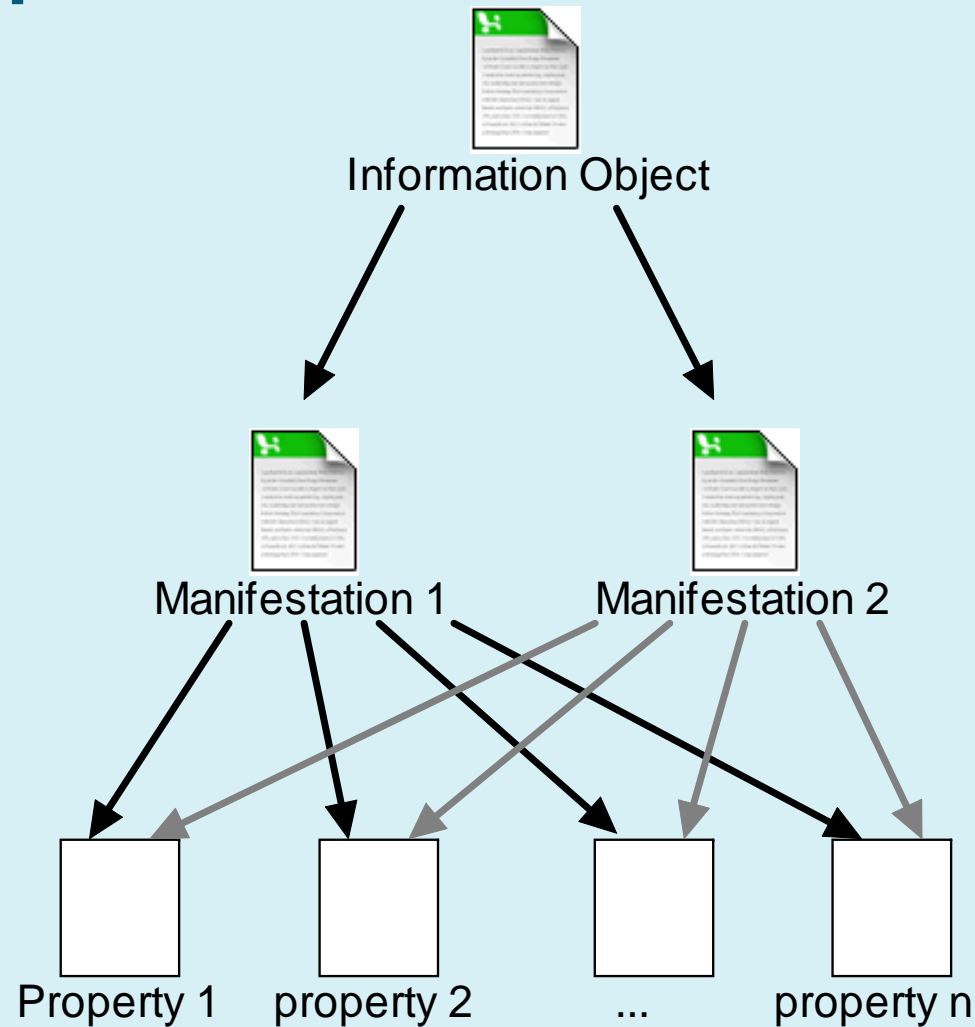


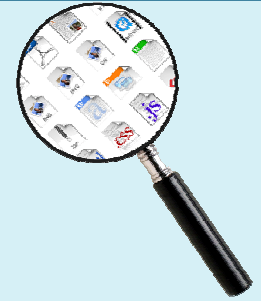
Compare and contrast





Compare and contrast





General Observations

- An understanding of significant properties is useful during creation and distribution of objects, in addition to long-term curation and preservation.
- Although some properties may be identified that are important to the form of resources, many decisions on significance require consideration of the context.
- Curators should begin to consider ways to capture and retain significant properties information on ingest into repository
- The success of preservation activity should be evaluated on the basis of the ability to maintain significant properties
- Further collaborative work between creators, archives/repositories and tool developers is required to provide a consistent approach.