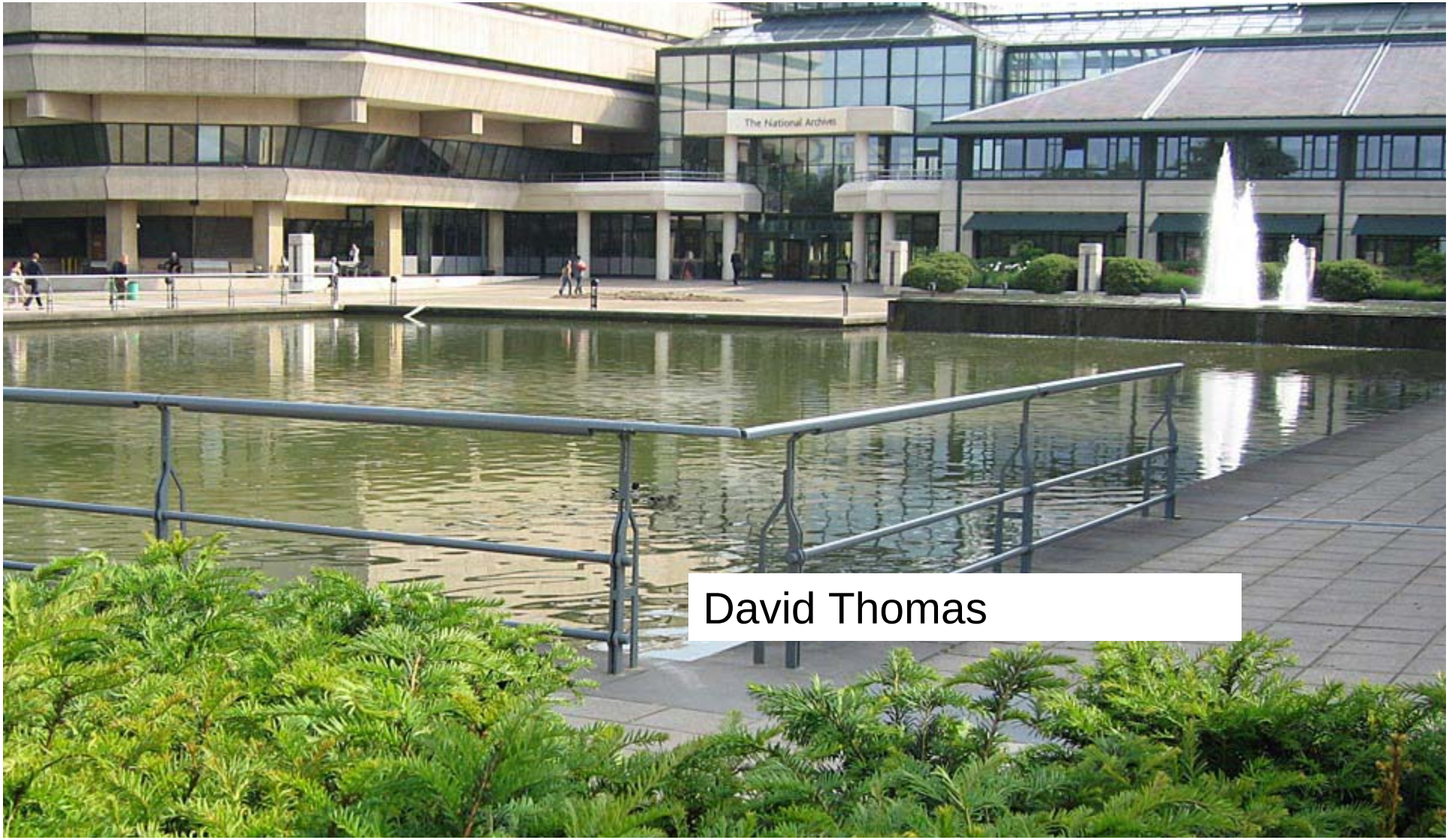
Digital Preservation in the UK



David Thomas

Digital Preservation in the UK

- Main focus on National Archives

  Websites

  Datasets

  Electronic records

Websites

Contract with European Archive http://www.europarchive.org/

Take about 65 sites – some weekly, some every six months and a few ‘current issues’ (Bird Flu; plane bombings etc)

Next - wikis

European Archive harvests, stores and displays websites

TNA takes back-up copy

Involved in UK Web Archiving Consortium a consortium of six leading UK working on a project to develop a test-bed for selective archiving of UK websites.

Datasets

- Social science datasets are held on our behalf by the UK Data Archive at Essex University http://www.data-archive.ac.uk/
- Other datasets are held by National Digital Archive of Datasets at London University http://www.ndad.nationalarchives.gov.uk/

the national archives

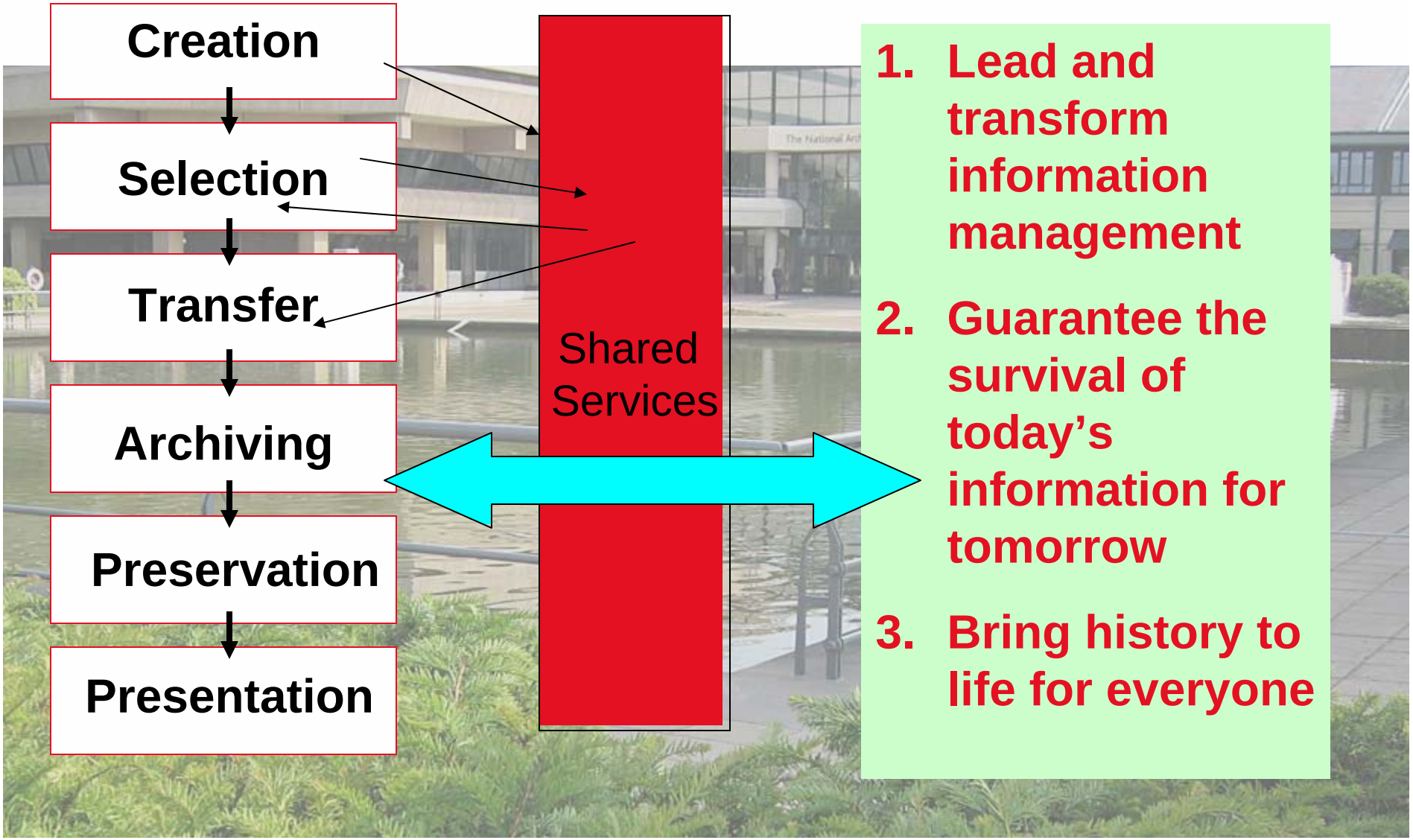
Which brings us conveniently to

Seamless Flow

Why Seamless Flow

1. Increasing use of electronic records
2. Increase in volume of electronic records
3. Inability to cope by manual means
4. Technology obsolescence
5. Need to be involved earlier in record life cycle

The Seamless Flow Programme



1. Lead and transform information management
2. Guarantee the survival of today's information for tomorrow
3. Bring history to life for everyone

Life Cycle Management

2. Approach

- Not big bang!
- Modular/projects
- Incrementally phased introduction
- Keep it simple
- Use of standards
- COTS
- Build on existing components!

Life Cycle Management

1. Objective

The Seamless Flow Programme is intended as far as possible to ensure a secure, managed, seamless, automated flow of electronic records from creation in government departments through their eventual preservation at the National Archives and delivery to the public over the Internet

Issues

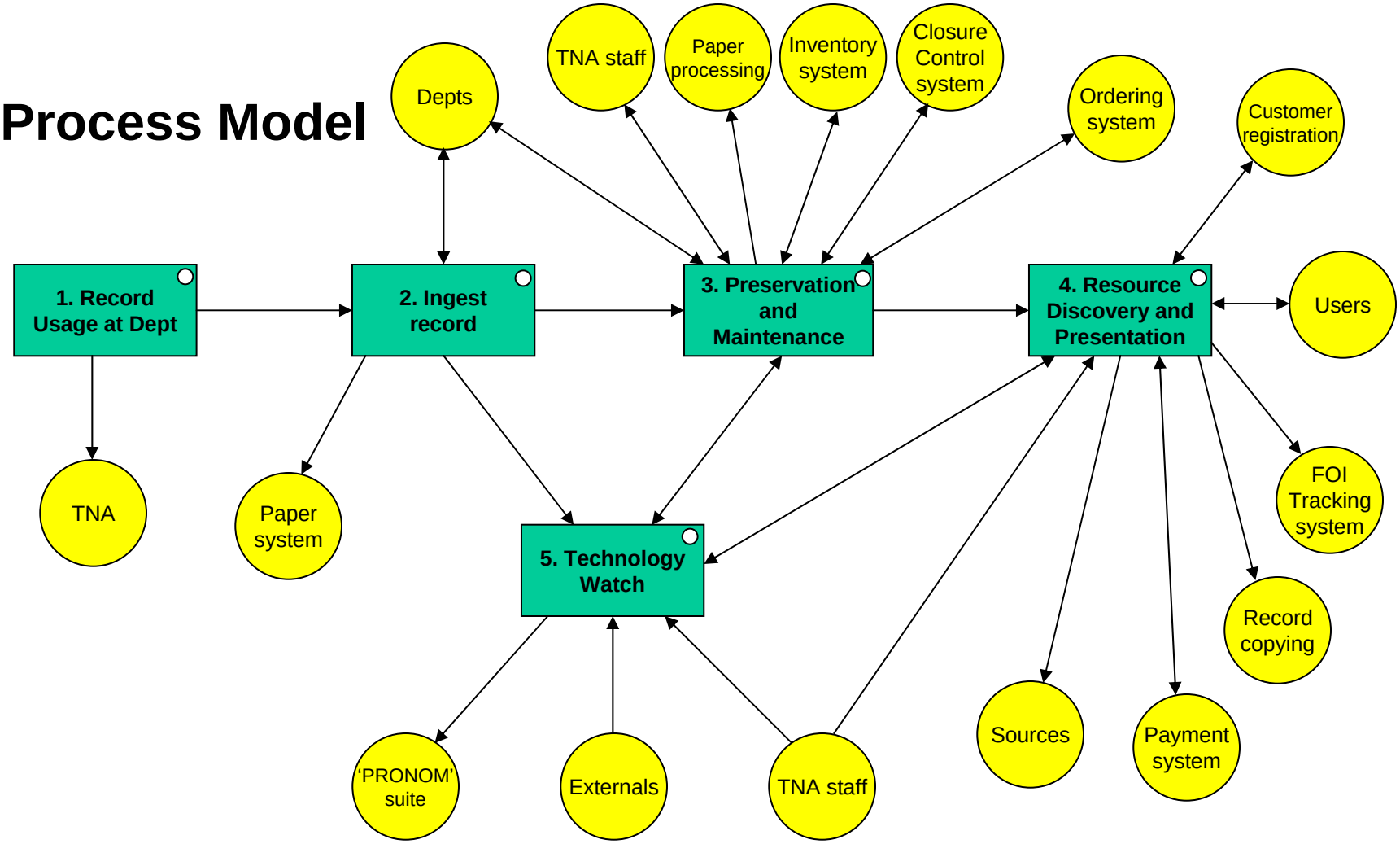
Managerial> technical

Social> managerial

Need to reduce costs

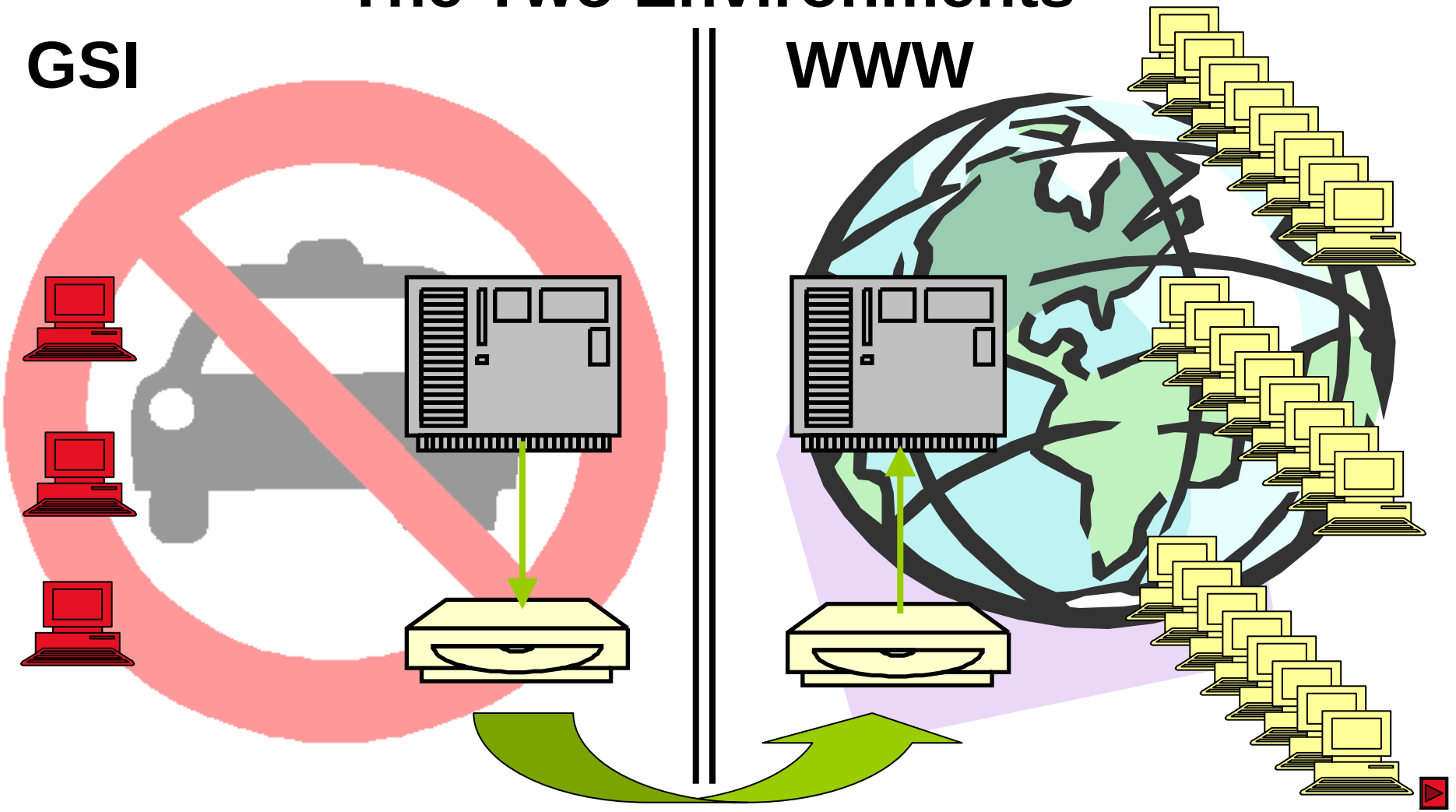
Quick wins

3. Process Model



(the life cycle of a record)

The Two Environments



Seamless Flow

Appraisal and Selection

Why Appraisal and Selection?

1. Early intervention to preserve the historical record
2. Secures the capture of context as well as information
3. Focus resources on records of known value
4. Historical appraisal a requirement to be built into disposal scheduling

How (on earth) do we appraise digital records?

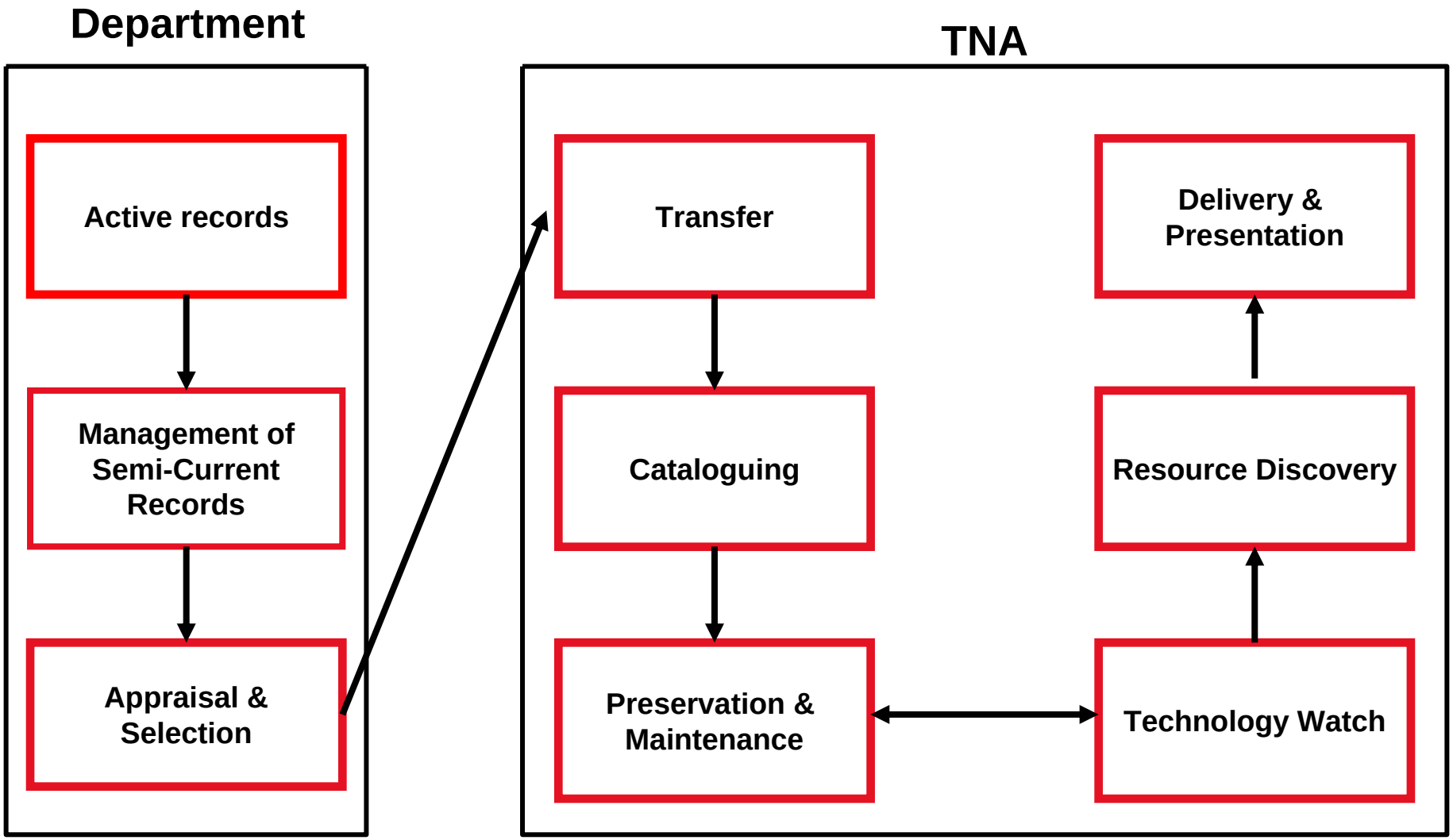
File Plan.obr

This is how:

- Move from review to appraisal;
- Use of macro-appraisal techniques;
- Appraisal decisions can be migrated as functions move within or between departments;
- Disposal scheduling is made easier by making the historical decisions;
- Still work to be done

Seamless Flow
Transfer to TNA

Seamless Flow Process



Transfer

Scope

- Physical transfer to TNA
- Pre-accession processes
- User Interface for editing metadata

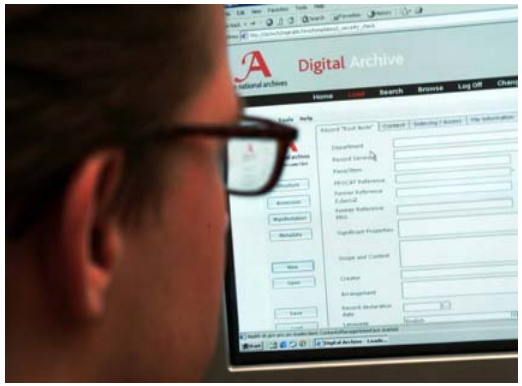
Transferring digital records raises new questions

- How do we best support government departments in the preparation of records for transfer?
- How do we manage the likely high volumes of digital records?
- How do we incorporate macro-appraisal?
- How to observe FOI, data protection, copyright and sensitivity issues?
- How do we actually move records from Departments to TNA?

Seamless Flow

Active Preservation

(Technology Watch)



Introduction

- Active Preservation is responsible for ensuing the continued accessibility of authentic records over time, and across changing technological environments. Its aims are:
- To generate and maintain accessible *preservation* copies of records
- To generate new *presentation* copies of records
- To provide facilities for secure redaction of electronic records
- To provide characterisation, preservation planning and migration services to support the above
- If possible, to make these services available for external reuse

PRONOM

- An online technical registry
- A resource for anyone requiring impartial and definitive information about the file formats, software products and other technical components required to support long-term access to electronic records and other digital objects of cultural, historical or business value
- A knowledge base to support automated preservation services

www.nationalarchives.gov.uk/pronom

the national archives

The technical registry
PRONOM

Welcome : About Add an entry
Search ? Help Information resources

? Help : detailed report on file format

Details: File format summary

File format | PRONOM Unique Identifier | Software | Vendor | Lifecycles

Details for: Portable Document Format 1.6

Save as... XML | CSV Print

Go to: Summary | Documentation > | Signatures > | Compression > | Character encoding > | Rights > | Reference files >

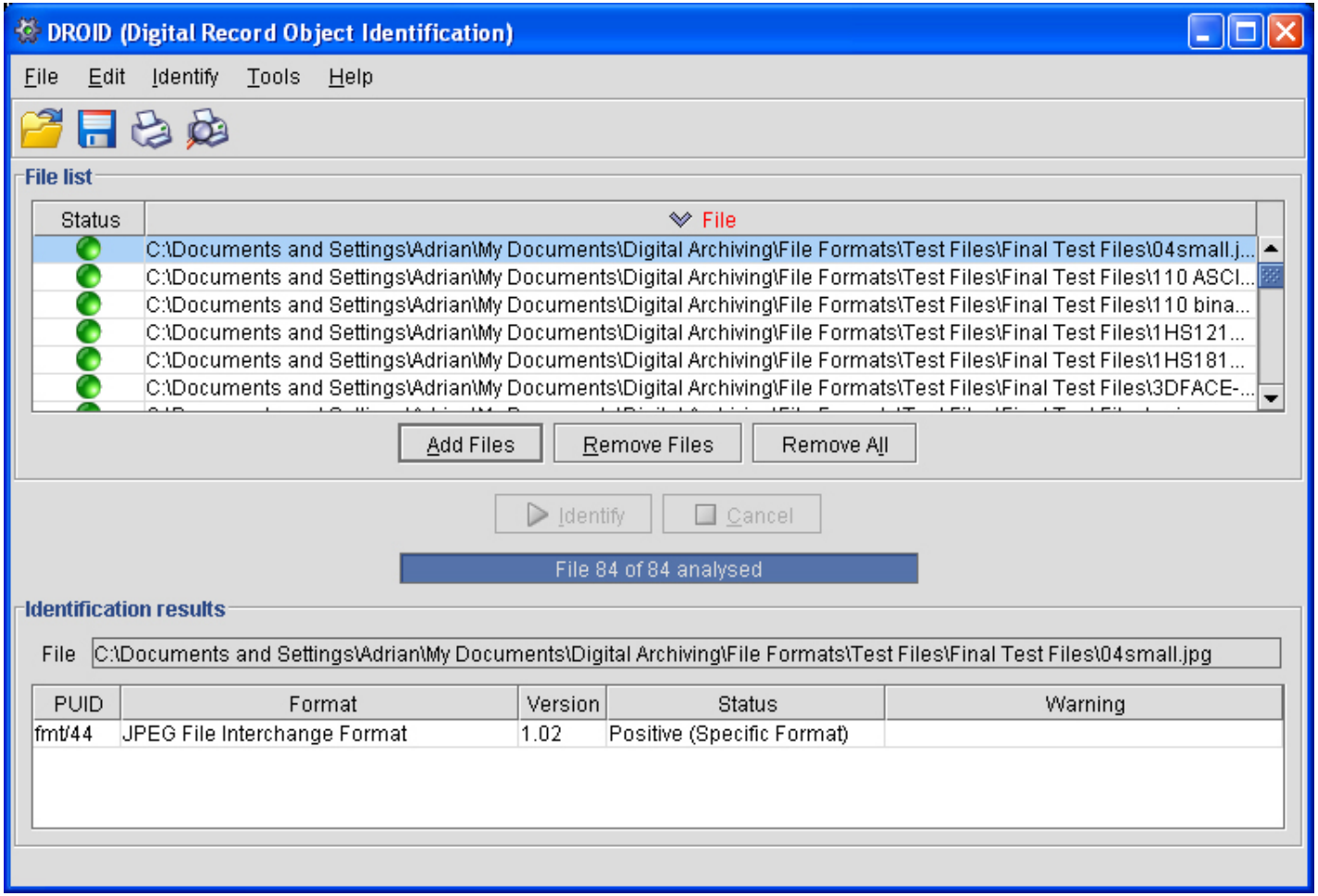
Summary

Name	Portable Document Format
Version	1.6
Other names	PDF (1.6)
Identifiers	PUID: fmt/20 MIME: application/pdf
Family	
Classification	Page Description
Disclosure	Full
Description	Portable Document Format is a platform-independent format for representing formatted documents, developed by Adobe Systems Incorporated. It is the native format of Adobe?s Acrobat family of software products, version 1.6 corresponding to the release of Acrobat 7.0. PDF is based on, and shares the same imaging model as, the PostScript page description language. A PDF file comprises a Header section, a Body section containing the objects which make up the document, a Cross Reference Table, and a Trailer section. PDF files can contain a wide variety of content, including text, images, video and audio.
Orientation	Binary
Byte order	Big-endian (Motorola)

Identification

- TNA has developed DROID (Digital Record Object Identification)
- uses automated analysis of the binary structure of a digital object, and comparison with predefined internal and external ‘signatures’ for specific formats.
- uses signature information stored in the PRONOM technical registry. PRONOM and DROID are both freely available on the web at http://www.nationalarchives.gov.uk/pronom.
- This method is currently limited to identification; full object characterisation, including validation and property extraction, is intended as a future enhancement.

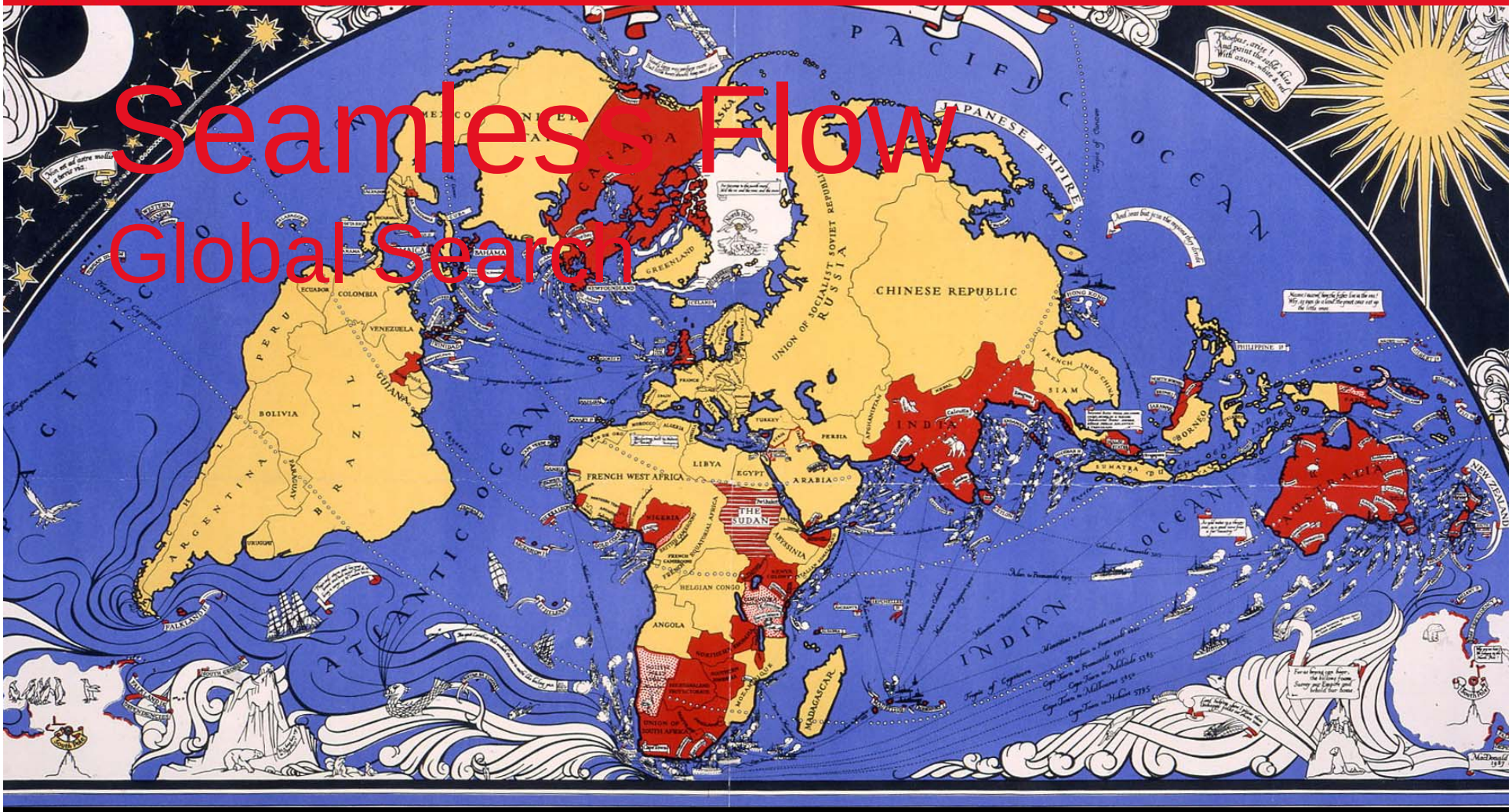
DROID


DROID (Digital Record Object Identification)
File Edit Identify Tools Help
File list
Status
File
C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\04small.j...
C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\110 ASCI...
C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\110 bina...
C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\1HS121...
C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\1HS181...
C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\3DFACE-...
Add Files
Remove Files
Remove All
Identify
Cancel
File 84 of 84 analysed
Identification results
File C:\Documents and Settings\Adrian\My Documents\Digital Archiving\File Formats\Test Files\Final Test Files\04small.jpg
PUID
Format
Version
Status
Warning
fmt/44
JPEG File Interchange Format
1.02
Positive (Specific Format)

Preservation Planning and action

- Identifying what action to take, when to take it, and how to enact it
  - Risk assessment
  - Technology watch
  - Impact assessment
  - Preservation plan generation
- Enacting the preservation plan and validating the results
  - Execute preservation plan
  - Validate results – characterise transformed objects and compare significant properties with source objects

the national archives



Global Search www.nationalarchives.gov.uk

Global Search

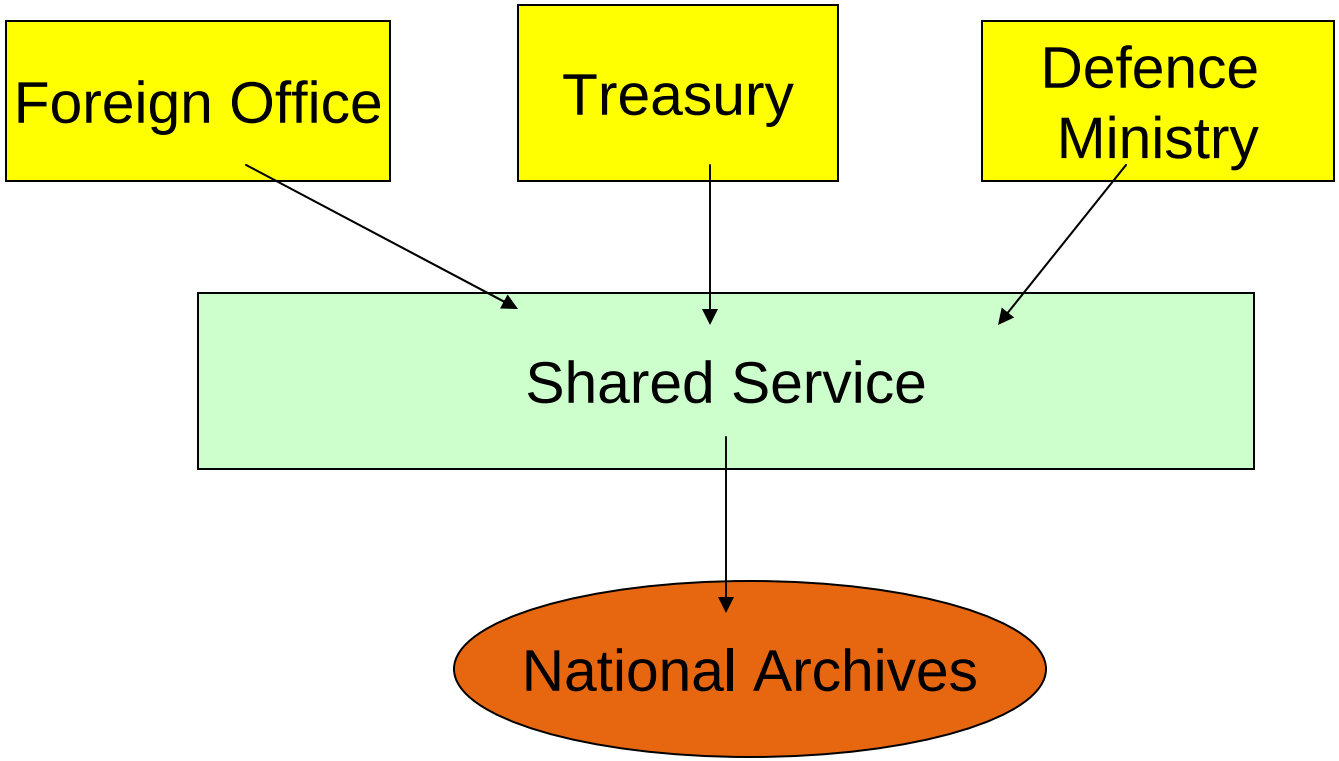
- Autonomy Search Engine

www.autonomy.com

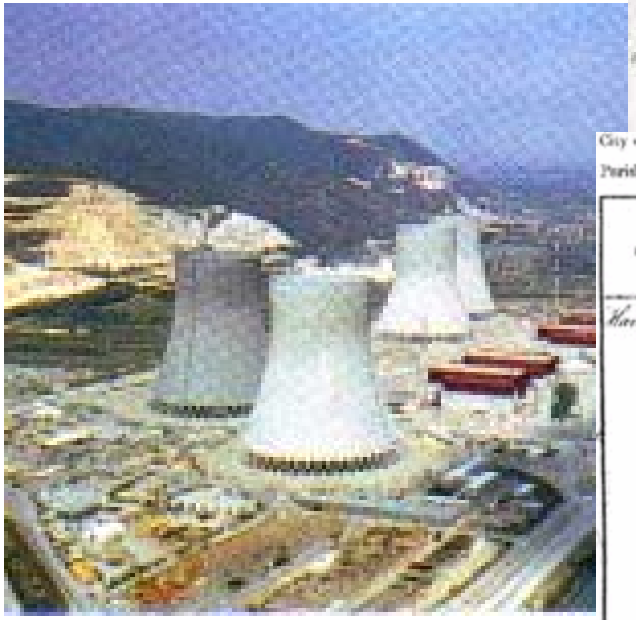
- Relevance ranking
- Results by Subject – Folksonomy
- Recommended links
- Personalised searching – My TNA

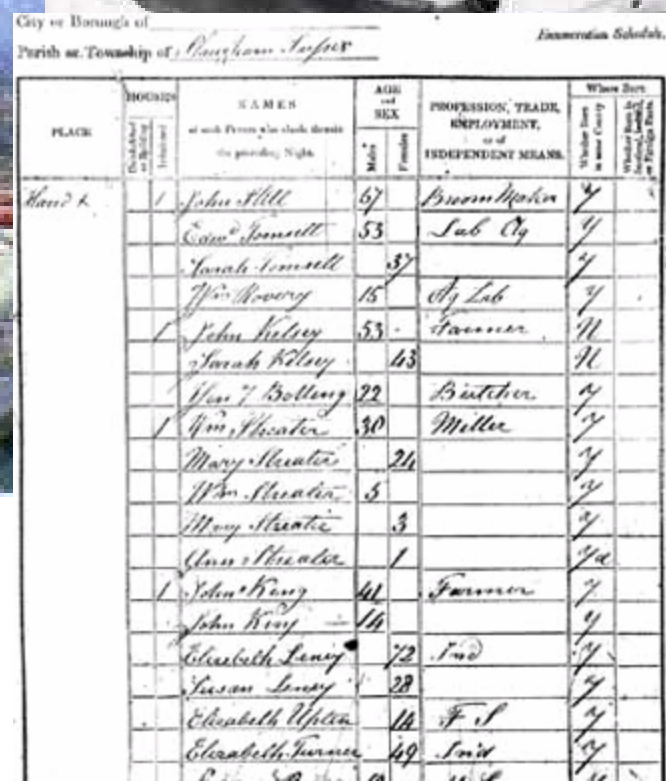
Shared services

- Two new initiatives :
  - Intermediate digital archives for records required for business purposes
  - Next generation EDRM systems


Foreign Office
Treasury
Defence Ministry
Shared Service
National Archives

Shared Service





City or Borough of

Parish or Township of Pingham Super

Enumeration Schedule.

PLACE	HOUSES Occupied or Building	HOUSES Inhabited	NAMES of each Person who shall dwell in the preceding Night	AGE and SEX Males	AGE and SEX Females	PROFESSION, TRADE, EMPLOYMENT, or of INDEPENDENT MEANS	Whose Born Whether Born in same County	Whether Born in Ireland, Scotland, or Foreign Parts
Hand &		1	John Hill	67		Broom Maker	Y	
			Edwd Tomsett	53		Lab Ag	Y	
			Sarah Tomsett		37		Y	
			Wm Rovery	15		Ag Lab	Y	
		1	John Kelsey	53		Farmer	N	
			Sarah Kelsey		43		N	
			Jen T Bolling	22		Butcher	Y	
		1	Wm Streater	30		Miller	Y	
			Mary Streater		24		Y	
			Wm Streater	5			Y	
			Mary Streater		3		Y	
			Ann Streater		1		Yd	
		1	John King	41		Farmer	Y	
			John King	14			Y	
			Elizabeth Leney		72	Ind	Y	
			Susan Leney		28		Y	
			Elizabeth Upton		14	F S	Y	
			Elizabeth Turner		49	Ind	Y	

the national archives
www.nationalarchives.gov.uk