

CoSA & Preservica Practical Digital Preservation 2015/16



Practical OAIS Digital Preservation Online Workshop Module 1

Practical Digital Preservation 2015/16

- **Welcome!**
- **PDP Online Workshops - with focus on records and email**
(Tuesdays 2-4pm Eastern)
 - Nov 10 2015 : Part 1
 - Nov 17 2015: Part 2
 - Mar 08 2016: Part 1
 - Mar 15 2016: Part 2
 - May 10 2016: Part 1
 - May 17 2016: Part 2
- **PDP “Hot Topic” Webinars**
(Tuesdays 2-3pm Eastern)
 - Achieving ISO Standards for your digital archive Oct 28 2015
 - Ingesting records from multiple sources and systems Dec 08 2015
 - Automating email archiving and preservation Feb 23 2016
 - New ways of providing public access to your archive Apr 26 2016
 - Real-world digital preservation and program/resources round-up Jun 21 2016



Sarah Grimm
Electronic Records Archivist
Wisconsin Historical
Society

<http://rc.statearchivists.org/Content/Electronic-Records/Education-Training/CoSA-Preservica-Practical-Digital-Preservation.aspx>

Workshop Objectives

Michael Hope
Preservica



Module 1 (today)

Understand the **fundamentals of Digital Preservation** - moving beyond the main acronyms and theory by illustrating topics with examples and demonstrations of practical real-world digital preservation workflows and processes

Module 2 (next Tuesday)

Understand **how Digital Preservation fits into the Information Governance lifecycle** – including content ingested from other systems (e.g. long-term records and emails) - as well as how to provide greater “transparency” through controlled access to information for internal and public users

Preservica Company Milestones



Preservica Users (partial list)

National & Pan-National



Libraries, Museums & Education



State & Government



Business & Corporate

Practical OAIS Digital Preservation



Jack O'Sullivan

Technical Consultant, Preservica

Agenda : Module 1

Digital Preservation Fundamentals

Module 1

Session 1

- Why do we need Digital Preservation?
- The fundamentals of preserving digital content

Questions

Session 2

- Understanding Metadata, Fixity and File Characterization
 - including example demonstration

Questions

- Next Steps and Close

Agenda : Module 2 (next week)

DP in Information Governance Lifecycle

Module 2

Recap

- Recap on Module 1 – and Intro to Module 2

Session 3

- Preservation Planning and Action
 - Ingest and Preservation long-term records from SharePoint

Questions

Session 4

- Controlling access to digital content – including practical demonstration
- Ingesting and Preserving Complex formats
 - Website harvesting (WARC files)
 - Emails and Attachments – demonstration of record classification & action

Questions & Close

- Next Steps and Close

Workshop demonstrations

The live practical examples use the Preservica Preservation system

Other systems are available:



DSPACE

Open Source
Digital Repository Application



Session 1: Part 1

Why do we need digital preservation? (exploring the problem)

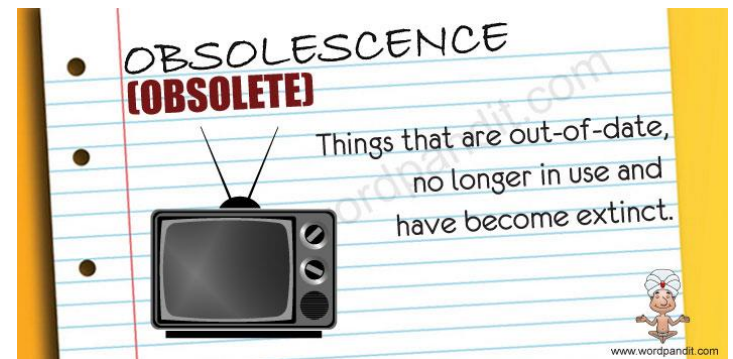
*HOW TO SAVE YOUR DIGITAL WORK FOR
THE POSTERITY?*



What is Digital Preservation?

***Digital preservation** is a process to ensure that digital information of continuing value remains accessible and usable.*

It involves planning, resource allocation, and application of preservation methods and technologies.



What are the Risks and Consequences?

Institutional risks posed by poor preservation strategy:

- Failure to meet mandate
- Legal & financial liability
- Regulatory fine
- Reputational damage
- Cost of re-digitization
- Born-digital material lost forever
- Loss of knowledge



But also benefits!

Institutional benefits of a good preservation strategy:

- Make digital information more accessible (transparent)
- Enable future reuse of information
- Preserve heritage or corporate memory (brand value)
- Save time searching for information
- Respond quickly to legal and compliance challenges
- Retire existing legacy systems (saving cost)
- Support digital ways of working



Digital Records: What's Important?

- Authenticity
- Provenance

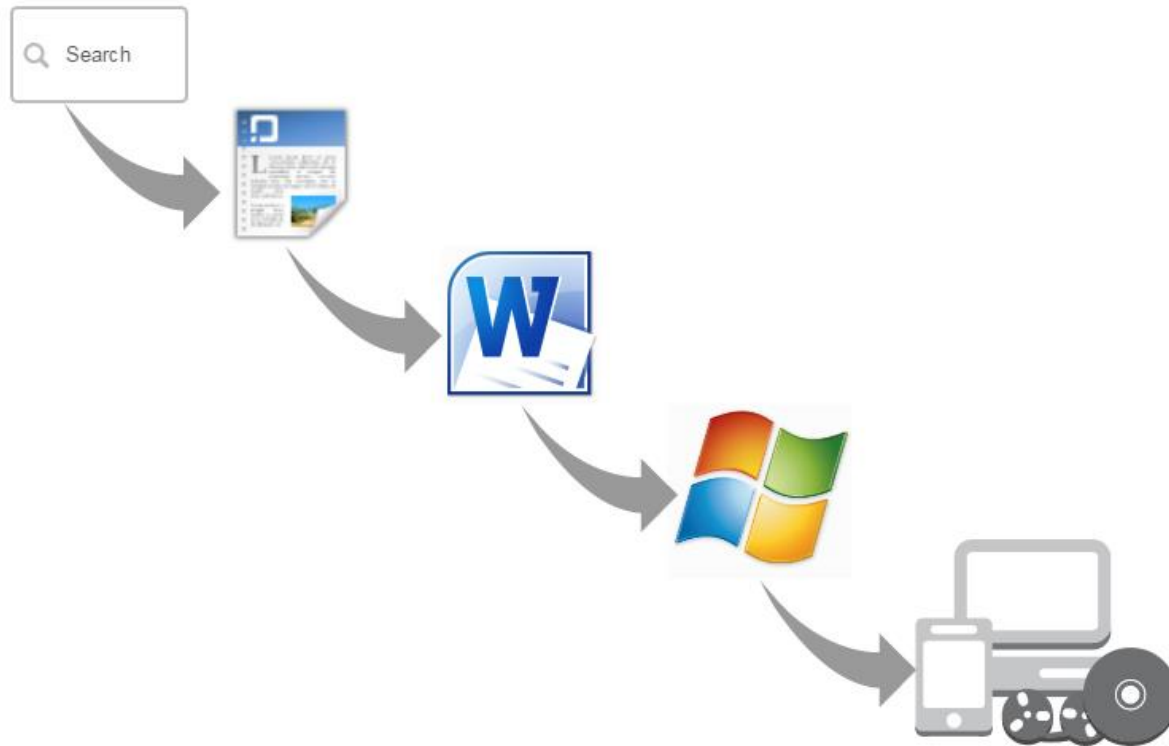
Today's sessions

- Preservation
- Retention & Disposition
- Access, Security & Privacy

Next sessions

The Fragility of Digital Content

- Information does not exist in isolation



- Each part can be obsolete within information's lifetime

Other Long Term Preservation Problems

Vital information
on removable or
unmanaged media



Lack of metadata to
interpret the data

1	3.4	2.079
0	5.1	7.386
0	1.8	3.826
...

Column headings incl. units?
Does 1 = male or female?
Temperature experiment run at?
When was it carried out?

Without organisation,
you cannot find
your information

my critical information

Your search - my critical information - did not match any documents.

Suggestions:

- Make sure all words are spelled correctly.
- Try different keywords.
- Try more general keywords.

Why not just Print to Paper?

- Lose some key advantages:
 - Content searching
 - Easy to copy without loss
 - Variable content (e.g. database, tracked changes)
 - All behaviour
 - Multiple views
- Some digital records just can't be printed!
 - CAD models, Video, Audio, GIS etc.

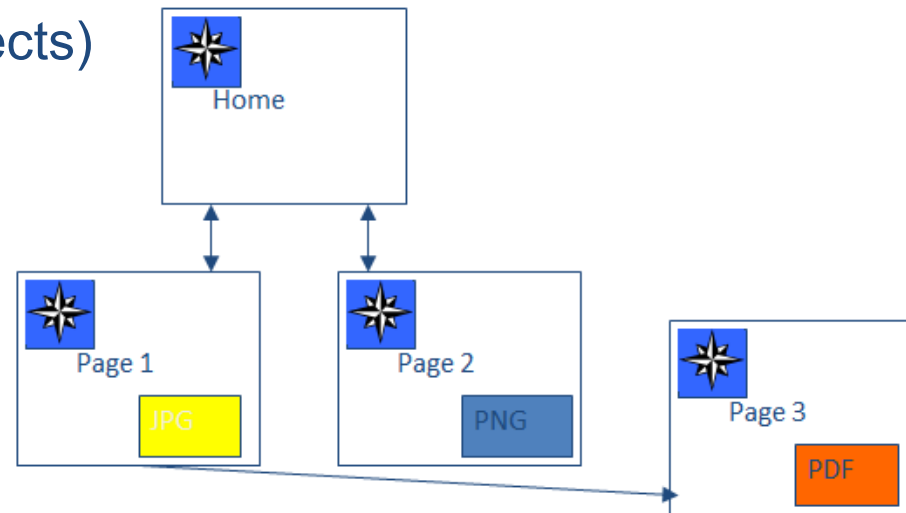


Complexity of Digital Content

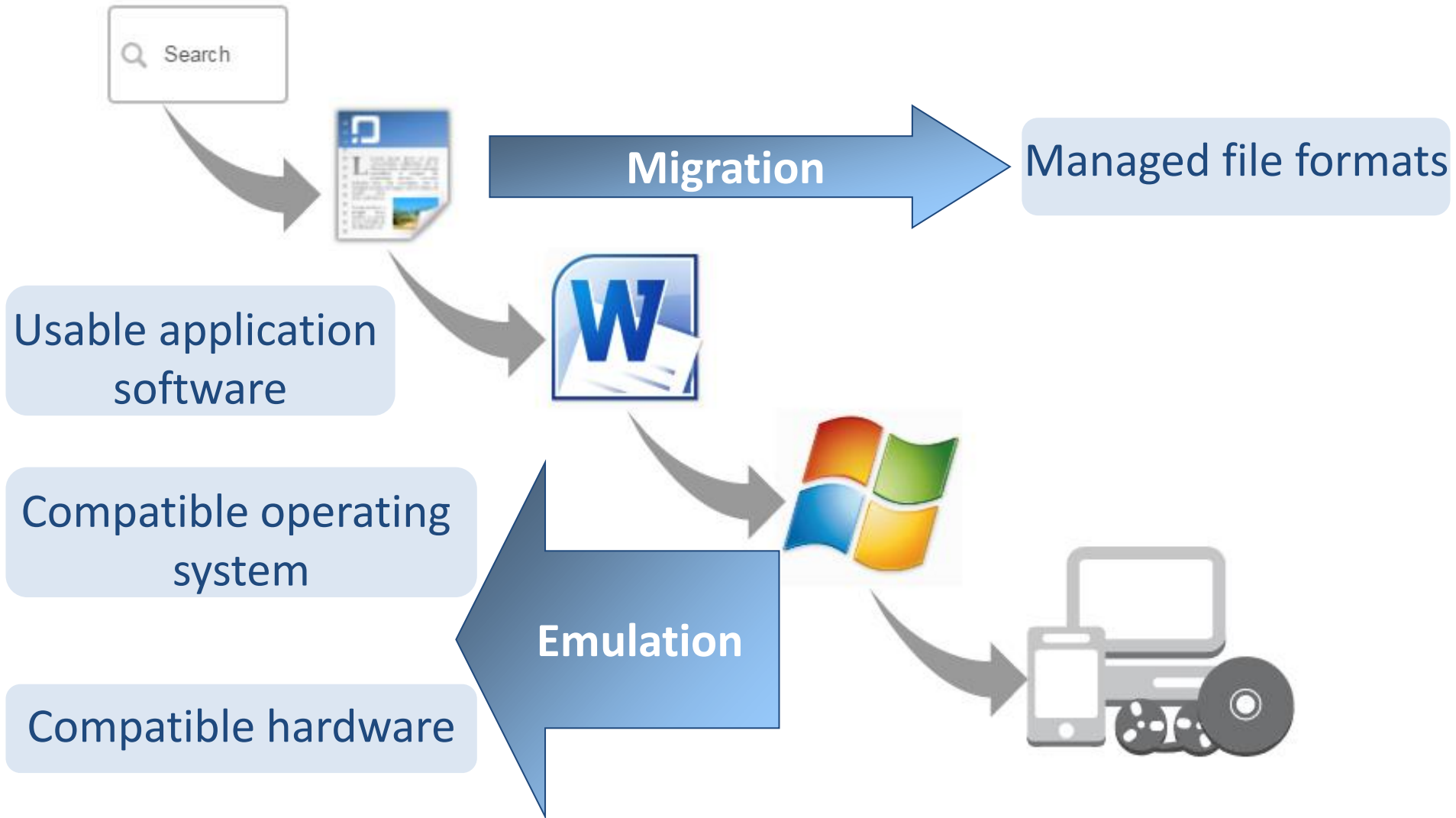
Example: Web page

- Really two structures:
 - “Physical” (digital objects)
 - Understood by machines
 - Technology dependent
 - May need to be migrated
 - “Conceptual” (information objects)
 - Understood by humans
 - Technology independent
 - Needs to be preserved

Home.html
Style.css
Logo.gif
Page1/Page1.html
Page1/Image.jpg
Page2/Page2.html
Page2/Image.png
Page3/Page3.html
Page3/Document.pdf



Preservation Strategies



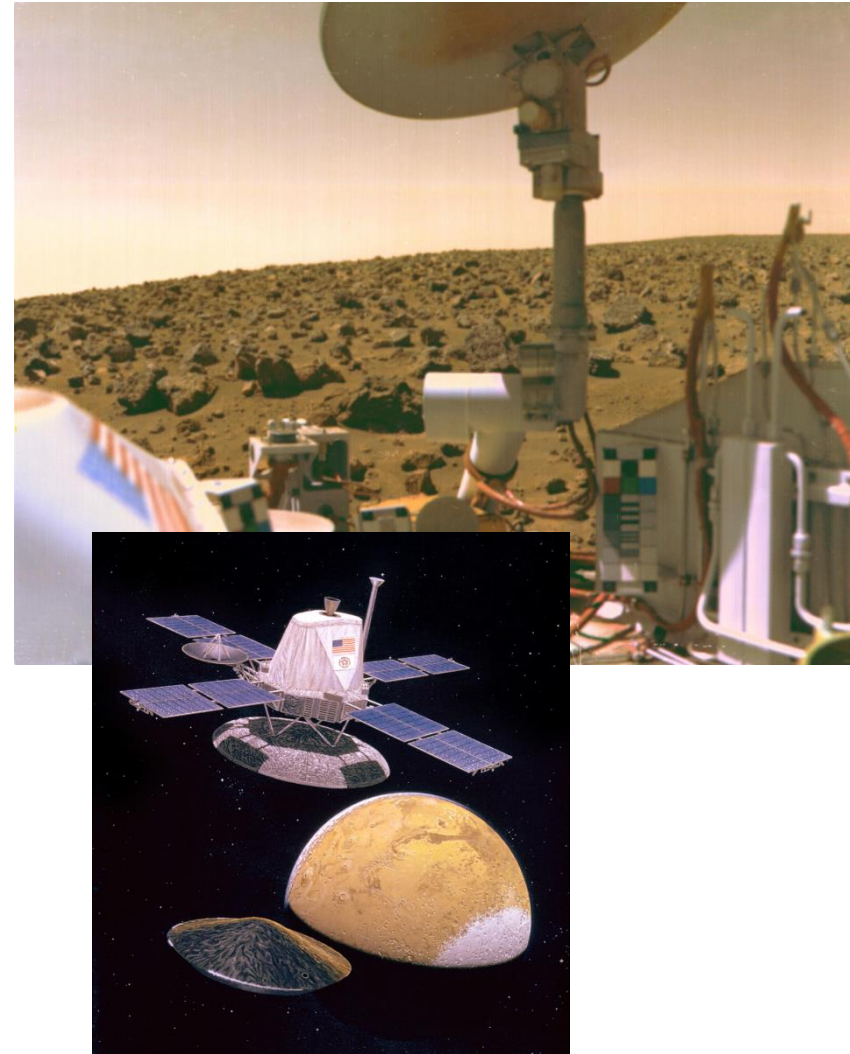
Case Study: Andy Warhol's Amiga Images

- Created in 1985 by Andy Warhol
- Commissioned by Commodore to promote the Amiga 1000
- State of the art technology
 - Pre-release hardware & software
- Unreadable at discovery in 2011



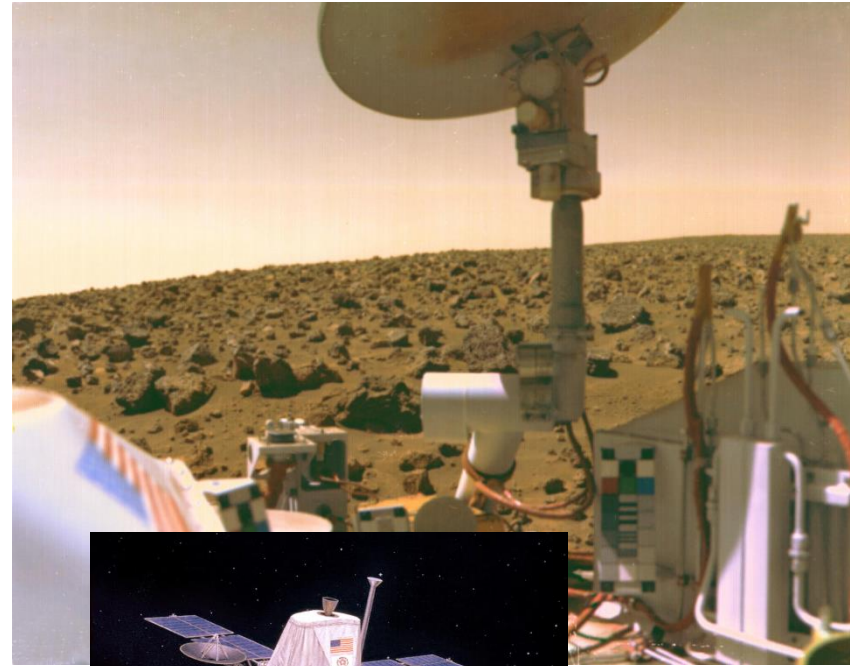
Viking Mars Lander Missions (1)

- NASA's 1975 Viking Lander probes collected data from the Martian surface
- Datasets were compiled and stored to magnetic tape for long term reuse
- Tapes were stored in climate controlled environments



Viking Mars Lander Missions (2)

- Tapes became brittle and cracked
- The formats could not be decoded in the 1990s
- Ultimately retyped using old printouts!
- Space community proposed what would become OAIS

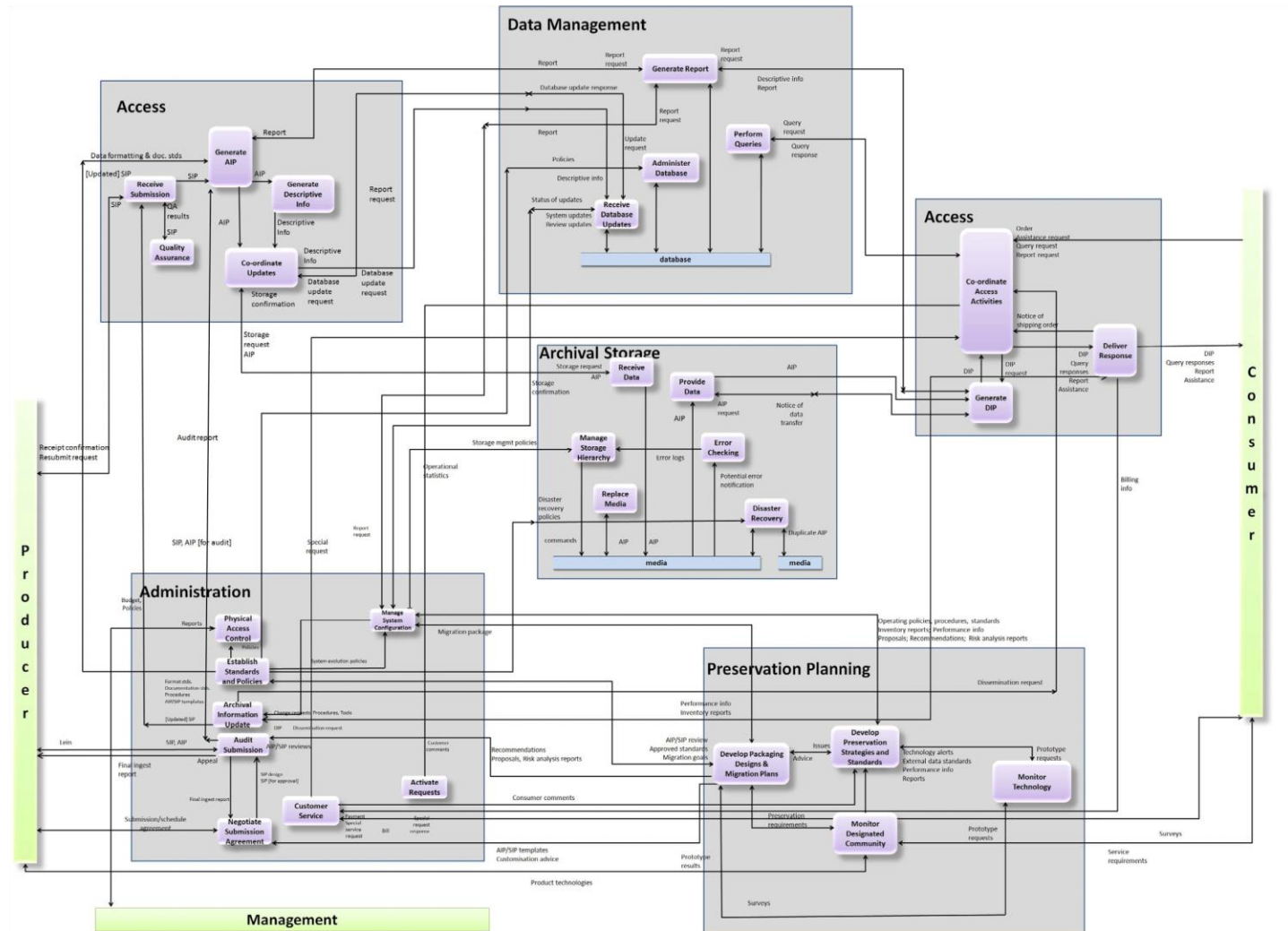


OAIS Model

CCSDS 650.0-P-1.1

A-2

August 2009



DRAFT CCSDS RECOMMENDED STANDARD FOR AN OAIS REFERENCE MODEL

Figure A-1: Composite of Functional Entities

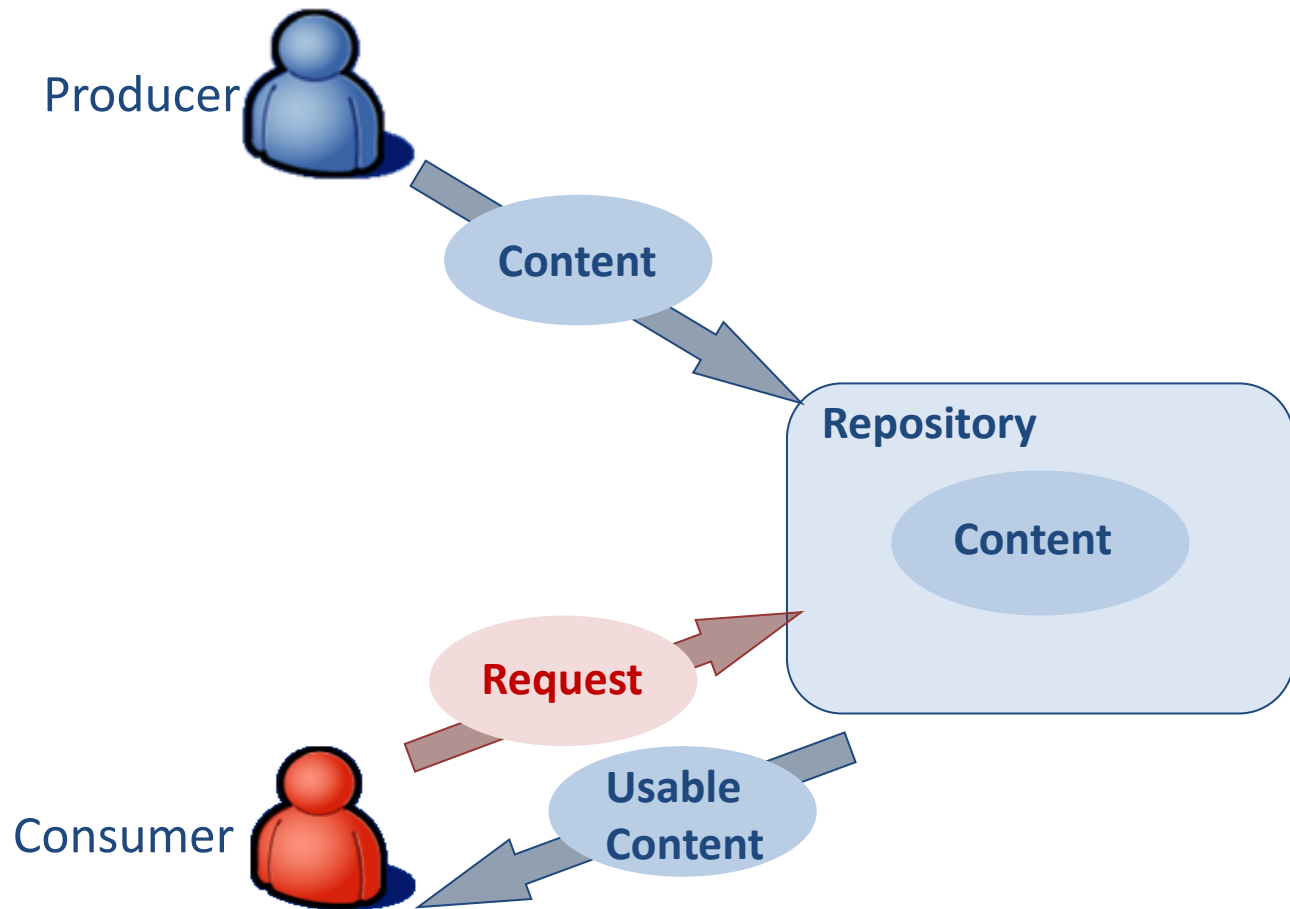
Session 1: Part 2

The fundamentals of Digital Preservation (exploring the solution)



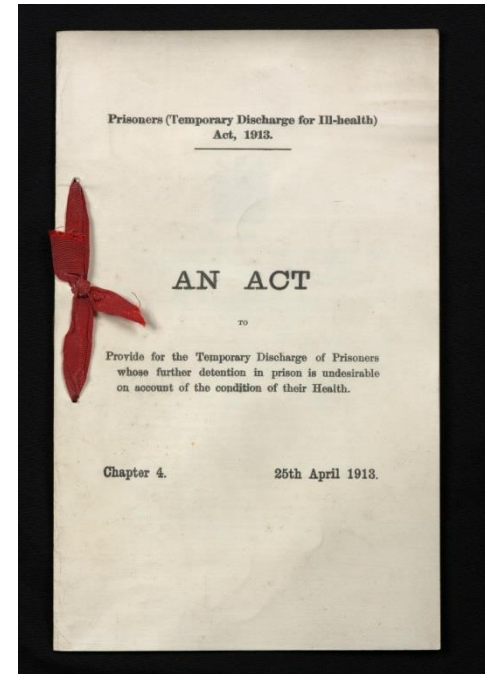
The Repository

- We need a preservation capable digital repository



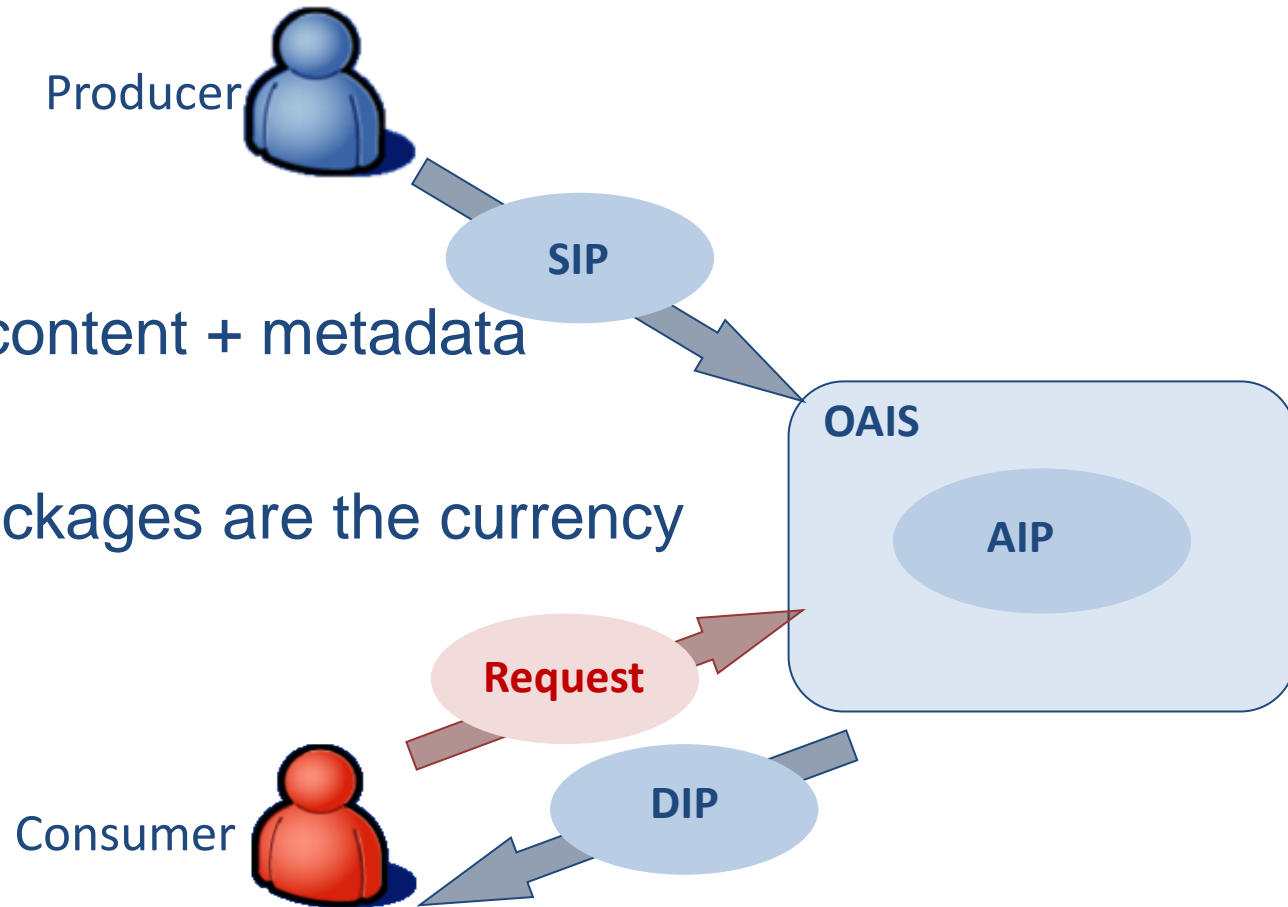
Content Types

- Digitised
 - Historic Materials
 - Potentially expensive to reproduce (but often possible)
- Born Digital
 - Records Management systems
 - Email
 - Web
 - Impossible to reproduce



OAIS Terms

- Open Archival Information System (ISO 14721)



- Information = content + metadata
- Information Packages are the currency

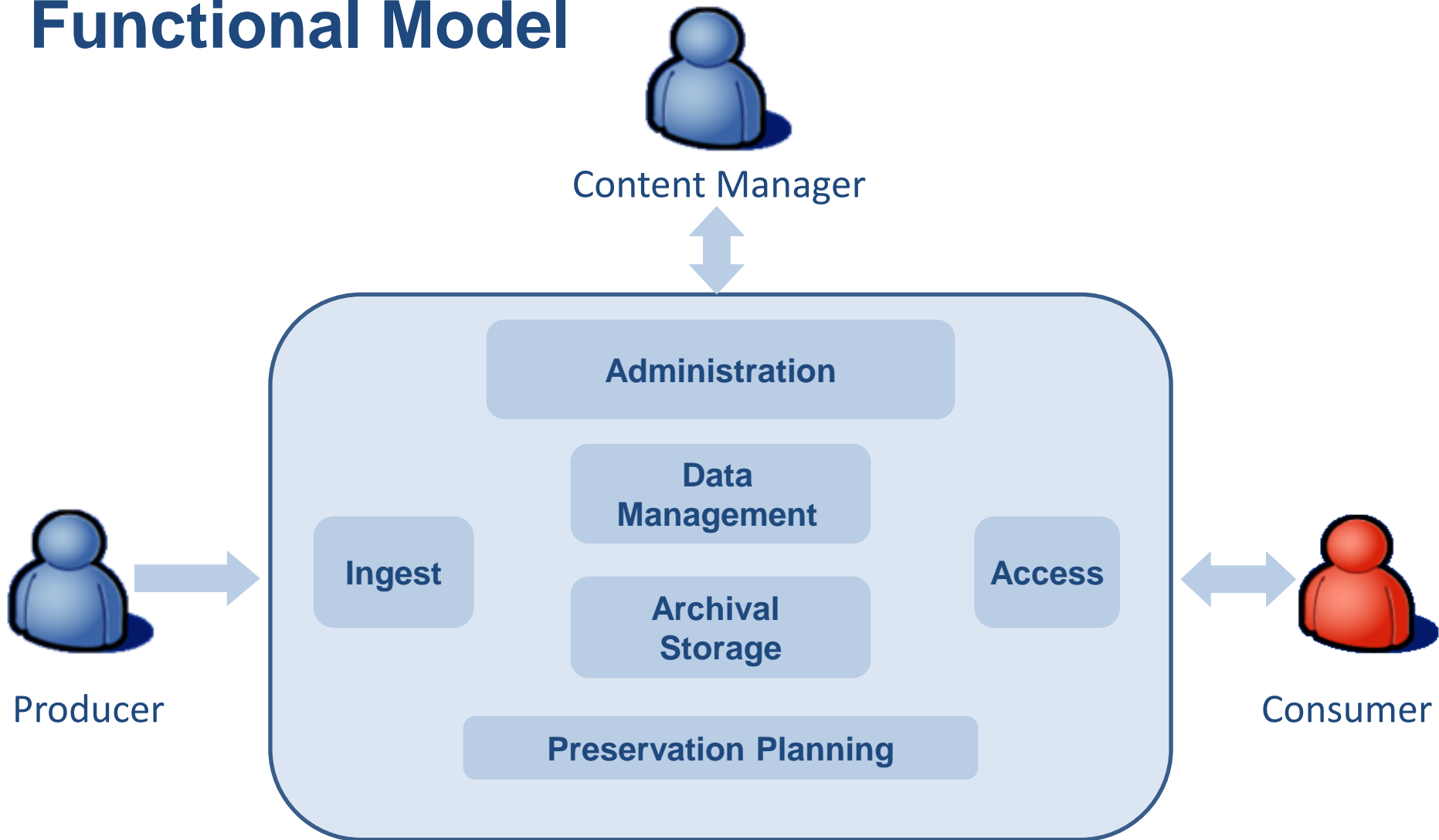
OAIS Information Model

- **SIP:** (Submission Information Package)
 - The information supplied to the repository by a content producer
- **AIP:** (Archival Information Package)
 - The information stored within the repository
- **DIP:** (Dissemination Information Package)
 - The information supplied to a user by the repository to satisfy a request
- These **may** be mapped 1:1

OAIS Actors

- Content Producers
 - The people generating the information and submitting it for long term preservation
- Content Consumers
 - The “designated community”, people who want to use the information
- Content Managers
 - Archivists, collection managers, preservation manager

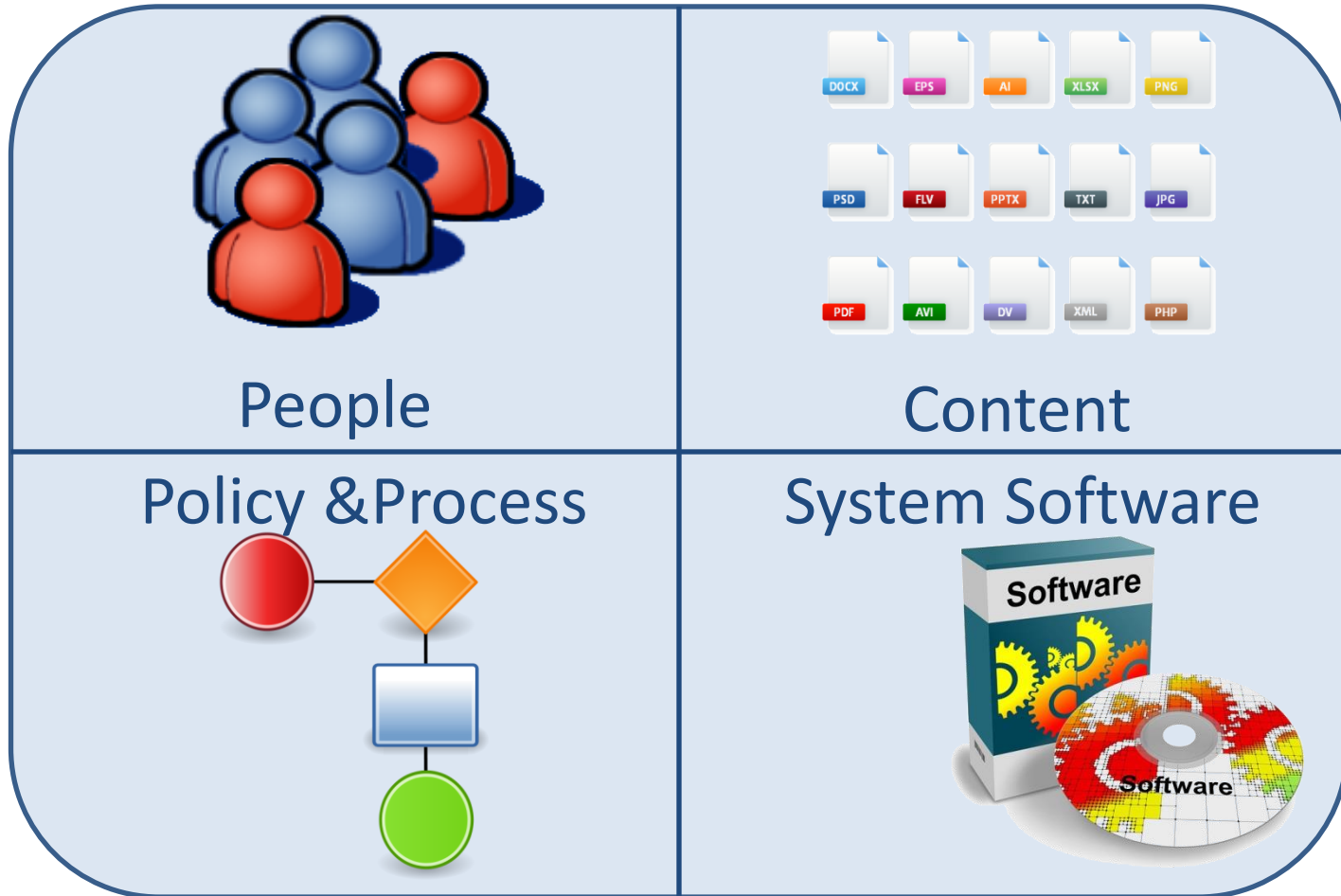
Functional Model



How do I know my system complies?

- OAIS ISO 14721
- Trusted Digital Repository (TDR)
 - ISO 16363
 - Requirements detailed as a checklist of metrics
 - *e.g. The repository shall assign and maintain persistent identifier of the AIP and its components so as to be unique with the context of the repository*
- Trustworthy Repositories Audit & Certification (TRAC)
 - ISO 16919
 - The process of auditing the repository for compliance to TDR

A Truly Trusted Repository



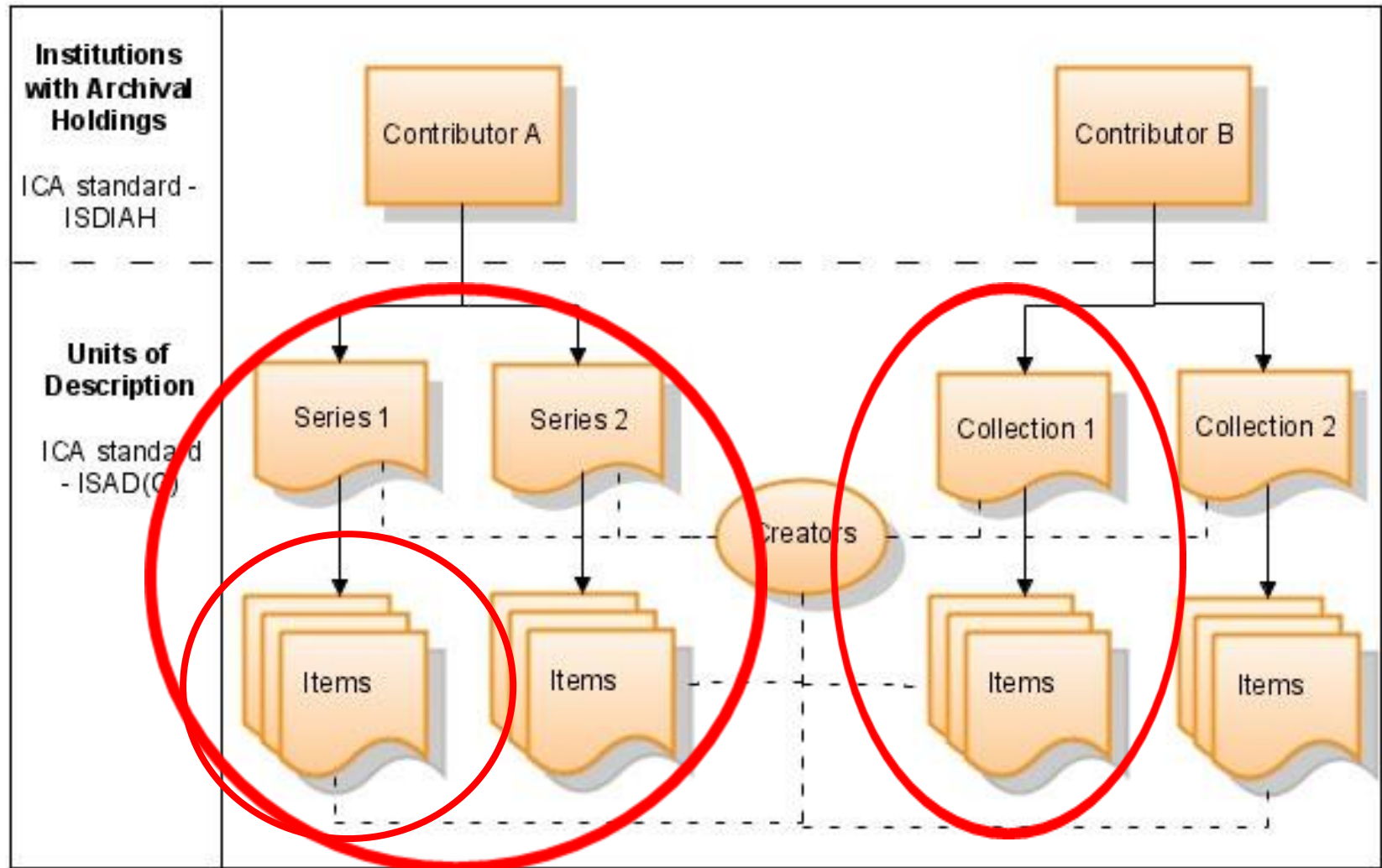
Questions?



Session 2

Understanding Metadata, Fixity and File Characterisation

Record Structure



DEMO : RECORD STRUCTURE AND HIERARCHY



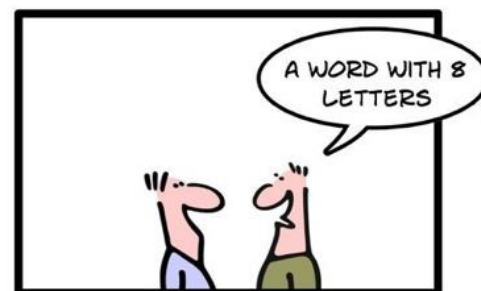
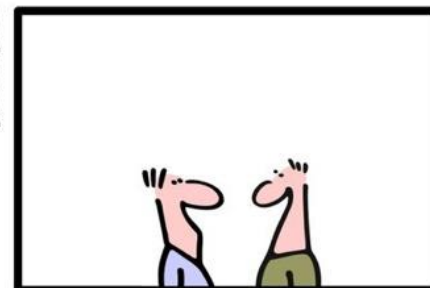
Metadata

- Descriptive/Structural:
 - Context to make the information usable
 - Usually created by and for humans
- Access Rights:
 - What to make available and to whom
- Preservation Description Information:
 - Proving authenticity and provenance
- Technical:
 - Enable long term preservation
 - Risk assessment & Validation

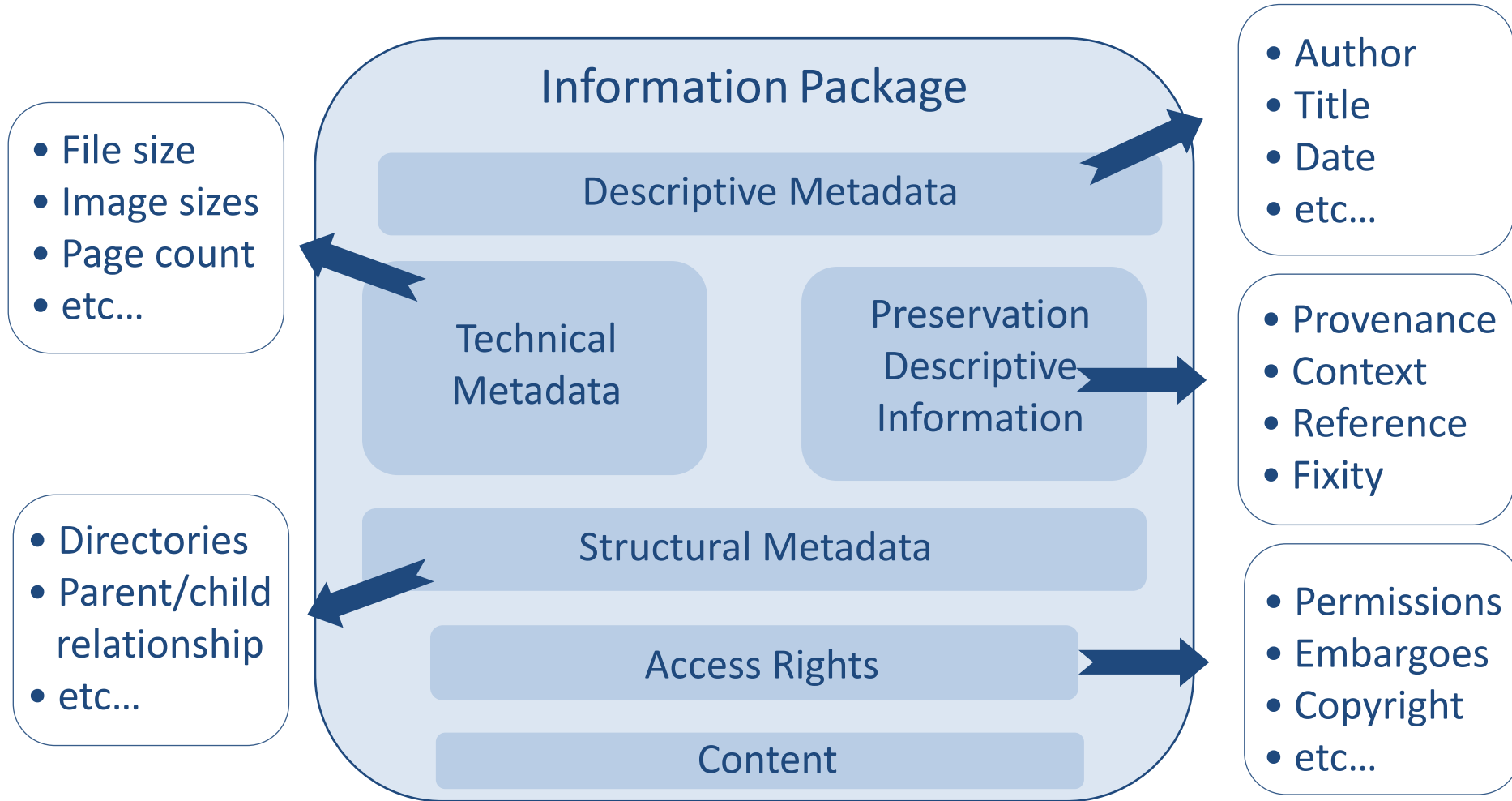
*SIMPLY EXPLAINED:
METADATA*



geek & poke



The Information Package



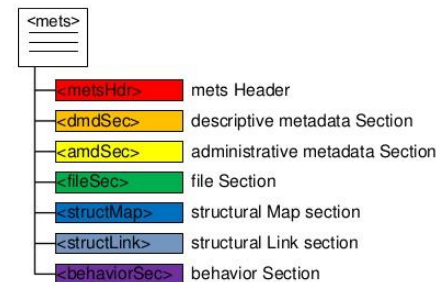
Descriptive Metadata

The Archival packages will contain descriptive metadata, this needs to be created by humans.

There is no one standard way of including descriptive metadata

Lots of Standards to choose

- ISAD(G) (EAD)
- Dublin Core
- METS
- MODS
- MARC
- ...



Seven Major Sections of a METS Document

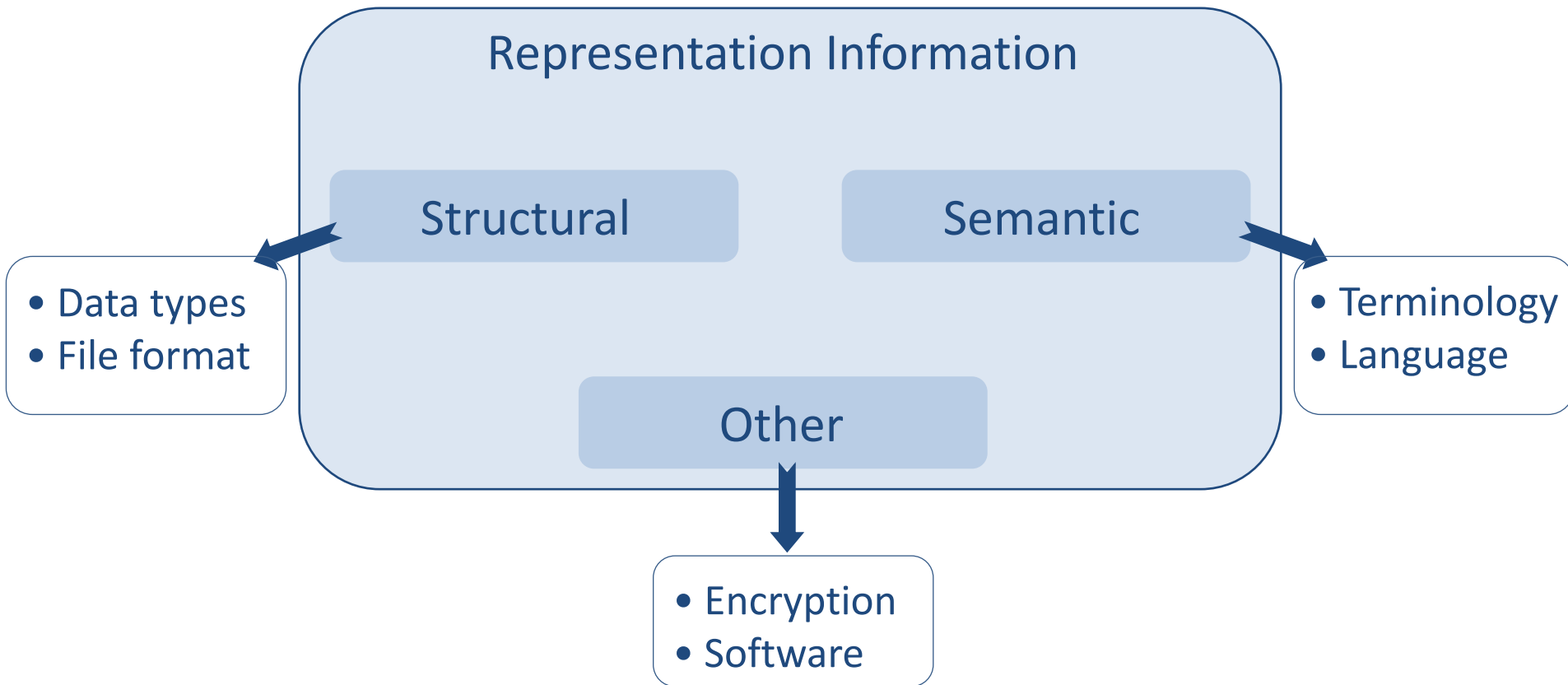
SOURCE: Bredenberg, K. (2011, June). METS. Communicating Archival Metadata conference and workshops in Stockholm 28 - 30 June 2011



OAIS Representation Information

Lives outside the package

- Tells us how to convert the binary representation of the content into something meaningful

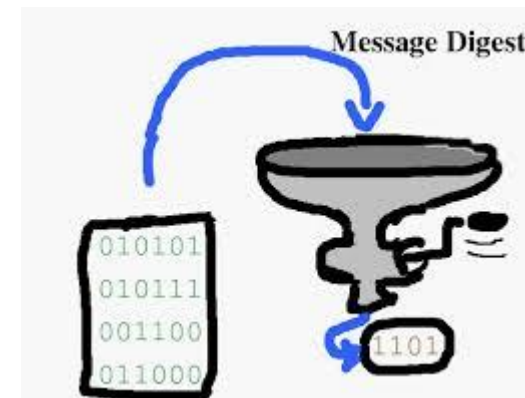


DEMO: FILE FORMAT REPRESENTATION INFORMATION - REGISTRY



Fixity

- Specific to Digital Material
- Measure to guarantee invariance of the bits
 - “Fingerprint”
- Sometimes referred to as:
 - Hash (or Cryptographic Hash)
 - Digest (or Message Digest)
 - Checksum
- Popular algorithms MD5 & SHA



FIXITY EXAMPLE

Which is Authentic?

The image displays two side-by-side Adobe Reader windows. The left window, titled 'Standard Services Contract Real.pdf', shows the original contract. The right window, titled 'Standard Services Contract Doctored.pdf', shows a modified version. Red boxes and lines highlight specific changes in the doctored version:

- Section 1.1.2.2 (Start Date):** The original text is 'The Supplier agrees to collect an initial shipment of records from the business address of the buyer on Monday 8th October 2007.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years, unless and until terminated'.
- Section 1.1.2.3 (Recurring Collections):** The original text is 'The Supplier agrees to collect further shipments on the first Monday of each month, or the first business day thereafter.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years unless and until terminated'.
- Section 1.2.1 (Charges):** The original text is 'The Buyer agrees to pay an initial fee of \$1,000 by 29th October 2007.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years unless and until terminated'.
- Section 1.2.1.1 (Initial Payment):** The original text is 'The Buyer agrees to pay a recurring monthly charge of \$1,000 by the 28th day of each month, or the first business day thereafter.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years unless and until terminated'.
- Section 1.2.1.2 (Recurring Payment):** The original text is 'The Buyer agrees to pay a recurring monthly charge of \$1,000 by the 28th day of each month, or the first business day thereafter.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years unless and until terminated'.
- Section 1.2.1.3 (Escalation):** The original text is 'The Buyer agrees that this monthly charge will be in effect for the first year of this contract, thereafter it is to be reviewed and agreed by both parties on an annual basis, the first revision to be agreed by the 3rd Day of October 2007.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years unless and until terminated'.
- Section 1.3 (Contract Period):** The original text is 'This agreement shall be effective from the date it is made and shall continue in force for a period of five (5) years from the date it is made, and thereafter for successive five (5) year terms, unless and until terminated by one year prior notice in writing by either party.' The doctored version changes this to 'the date it is made and shall continue in force for a period of five (5) years unless and until terminated'.

Both contracts include a signature line at the bottom: 'Signed by: _____ On behalf of the Provider' and 'Signed by: _____ On behalf of the Buyer'.

Check the Fingerprints

Standard Services Contract Real.pdf - Adobe Reader

File Edit View Window Help

Standard Services Contract

This agreement was made between The General Services Company Inc (the "Supplier") and the Commonwealth of West Atlanta (the "Buyer") on the 3rd day of October 2007.

Terms

1.1 Services

1.1.1 Provision of Services

The Supplier agrees to deliver services relating to the collection and storage of records.

1.1.2 Delivery of Services

1.1.2.1 Nature of Service

The Supplier agrees to collect records from the business address of the Buyer on a periodic basis; such records to be transferred to a secure storage facility provided by the Supplier.

1.1.2.2 Start Date

The Supplier agrees to collect an initial shipment of records from the business address of the buyer on Monday 8th October 2007.

1.1.2.3 Recurring Collections

The Supplier agrees to collect further shipments on the first Monday of each month, or the first business day thereafter.

1.2 Payment

1.2.1 Charges

1.2.1.1 Initial Payment

The Buyer agrees to pay an initial fee of \$1,000 by 29th October 2007.

1.2.1.2 Recurring Payment

The Buyer agrees to pay a recurring monthly charge of \$1,000 by the 28th day of each month, or the first business day thereafter.

1.2.1.3 Escalation

The Buyer agrees that this monthly charge will be in effect for the first year of this contract, thereafter it is to be reviewed and agreed by both parties on an annual basis, the first revision to be agreed by the 3rd Day of October 2008.

1.3 Contract Period

This agreement shall be effective from the date it is made and shall continue in force for a period of five (5) years from the date it is made, and thereafter for successive five (5) year terms, unless and until terminated by one year prior notice in writing by either party.

Signed by: _____
On behalf of the Provider On behalf of the Buyer

MD5: 31d49ed2b739f0ae0e639ad1a7bba162
SHA-1: f0eb43942966b7dfd6c3ec236a074652c6810d0b

Standard Services Contract Doctored.pdf - Adobe Reader

File Edit View Window Help

Standard Services Contract

This agreement was made between The General Services Company Inc (the "Supplier") and the Commonwealth of West Atlanta (the "Buyer") on the 3rd day of October 2007.

Terms

1.1 Services

1.1.1 Provision of Services

The Supplier agrees to deliver services relating to the collection and storage of records.

1.1.2 Delivery of Services

1.1.2.1 Nature of Service

The Supplier agrees to collect records from the business address of the Buyer on a periodic basis; such records to be transferred to a secure storage facility provided by the Supplier.

1.1.2.2 Start Date

The Supplier agrees to collect an initial shipment of records from the business address of the buyer on Monday 8th October 2007.

1.1.2.3 Recurring Collections

The Supplier agrees to collect further shipments on the first Monday of each month, or the first business day thereafter.

1.2 Payment

1.2.1 Charges

1.2.1.1 Initial Payment

The Buyer agrees to pay an initial fee of \$1,000 by 29th October 2007.

1.2.1.2 Recurring Payment

The Buyer agrees to pay a recurring monthly charge of \$1,000 by the 28th day of each month, or the first business day thereafter.

1.2.1.3 Escalation

The Buyer agrees that this monthly charge will be in effect for the first year of this contract, thereafter it is to be reviewed and agreed by both parties on an annual basis, the first revision to be agreed by the 3rd Day of October 2008.

1.3 Contract Period

This agreement shall be effective from the date it is made and shall continue in force for a period of five (5) years from the date it is made, and thereafter for successive five (5) year terms unless and until terminated by one year prior notice in writing by either party.

Signed by: _____
On behalf of the Provider On behalf of the Buyer

MD5: 078dd7ddff28b1bc4801c737d2ff1dee
SHA-1: be5d0f673fa884df3f07cdf4e7652fa333aec9a5

FILE CHARACTERISATION

What is our Information Object?

- What does the object claim to be?
- Do we believe it?

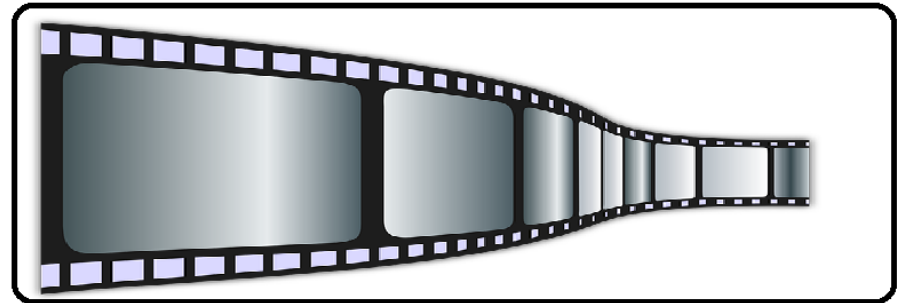
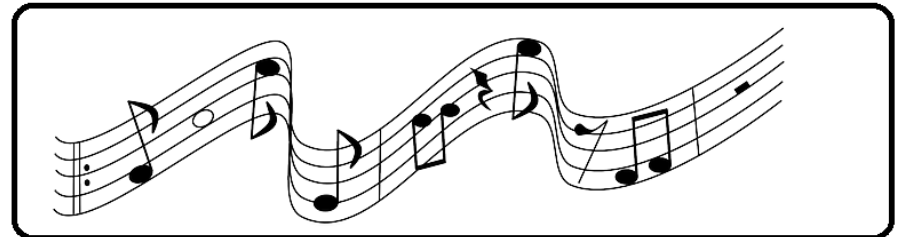
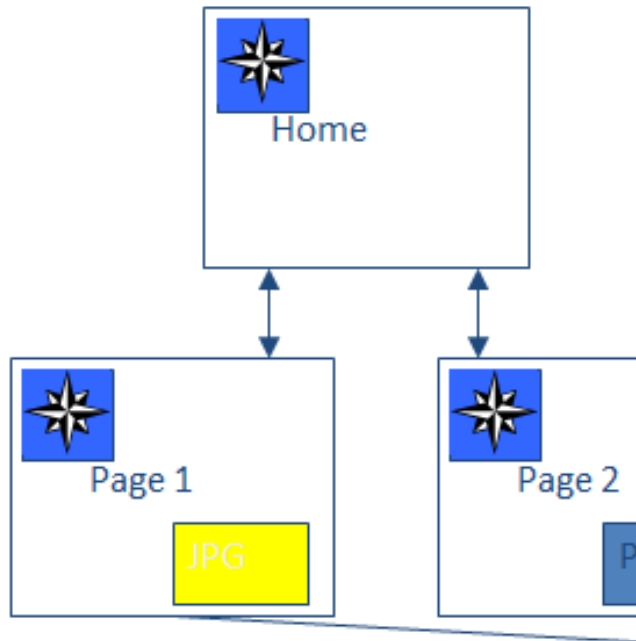


What else do we know about it?

- What makes our object different from other objects of the same type?



Above and below the file level



So what?

- Data is not information!
- Data is the bit-stream
 - The conveyor of information
- Information is what we glean from this data
- The intrinsic characteristics are a proxy for information

DEMO: INGEST & CHARACTERISATION



Questions ?



Next Steps

- **Module 2:** 2-4pm Eastern, Tuesday Nov 17 2015
- Next webinar: 2-3pm Eastern , Tuesday December 08 2015
 - Ingesting in multiple formats and from multiple systems
- Achieving a Step Change in Digital Preservation Capability
<http://preservica.com/resource/essential-guide-achieving-step-change-digital-preservation-capability/>
- Safeguarding your vital long-term electronic records
<http://preservica.com/resource/electronic-records-preservation/>
- www.preservica.com/resources



Next Steps

Workshop Objectives

Understand the **fundamentals of Digital Preservation** - moving beyond the main acronyms and theory by illustrating topics with examples and demonstrations of practical real-world digital preservation workflows and processes

- **We value your feedback ;-)**
- **Please complete the short evaluation....**

Thank you!

www.preservica.com

info@preservica.com

@preservica

@dPreservation

www.statearchivists.org/