Small steps and lasting impact: making a start with preservation

or

It’s not all NASA

Patricia Sleeman
Digital Archives and Repositories
University of London Computer Centre
(ULCC)

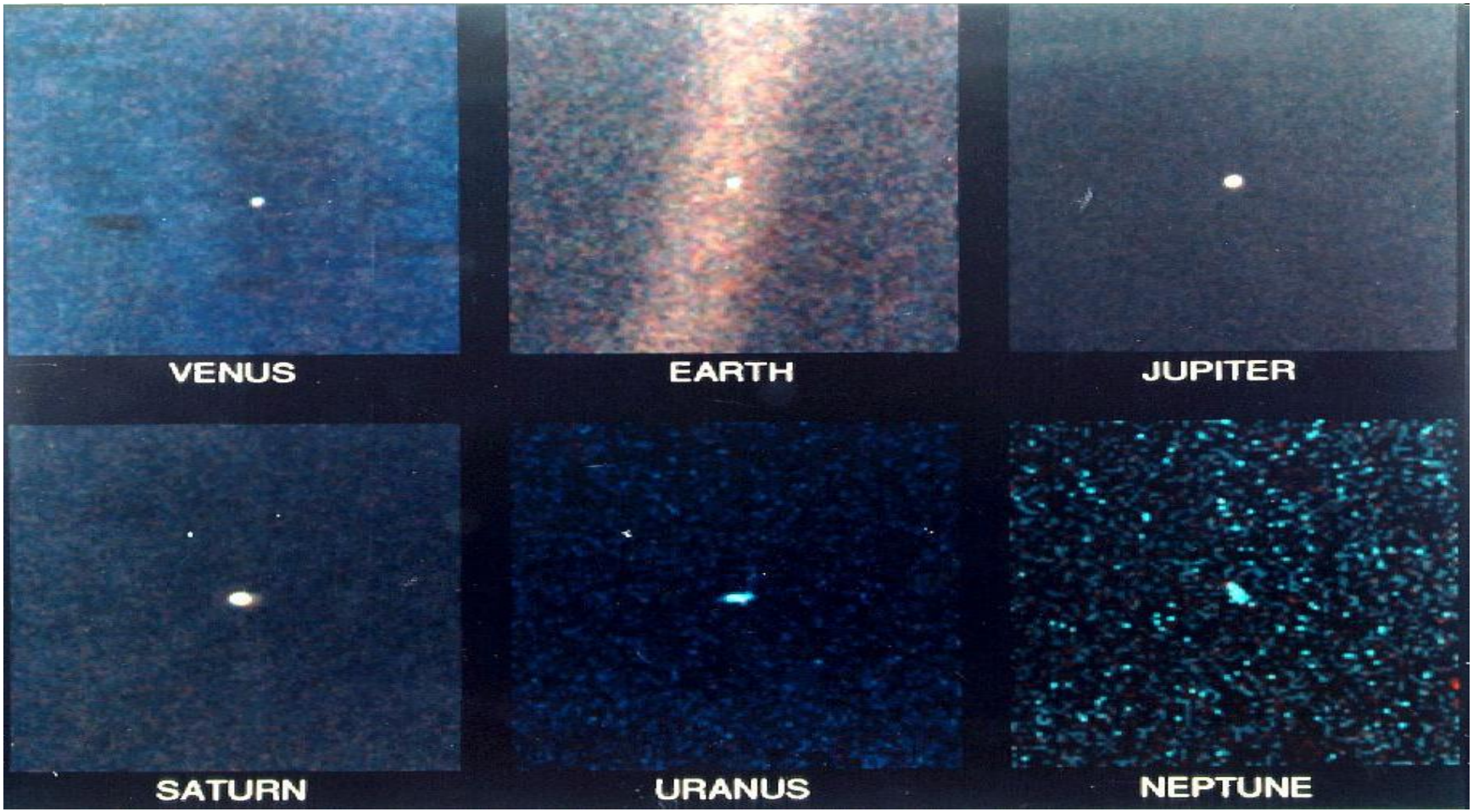
ulcc
University of London Computer Centre

Overview of session

- Not all NASA models
- But it helps – OAIS
- You could be using it already
- Some preservation strategies
- Some useful tools
- How and where do I start!


Flickr Accidentally Wipes Out Account: Five Years And 4,000 Photos Down The Drain - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://techcrunch.com/2011/02/02/flickr-accidentally-wipes-out-account-five-years-and-4000-photos-d
facebook
Most Visited
Digital Archives and R...
YouTube home
YouTube - Greensleev...
http://www.freeguita...
YouTube home
nasa site of the day - ...
(3) Twitter /...
Flickr ...
Turbulence ...
Turbulence ...
Digital Archi...
It's not all N...
When Irish ...
Al Jazeera E...
(53) Twitter...
Tech
Gadgets
Mobile
Enterprise
GreenTech
CrunchBase
TechCrunch TV
The Crunchies
More
Google Custom Search
Search
TechCrunch
Shirts
Boots
Sneakers
Tot bags
Sunglasses
Handbags
GROUPON
Shopping
up to 70%
To the Deals
What's Hot: Android
Apple
Facebook
Google
Groupon
Microsoft
Twitter
Zynga
Classics
Subscribe:
Flickr Accidentally Wipes Out Account: Five Years And 4,000 Photos Down The Drain
Robin Wauters
10 hours ago
Like 1K
Buzz 47
Tweet 1,891
22
Digg
154 Comments
flickr from YAHOO!
Home
The Tour
Sign Up
Explore
Upload
bindermichi's photostream
pro
Collections
Sets
Galleries
Tags
People
Archives
Favorites
Profile
bindermichi doesn't have anything available to you.
Got a tip? Building a startup? Tell us
ALL YOU CAN EAT
in London
up to
70%
cheaper
Here
GROUPON
Most Popular
Now
Commented
Facebook
Find: phili
Next
Previous
Highlight all
Match case
Phrase not found
Transferring data from api.conduit.com...
start
Flickr Accidentally ...
WL
2 Microsoft Offic...
2 Microsoft Offic...
Microsoft Office Pi...
21:39

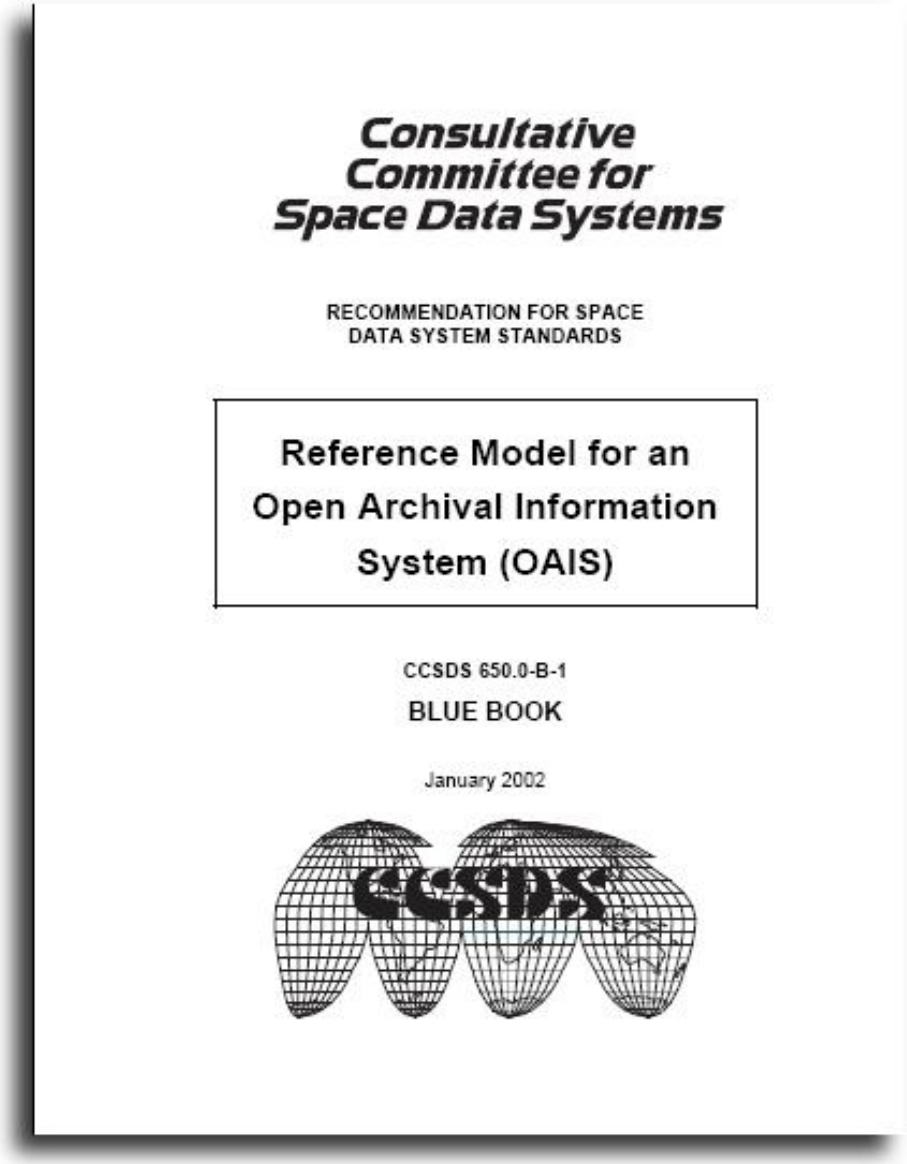
NASA family portrait



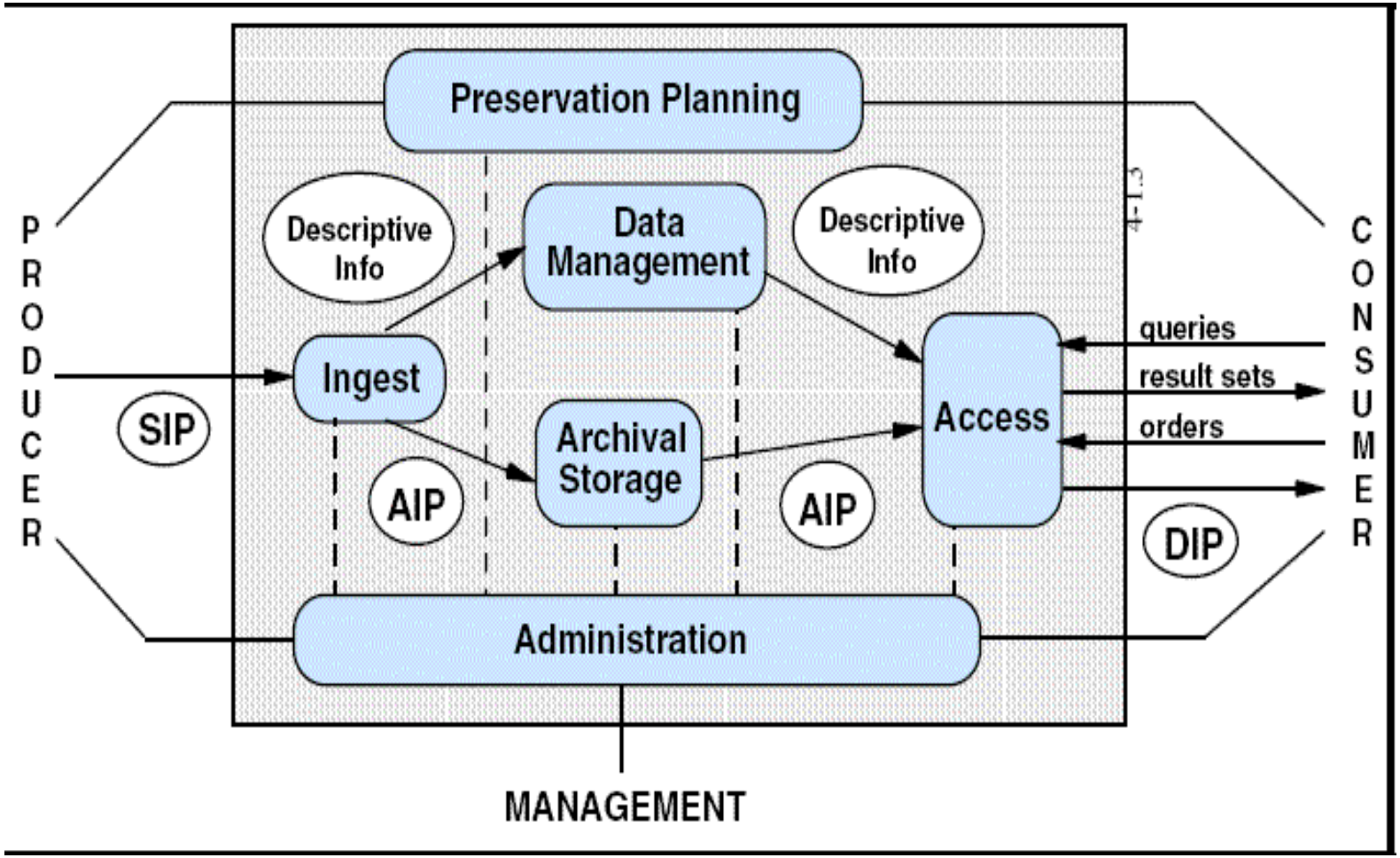
NASA/ P-36089C, 90-HC-380

NASA doesn’t always get it right

- About 20 percent of the data collected for NASA's 1976 Viking Mars landing is completely unreadable and lost forever. With over 1000 people working on the landing alone can you imagine how much they spent to get that 20% of data in the first place?



Open Archival Information System (OAIS)

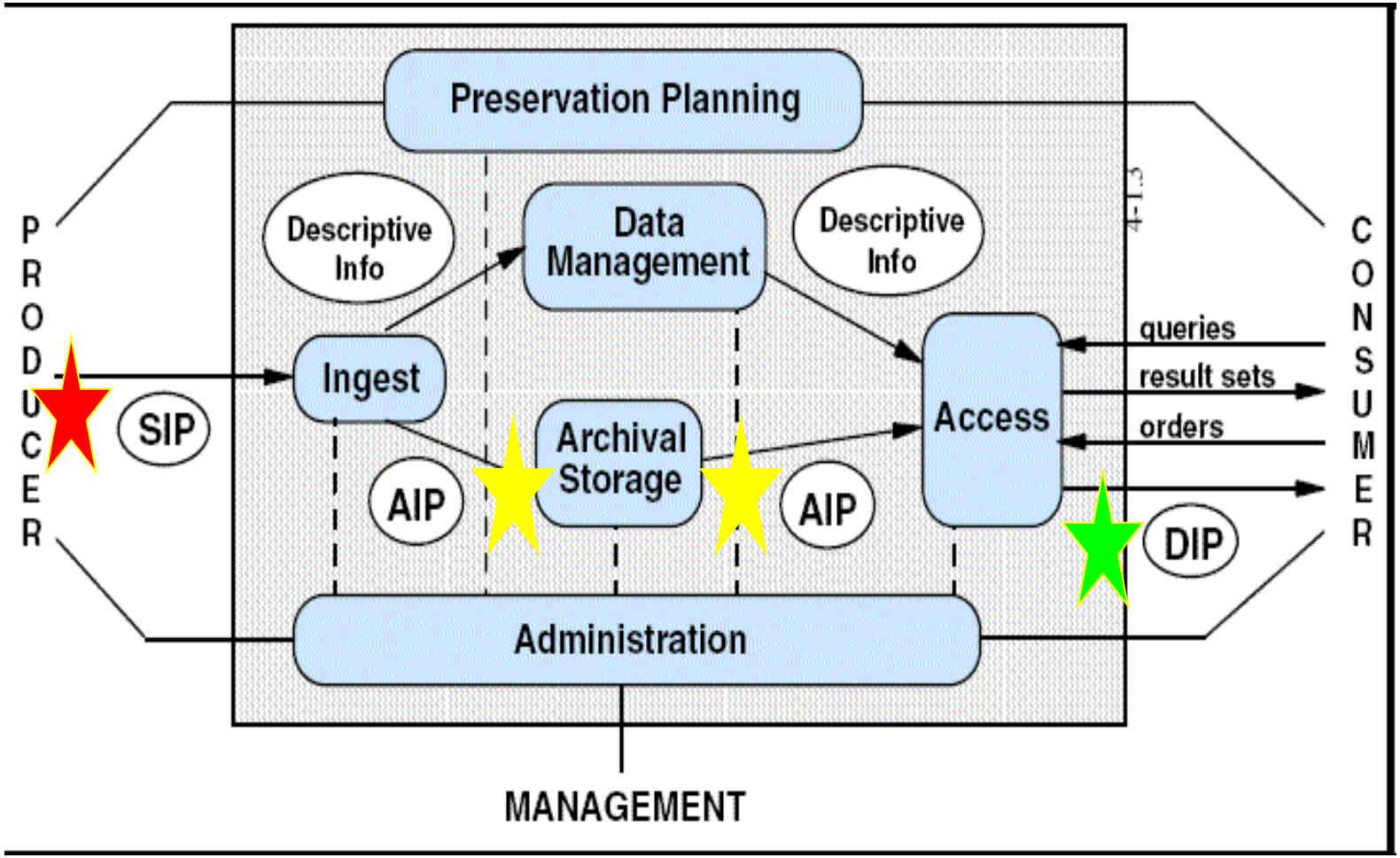

Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
Ingest
SIP
Archival Storage
Access
queries
result sets
orders
CONSUMER
AIP
AIP
DIP
Administration
MANAGEMENT

An information package (IP)

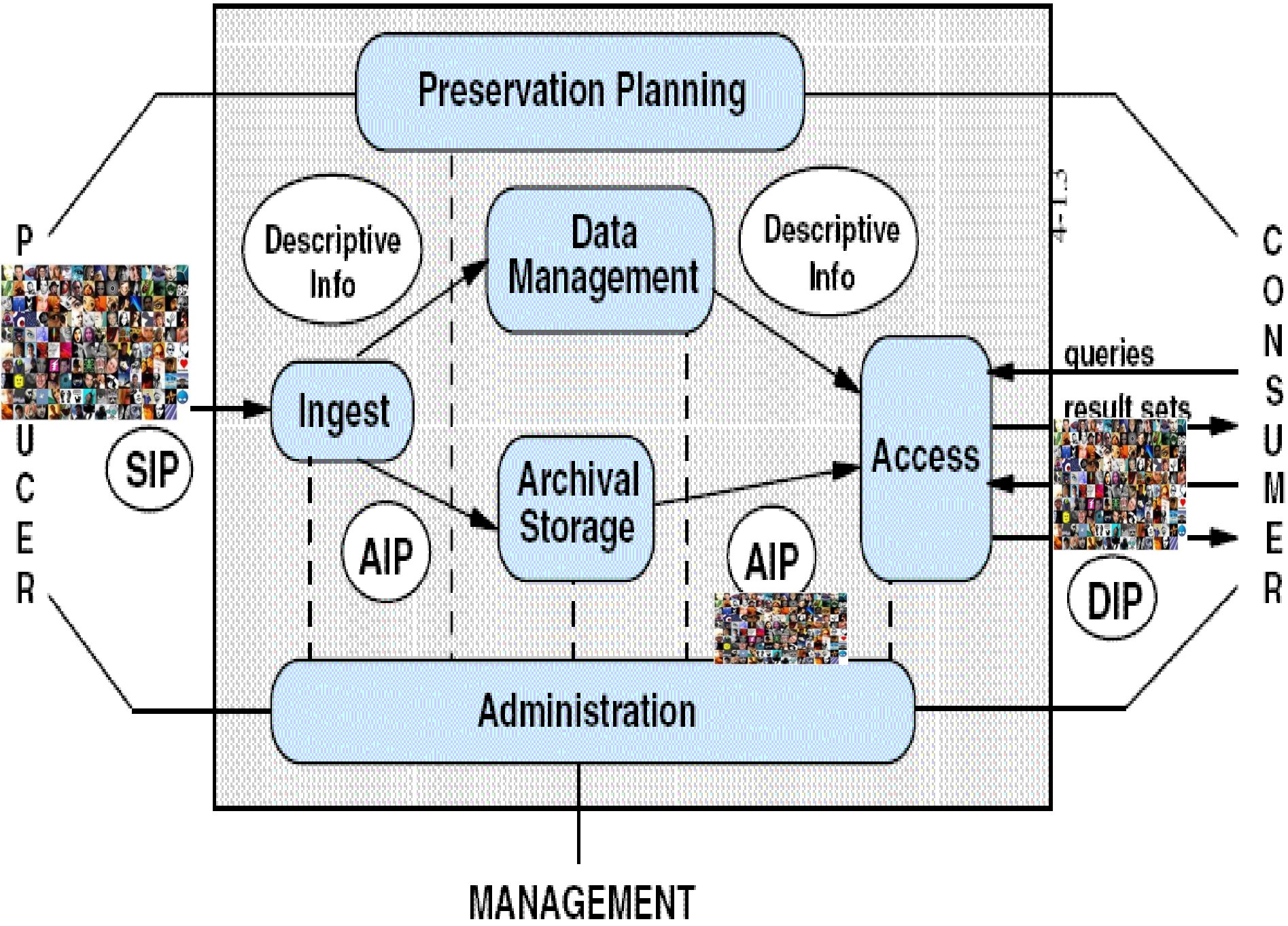
- Contains
  - The digital object(s) to be preserved.
  - The metadata required at that point in the system.
  - The Packaging Information which relates 1 and 2.

Information packages in OAIS

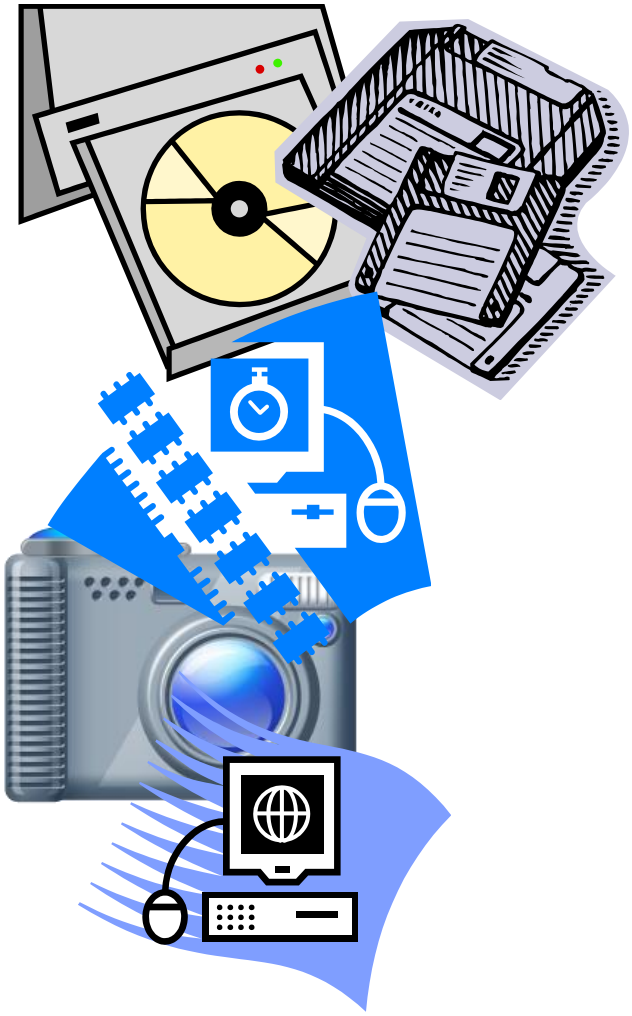
- OAIS outlines three types of Information Package:
  - Submission Information Packages (SIPs).
  - Archival Information Packages (AIPs).
  - Dissemination Information Packages (DIPs).

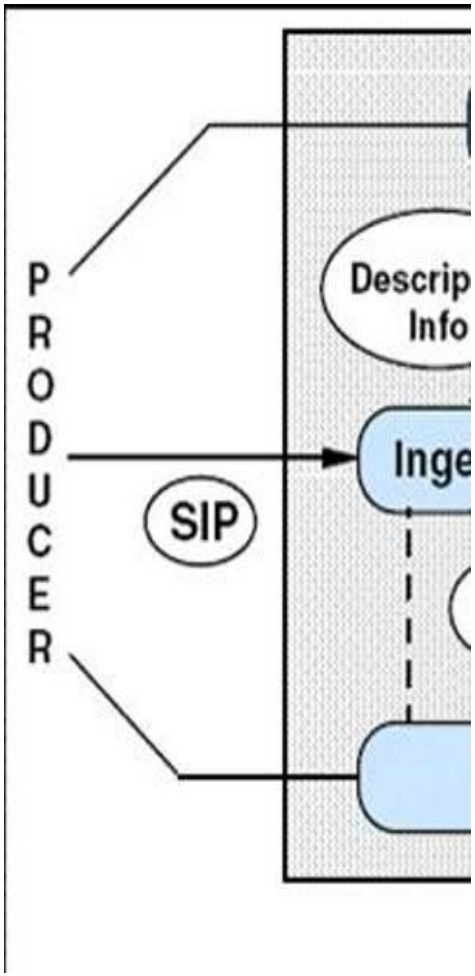

Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
SIP
Ingest
Archival Storage
Access
queries
result sets
orders
CONSUMER
AIP
AIP
DIP
Administration
MANAGEMENT

Personal archiving


Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
Ingest
SIP
AIP
Archival Storage
AIP
Access
queries
result sets
DIP
CONSUMER
Administration
MANAGEMENT

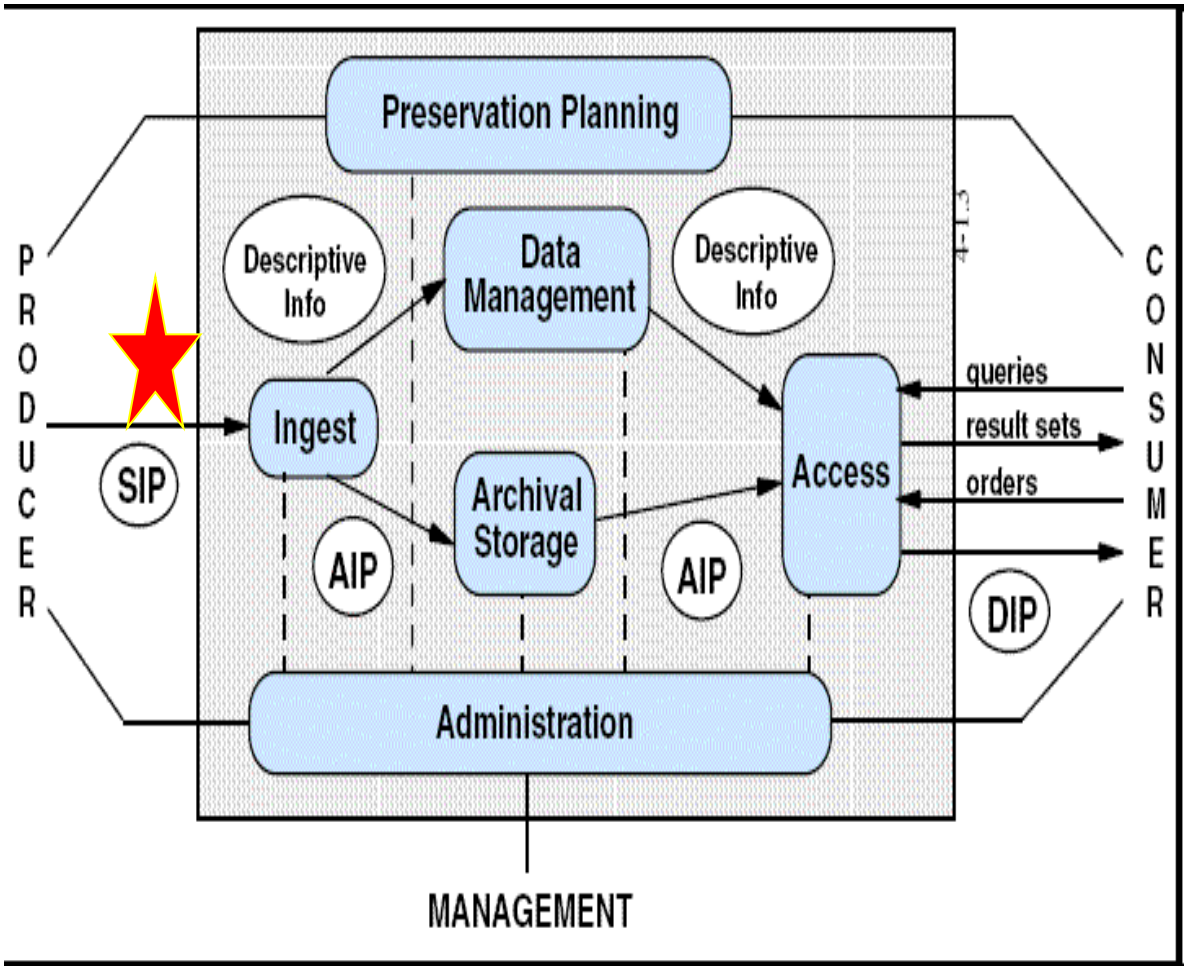
Identify





Here's your SIP

Metadata + digital photo.jpeg =SIP

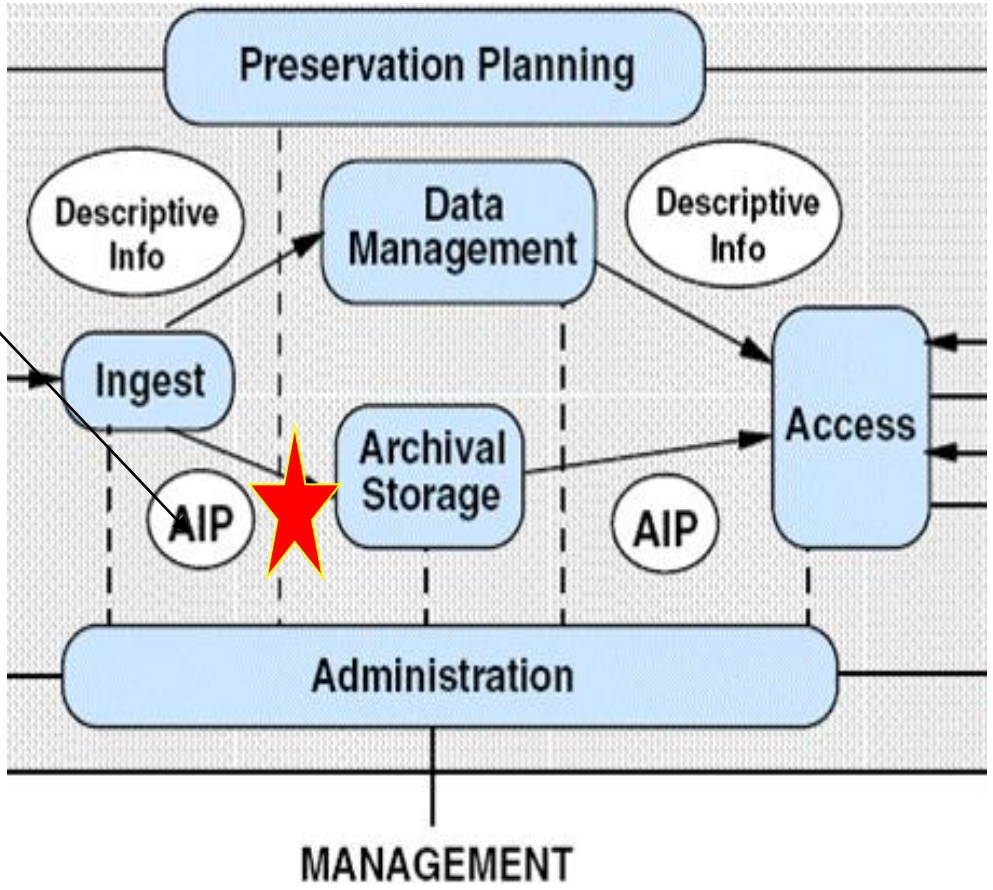


AIP

Ref/name,
Context info

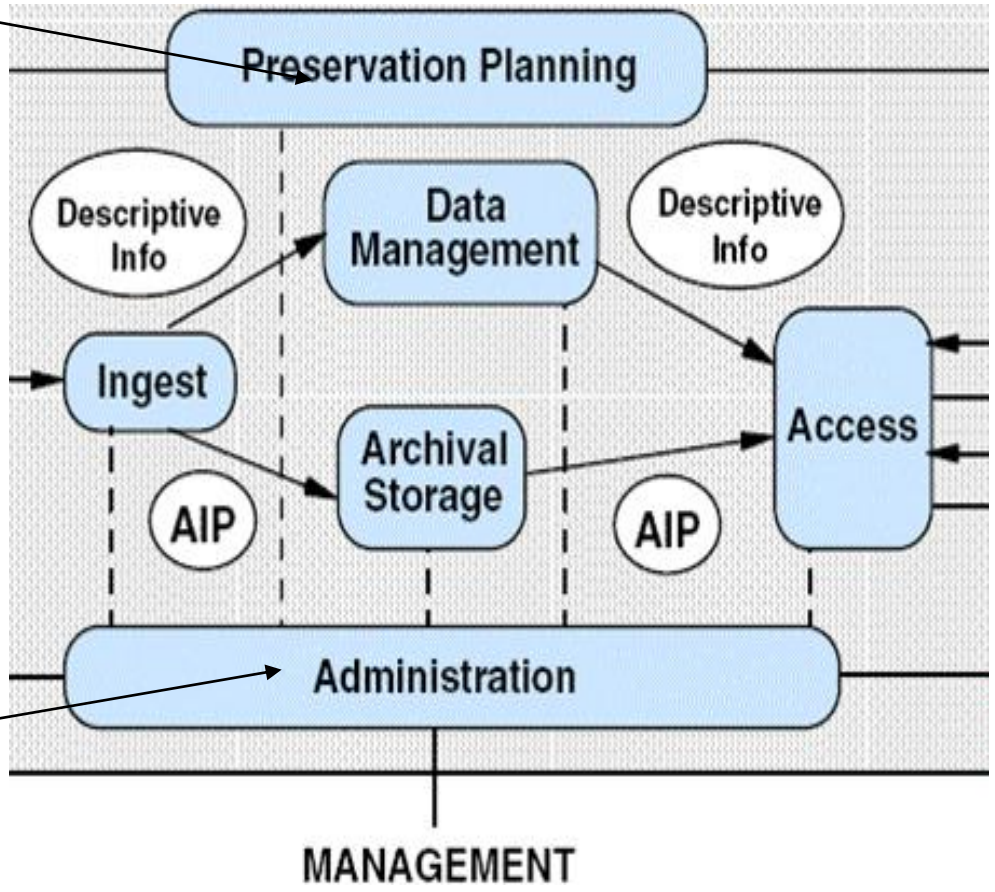
Fixity info

+ object =digital photo.TIFF =AIP



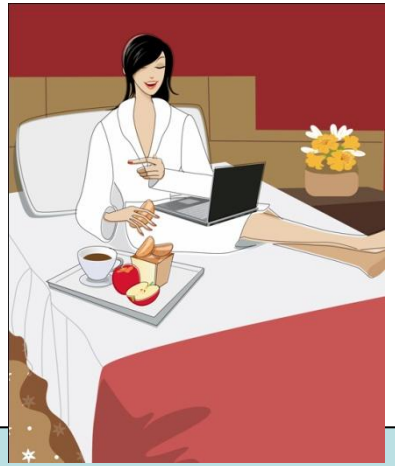
Admin/Management.

Decide how you will preserve your photos over time.



Make sure you look after all the information about your photos and their preservation needs

Here's your DIP



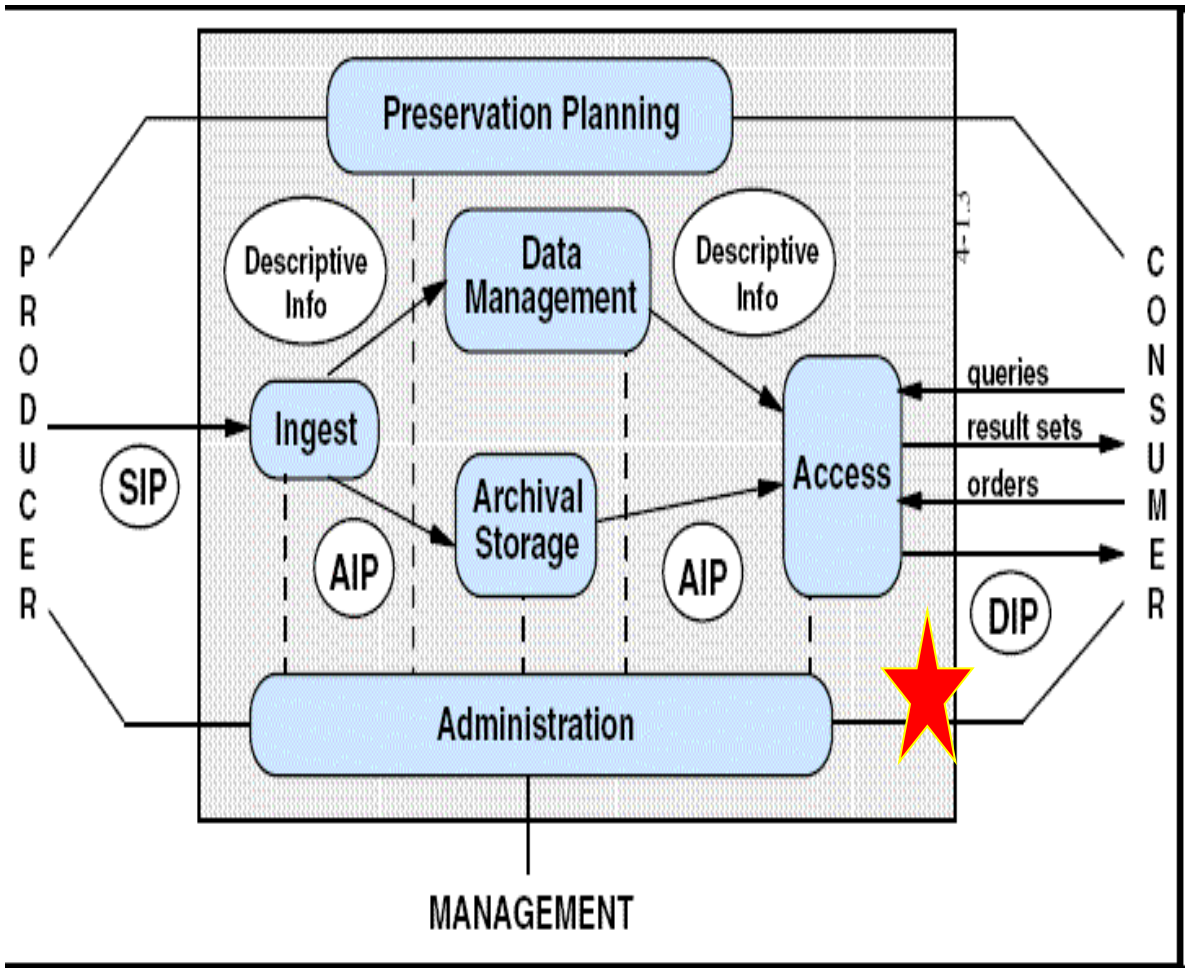
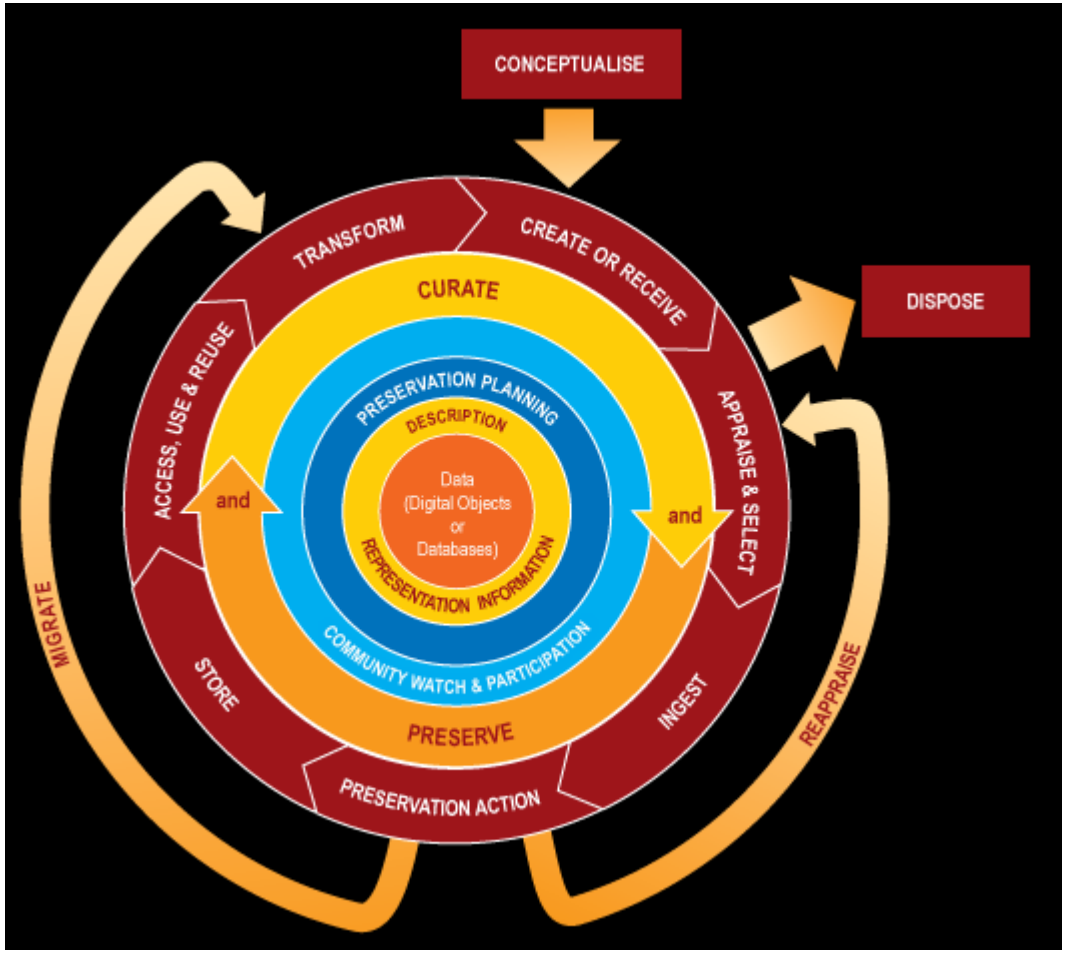


Photo =digital object asJpeg

Descriptive information

Other ways of looking at digital preservation workflow

DCC Lifecycle model



Some simple preservation strategies

- Build policy foundations
- Develop an internal system
- Use external services
- Learn by doing
- Collaborate

Some simple preservation strategies

- Bit preservation
- Migration
- Emulation

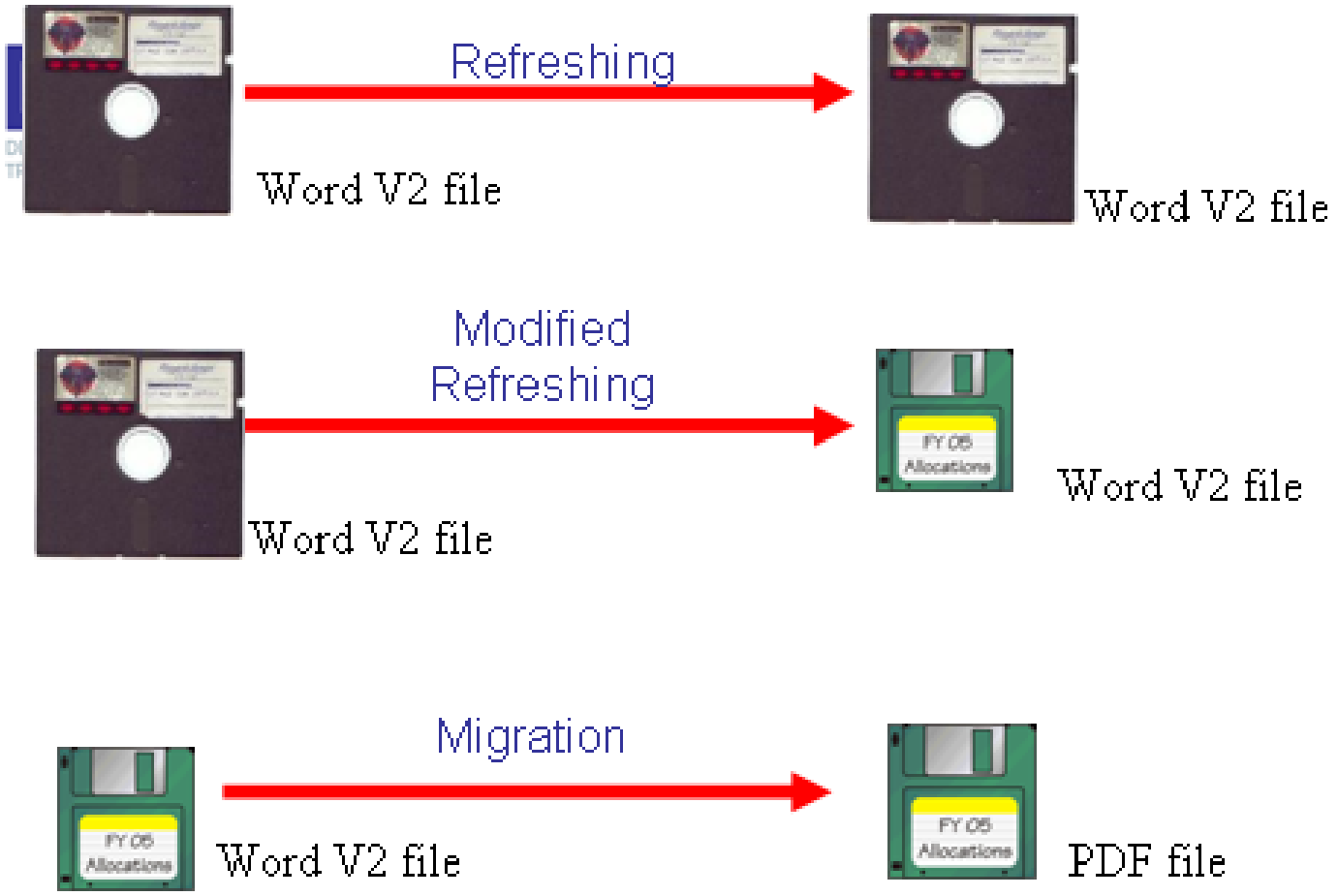
Bit preservation

- Capture information in its original form and focus on maintaining data integrity.
  - Advantages: scalable, practicable; works well (so far).
  - Disadvantage: useful life of data unclear (formats may go obsolete, for example).

Migration

- Transform/normalize data into formats and structures that are optimal for preservation.
  - Advantages: Homogeneous data easier to manage, access; possible to solve preservation issues once and for all if data never has to be transformed again.
  - Disadvantages: Complex initial processes; potential loss of data, functionality; scalability, practicality not proven.

Refreshing is NOT migration



Emulation

- Use software to mimic behavior of obsolete systems to access and use original data.
  - Advantages: look and feel preserved; no need to process original files.
  - Disadvantages: development of emulators can be costly and needs to account for application and operating system, both of which can have many variations; always some uncertainty if the emulation is completely correct.

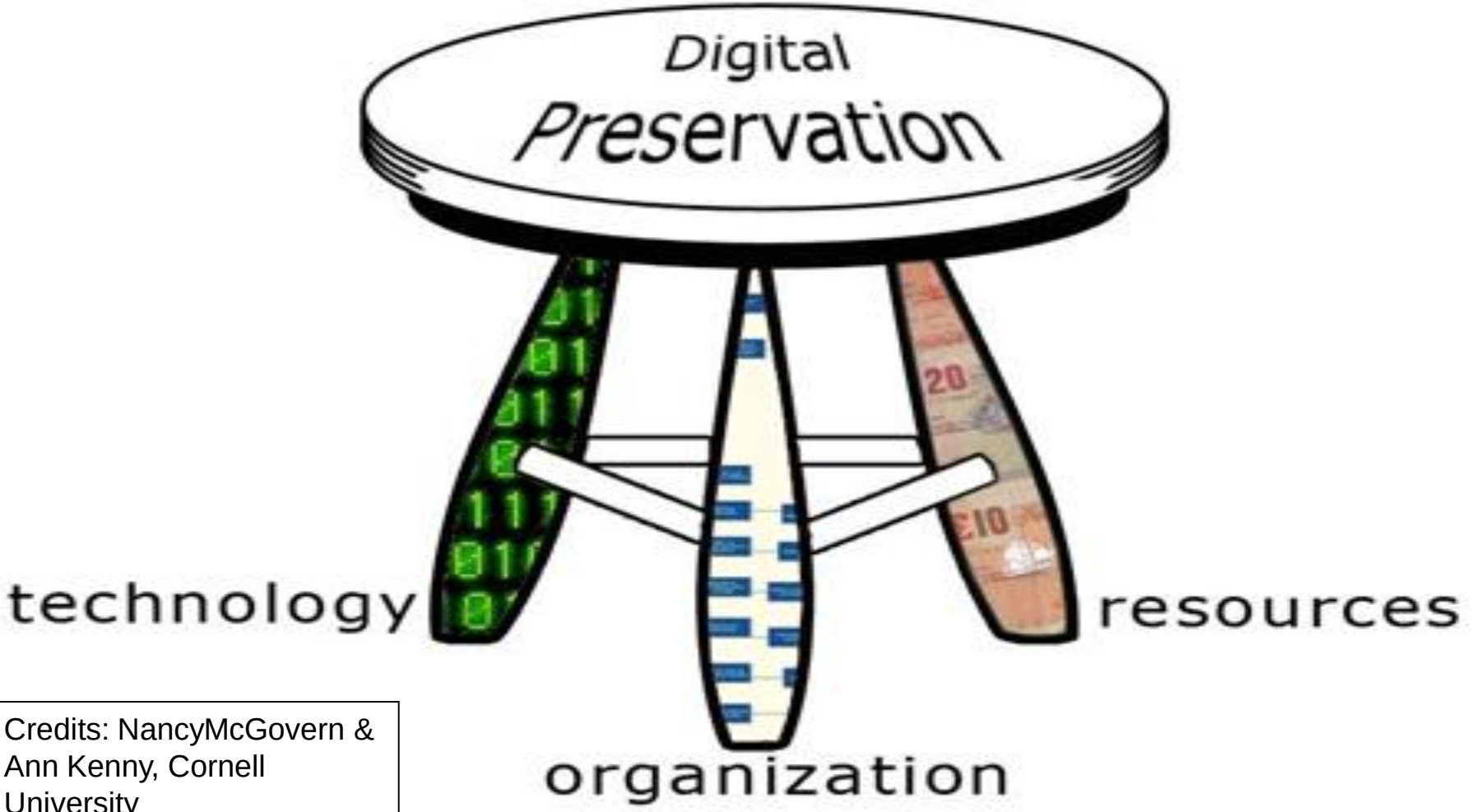
You are not alone

- Collaboration is key to digital preservation
  - Use your contacts and identify expertise which exists elsewhere which you may need
  - Funding bids
  - Examples of collaboration include
    - NDAD/TNA project
    - DPTP training project

Functions which tools are needed for

- Preservation planning including policy and strategy planning, risk assessment, software decisions, implementing standards
- Data objects are appropriately ingested, archivally stored and and managed
- Administrative procedures are in place for the overall operation of the archive
- Data objects continue to be accessible to those who need to use them over time

Planning: where and how do I start?

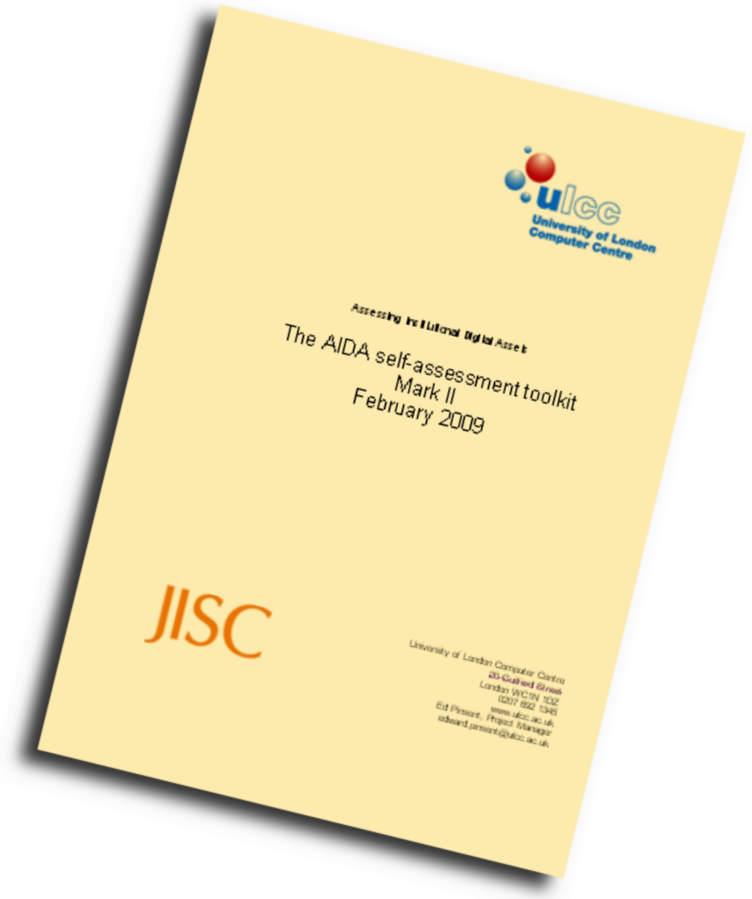


Credits: NancyMcGovern & Ann Kenny, Cornell University

AIDA

- AIDA self assessment toolkit
  - Measures capacity
  - Readiness
  - Capability

http://aida.jiscinvolve.org/wp/



Some tools to look at

- Bagit - the Library of Congress tool
- DROID - National Archives' tool for file format identification
- Jhove - developed by JSTOR and Harvard for object identification, validation and metadata extraction
- NLNZ metadata extraction tool - extracts basic metadata from some popular formats

Tools – good sites to visit

Comprehensive advice sites

- Digital Preservation Coalition
  http://www.dpconline.org
- PADI
  - http://www.nla.gov.au/padi/topics/535.html
- Library of Congress
  - http://www.digitalpreservation.gov/

Training

- Training opportunities
  - www.dptp.org

In the meantime…small steps

- Decide how you will preserve your objects over time.
- Think about the best preservation format.
- Make at least two copies of your object—more copies are better.
- Store copies in different locations that are as physically far apart as practical. If disaster strikes one location, your photographs in the other place should be safe.
- Put a copy of the inventory in a secure location.
- Check your objects at least once a year to make sure you can read them.
- Create new media copies every five years or when necessary to avoid data loss.

In the meantime

Good practise

- Use open, non-proprietary data formats.
- Try to make sure your metadata in conformance with emerging standards and documentation aimed at facilitating future use and future management of the resource.

What else?

- Look at sucesses

  Gloucestershire archives:

  http://futurearchives.blogspot.com/2010/03/scat-gloucestershire-archives.html

- Look at failures
  - Both are learning opportunities

- Look at DPC case notes series

  http://www.dpconline.org/advice/case-notes

- Share your sucesses/failures

And remember

- You are probably doing a lot of it already…

Thank you

Patricia Sleeman

P.sleeman@ulcc.ac.uk

- http://www.dptp.org
- http://dablog.ulcc.ac.uk/
- http://www.ulcc.ac.uk

University of London Computer Centre