Preservation Strategies of the *Koninklijke Bibliotheek*

Hilde van Wijngaarden
Digital Preservation Officer

Koninklijke Bibliotheek/ National Library of the Netherlands
www.kb.nl/e-depot

Digital archiving at the KB: the *e*-Depot

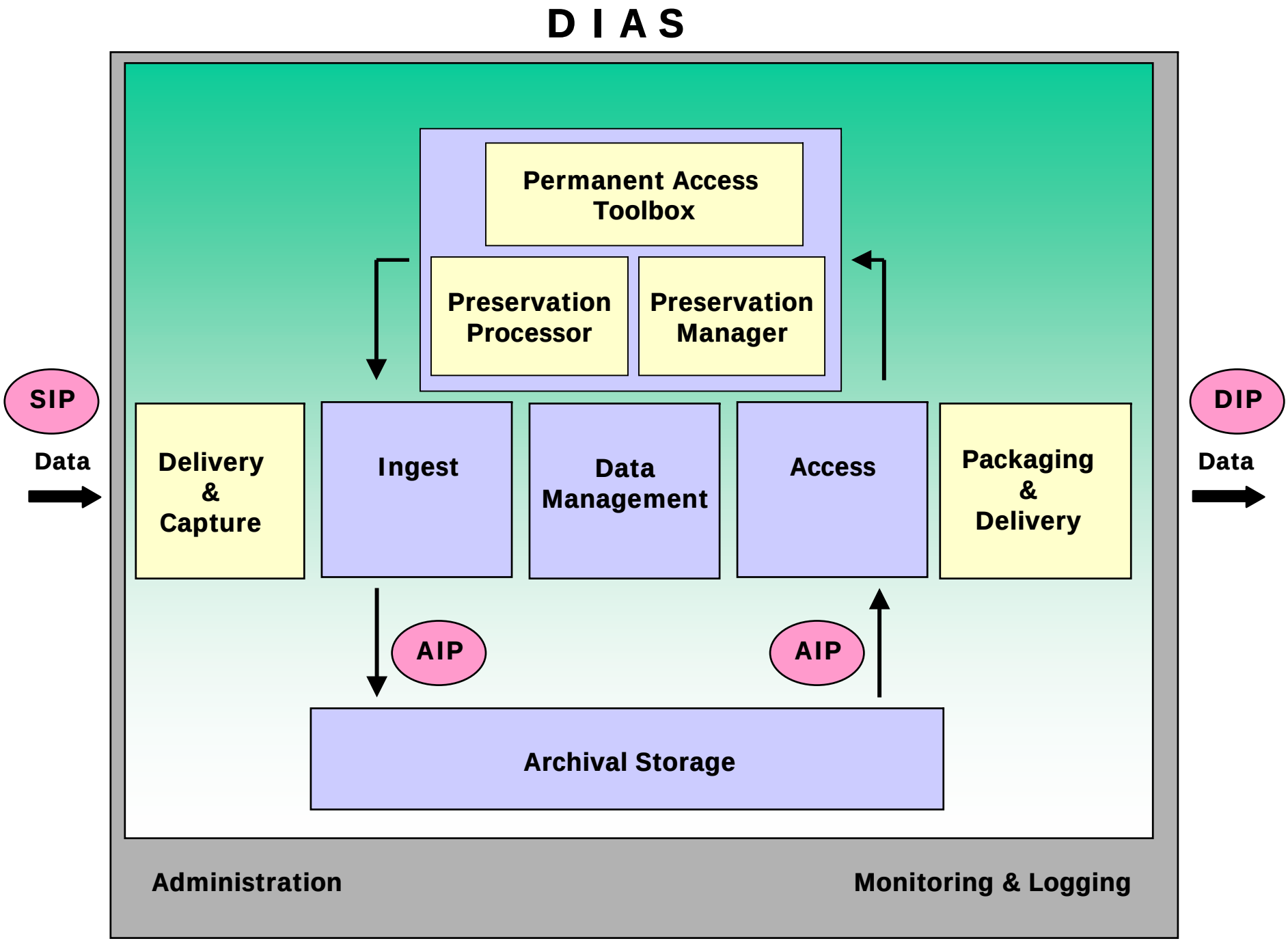
- Electronic version traditional depository
- Developed in collaboration with IBM
- Technical heart: DIAS (OAIS-compliant)
- Integrated with other library modules
- Ingest of online journal articles, e-books, and CD-roms (installables)
- Operational since March 17, 2003
- Over 2 million electronic publications processed

LTP studies 2002

- While building DIAS and the e-Depot, developing Long Term Preservation functionality was not possible yet
- KB/IBM Projectgroup to study LTP issues
- Six studies presented December 2002
- Memorandum of Understanding with IBM

Two projects for 2003

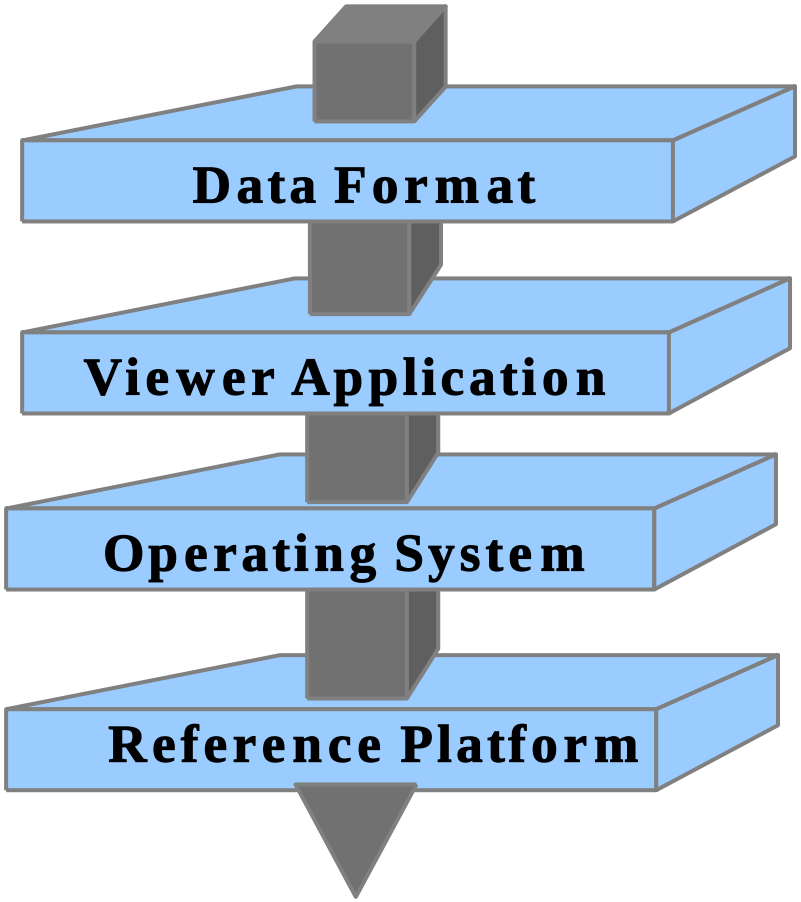
- Preservation Manager
- Operational UVC and Preservation Processor
- Projects were finished in April 2000


DIAS
Permanent Access Toolbox
Preservation Processor
Preservation Manager
SIP
Data
DIP
Data
Delivery & Capture
Ingest
Data Management
Access
Packaging & Delivery
AIP
AIP
Archival Storage
Administration
Monitoring & Logging

The Preservation Manager

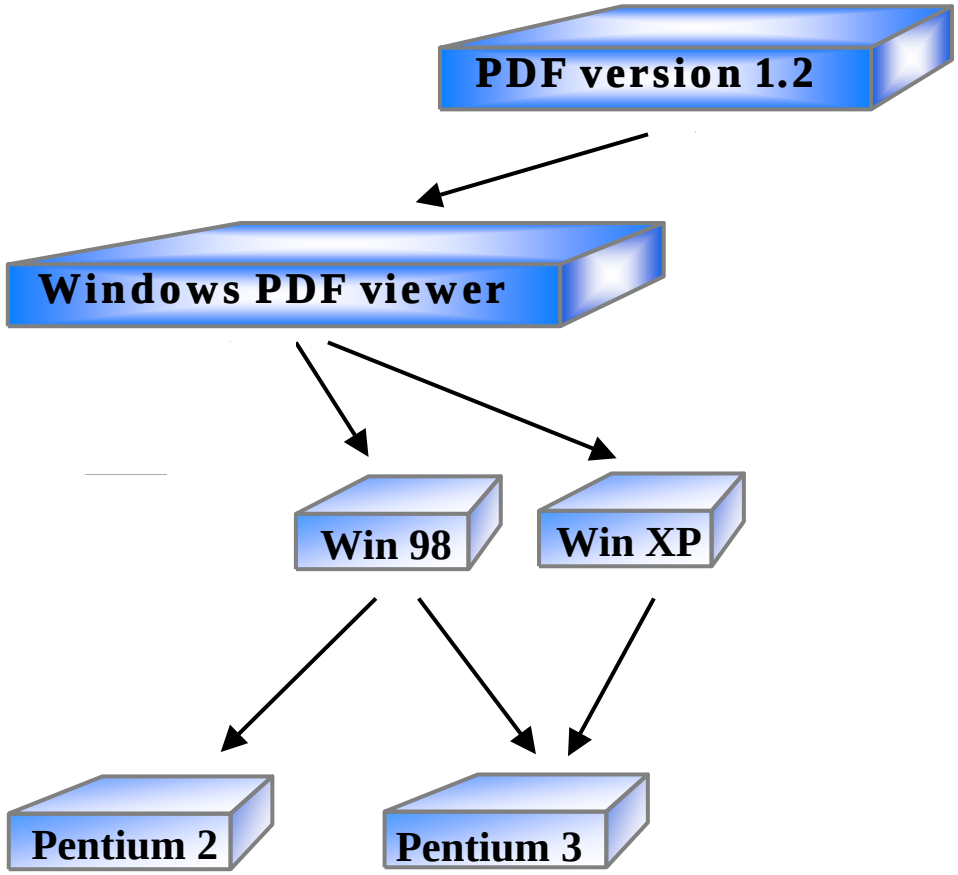
- Storing information on file formats
- Control mechanism for changing technologies
- Possible interaction with international format registries
- Hardware and software specifications are described as layers
- Layers are the building blocks for a **Preservation Layer Model** (PLM)
- **View Paths** are instantiations of PLMs
- Every file format is connected to one or more View Paths

Autonomous Digital Item



- The **Data Format** identifies the structure and meaning of a bit-stream, like a .PDF file
- The structure and meaning of the bit-stream are defined within the application logic of a specific **Viewer Application**
- The **Operating System** contains the functionalities that all viewer applications need like access to a printer or scanner, and fundamental requirements like file structure
- The **Reference Platform** specifies how the bits and bytes are transformed to a physical representation like for example on a screen

Requested Information Object



So... Windows^{95} gets obsolete..

And... Acrobat for AIX is no longer supported..

Planning Permanent Access Strategies:

- What do you want to preserve?
- Why do you want to preserve?
- What do you want to render in the future?

KB: Choice for keeping the original

Consequence of choosing to keep the original:

- Limited number of strategies
- Emulation is the preferred strategy

Emulation:

- Never operationalised in a digital archiving environment
- Project will start this year, in co-operation with the Dutch National Archives

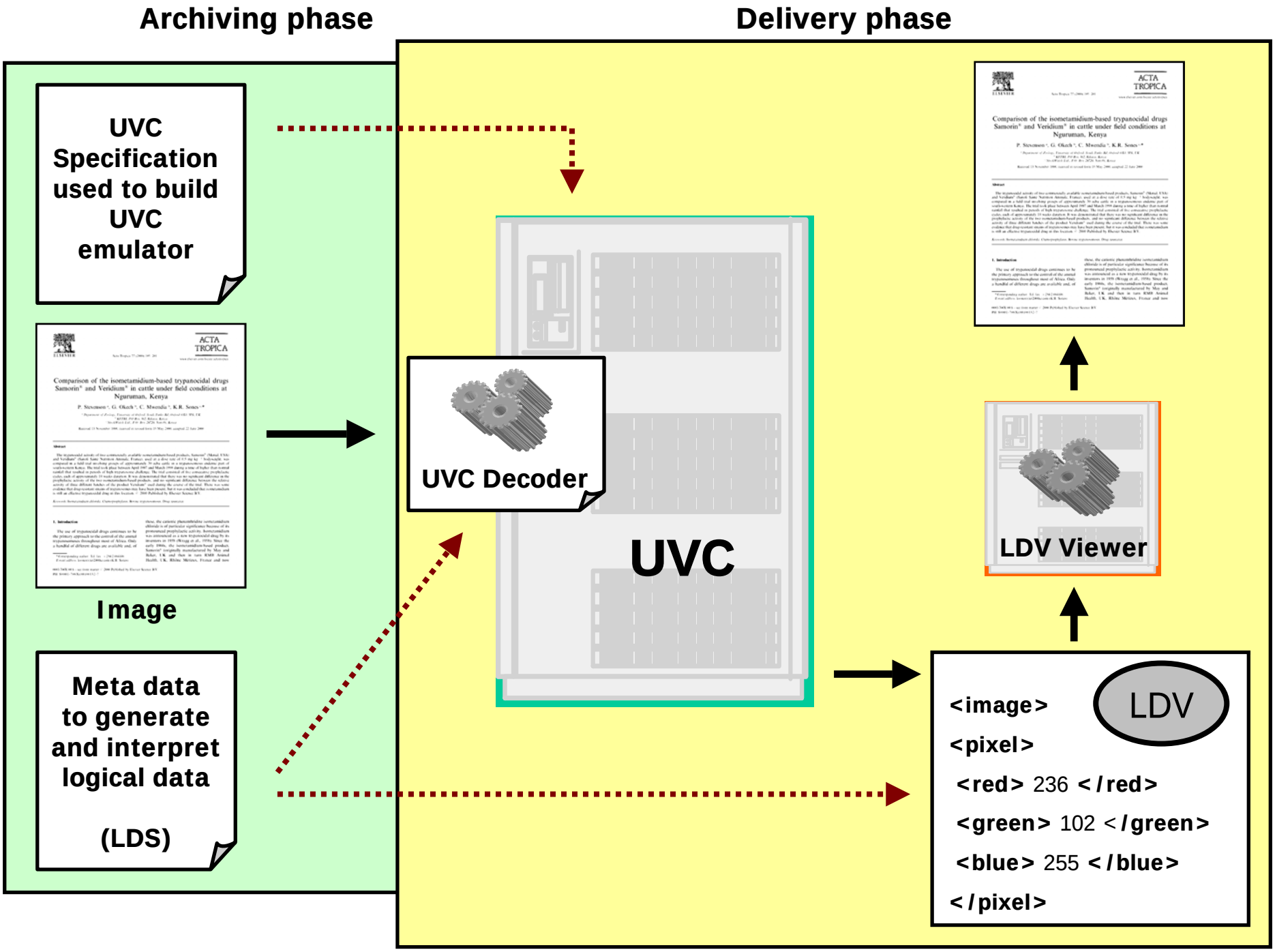
Universal Virtual Computer:

- Combination of emulation and migration (on-the-fly)
- Joint project with IBM to develop operational UVC
- First operational UVC for JPEG

KB

Universal Virtual Computer (UVC)

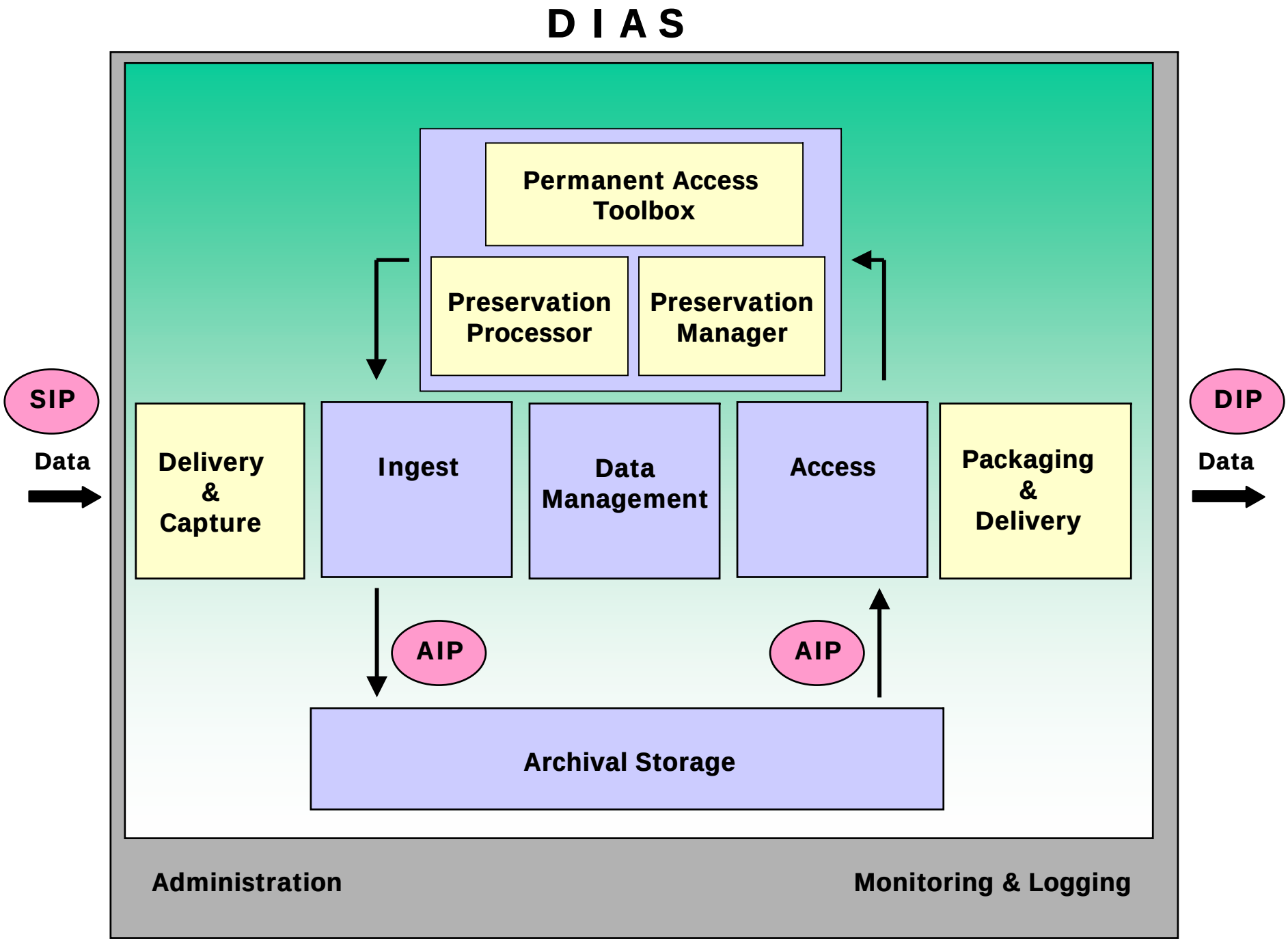
- Approach developed by Raymond Lorie (IBM)
- A specification of a virtual intermediate platform
- Simple enough to be implemented on any future platform
- A UVC **Decoder** translates objects into a Logical Data View (XML-like)
- A schema explains the **Logical Data View** to enable future viewing


Archiving phase
Delivery phase
UVC Specification used to build UVC emulator
Image
Meta data to generate and interpret logical data
(LDS)
UVC Decoder
UVC
LDV Viewer
<image>
<pixel>
<red> 236 </red>
<green> 102 </green>
<blue> 255 </blue>
</pixel>
LDV

An operational UVC for JPEG

- UVC for PDF too complicated
- To review this strategy we need an operational tool
- All the components for a working solution needed
- Convert PDF to JPEG
- The ‘safety-net’ solution: UVC for images

DEMO
of the
UVC for JPEG


DIAS
Permanent Access Toolbox
Preservation Processor
Preservation Manager
SIP
Data
DIP
Data
Delivery & Capture
Ingest
Data Management
Access
Packaging & Delivery
AIP
AIP
Archival Storage
Administration
Monitoring & Logging

The Preservation Processor

- Module for selection and processing of stored digital objects

- First: specific for UVC for Jpeg
  - Select pdf-files
  - Convert to Jpeg
  - Re-ingest converted AIP

- Plans for generic version:
  - Selection: objects and/or file formats in *e*-Depot
  - Processing: copying to new format if required
  - Support Permanent Access Strategy
  - Interface with Ingest and Access

Future plans

- Evaluate the UVC for JPEG with international experts
- Extend approach to UVC for TIFF and PDF

Conclusions

- Permanent Access Tool in place, embedded in operational system
- The UVC approach is viable (and can be demonstrated!)
- UVC for JPEG is a ‘safety net’ for the KB *e*-Depot

Any questions?