European **A**rchival **R**ecords and **K**nowledge Preservation

Digital Archiving in the E-ARK Project

Karin Bredenberg, National Archives of Sweden/ ES Solution

Is a Data Warehouse a Data Archive and Why Does It Matter?

19th of February 2015

University of Portsmouth

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA KULTURO
ARHIV REPUBLIKE SLOVENIJE

DLM FORUM FOUNDATION

STATENS ARKIVER
THE DANISH NATIONAL ARCHIVES

DigitalPreservationCoalition

E-ARK

THE E-ARK PROJECT IS CO-FUNDED BY THE EUROPEAN COMMISSION UNDER THE ICT-PSP PROGRAMME

www.eark-project.eu

GOBIERNO DE ESPAÑA
MINISTERIO DE HACIENDA Y ADMINISTRACIONES PÚBLICAS

ICTPSP

ama

KEEPSOLUTIONS
University of Minho SPIN-OFF

MAGENTA aps

ES Solutions

THE NATIONAL ARCHIVES OF NORWAY

RAHVUSARHIIV
THE NATIONAL ARCHIVES OF ESTONIA

archívum
Magyar Nemzeti Levéltár

TÉCNICO LISBOA

Outline

- Back to basics, the challenge of digital preservation
- Specifications
- Database archiving in E-ARK

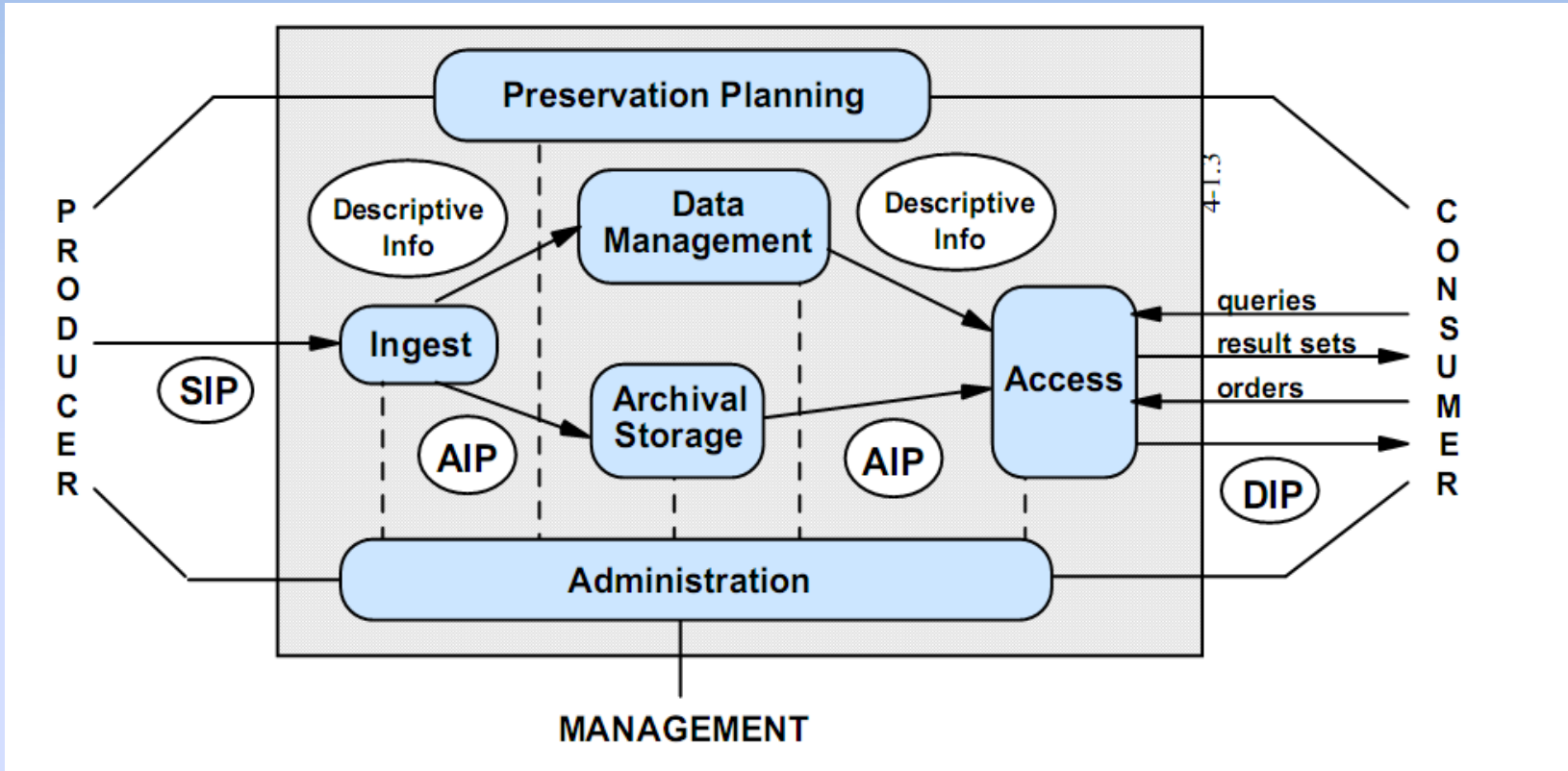
The Challenge of Digital Preservation

E-ARK

ICTPSP

OAIS

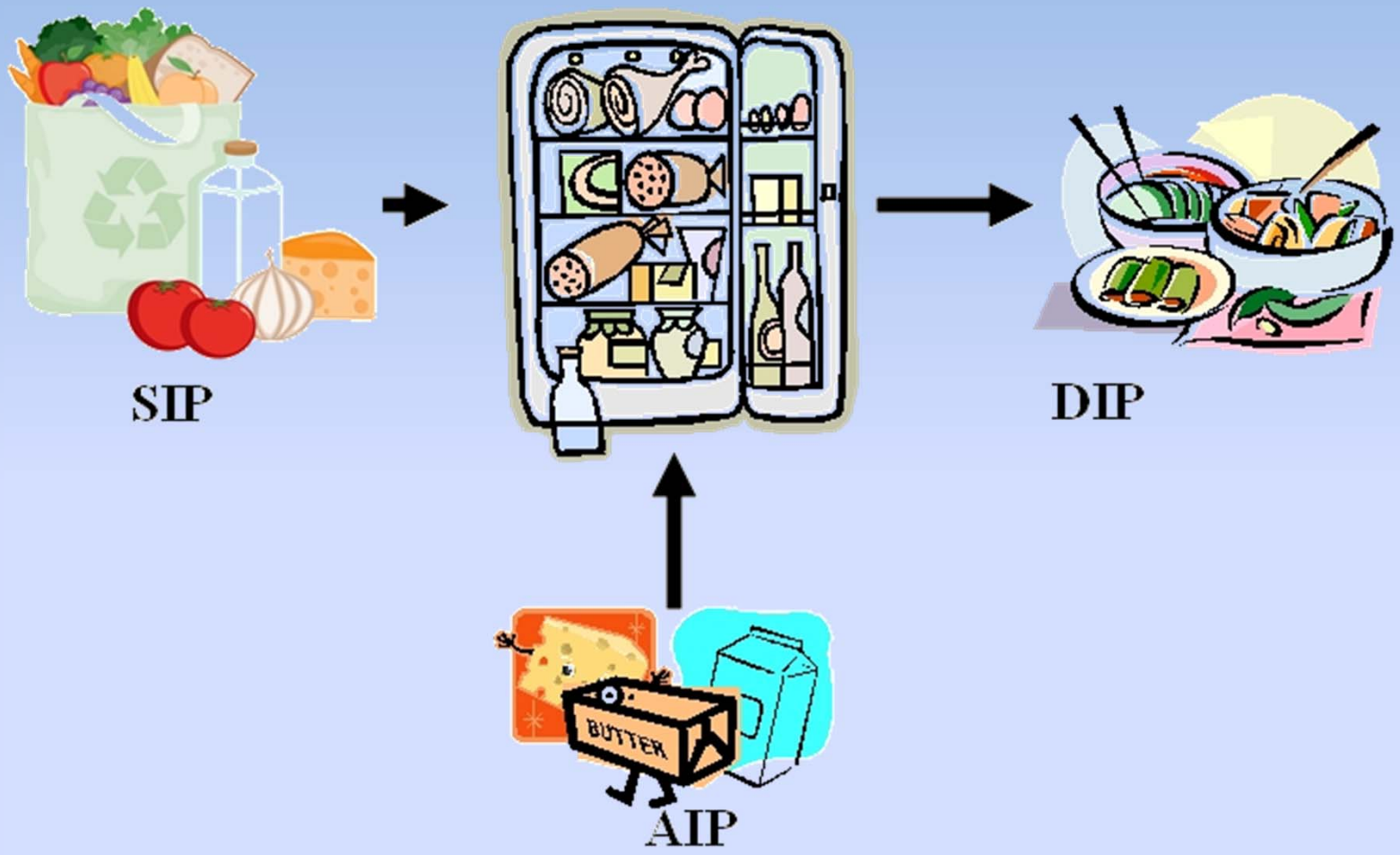
http://public.ccsds.org/publications/archive/650x0m2.pdf



Open Archival Information System
or
the ISO OAIS Reference Model for an OAIS

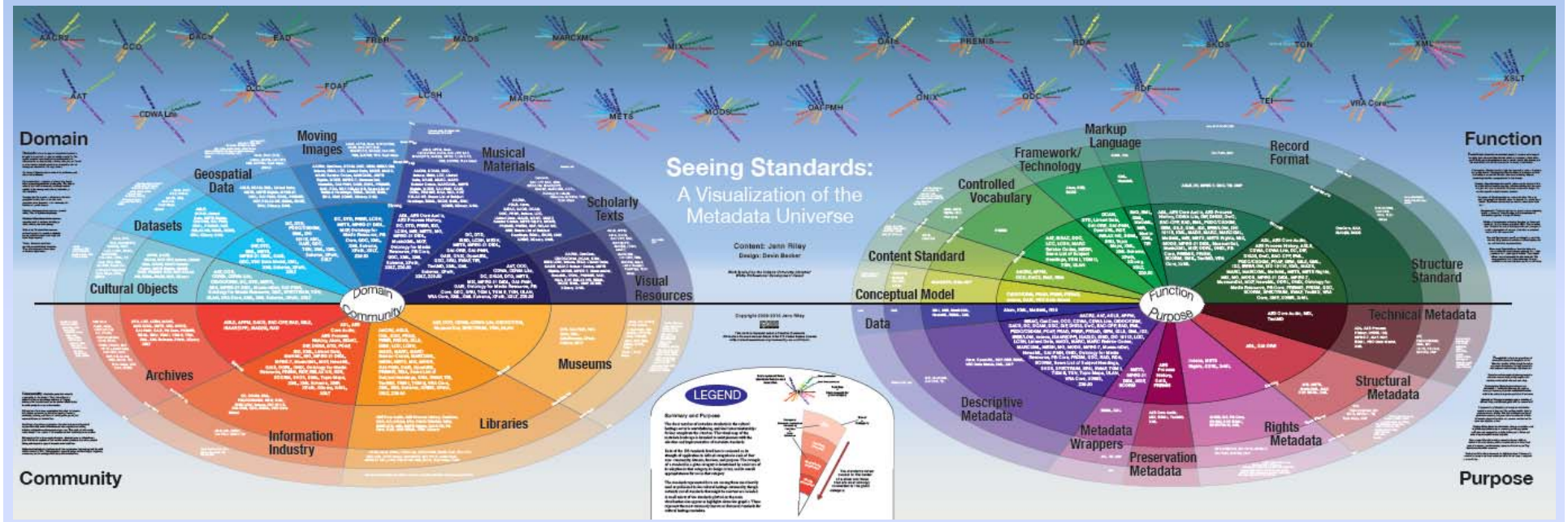
SIP (Submission Information Package)
AIP (Archival Information Package)
DIP (Dissemination Information Package)

OAIS popular version



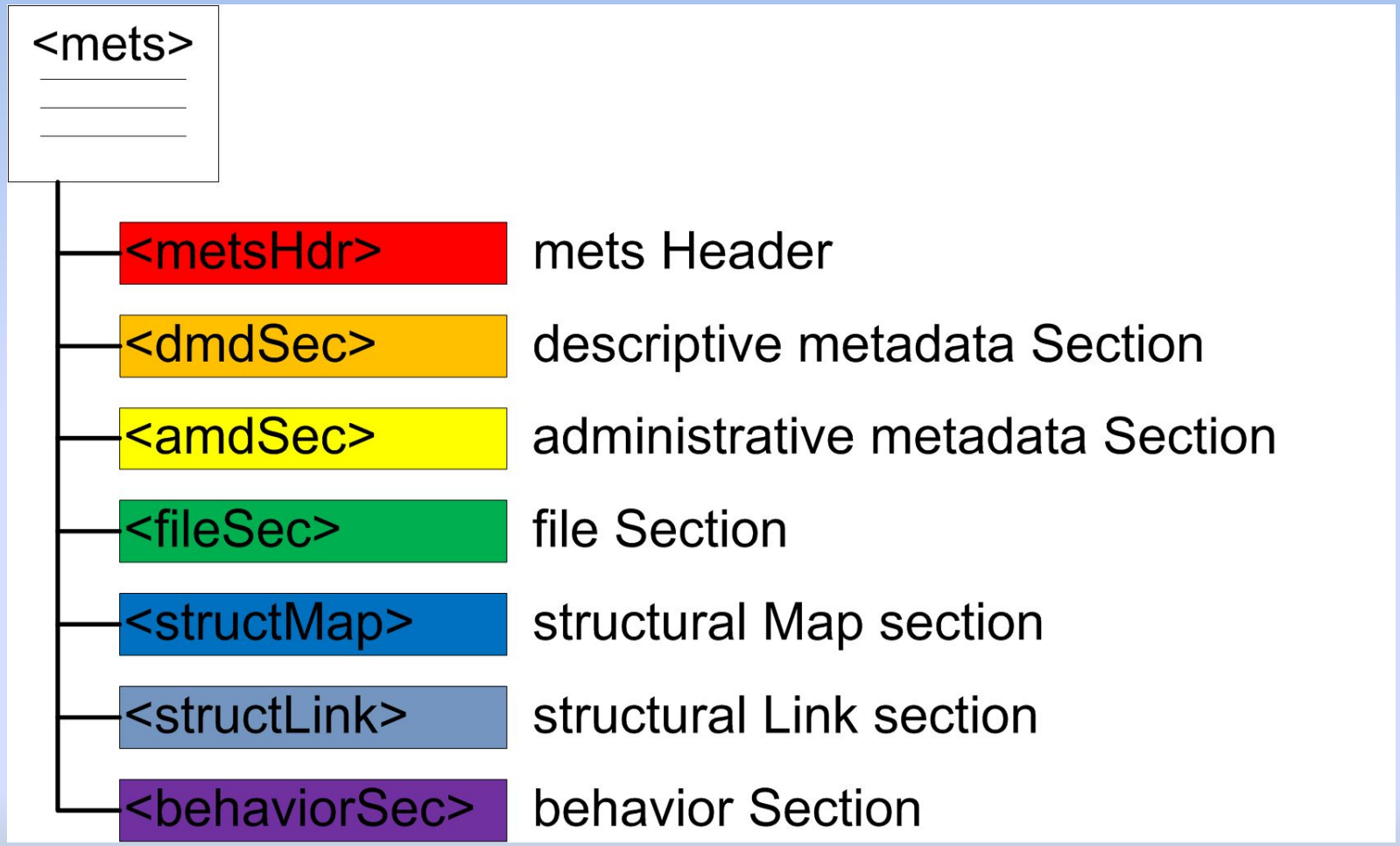
Richard Pearce-Moses and more.

Standards

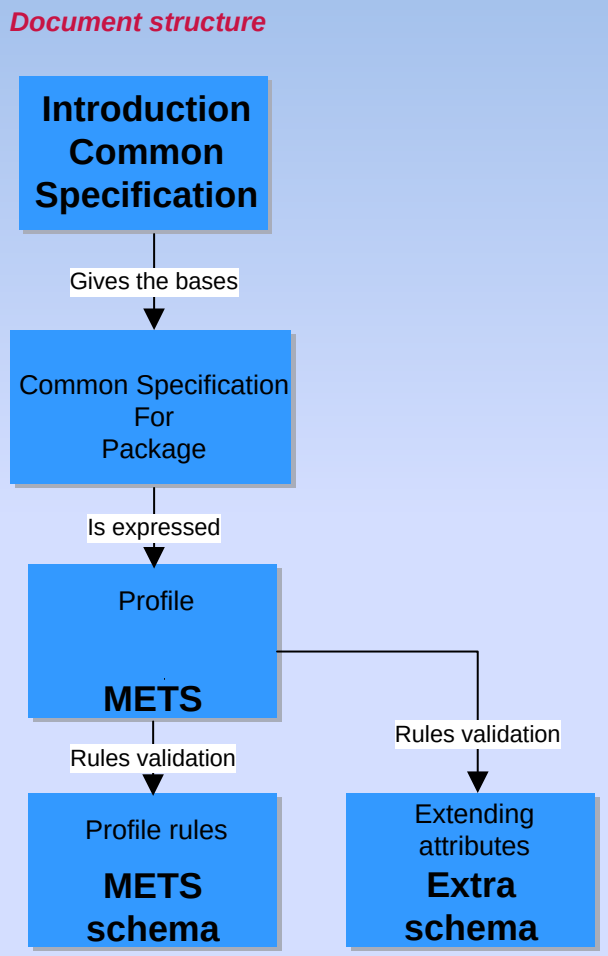


•http://www.dlib.indiana.edu/~jenlrile/metadatamap/

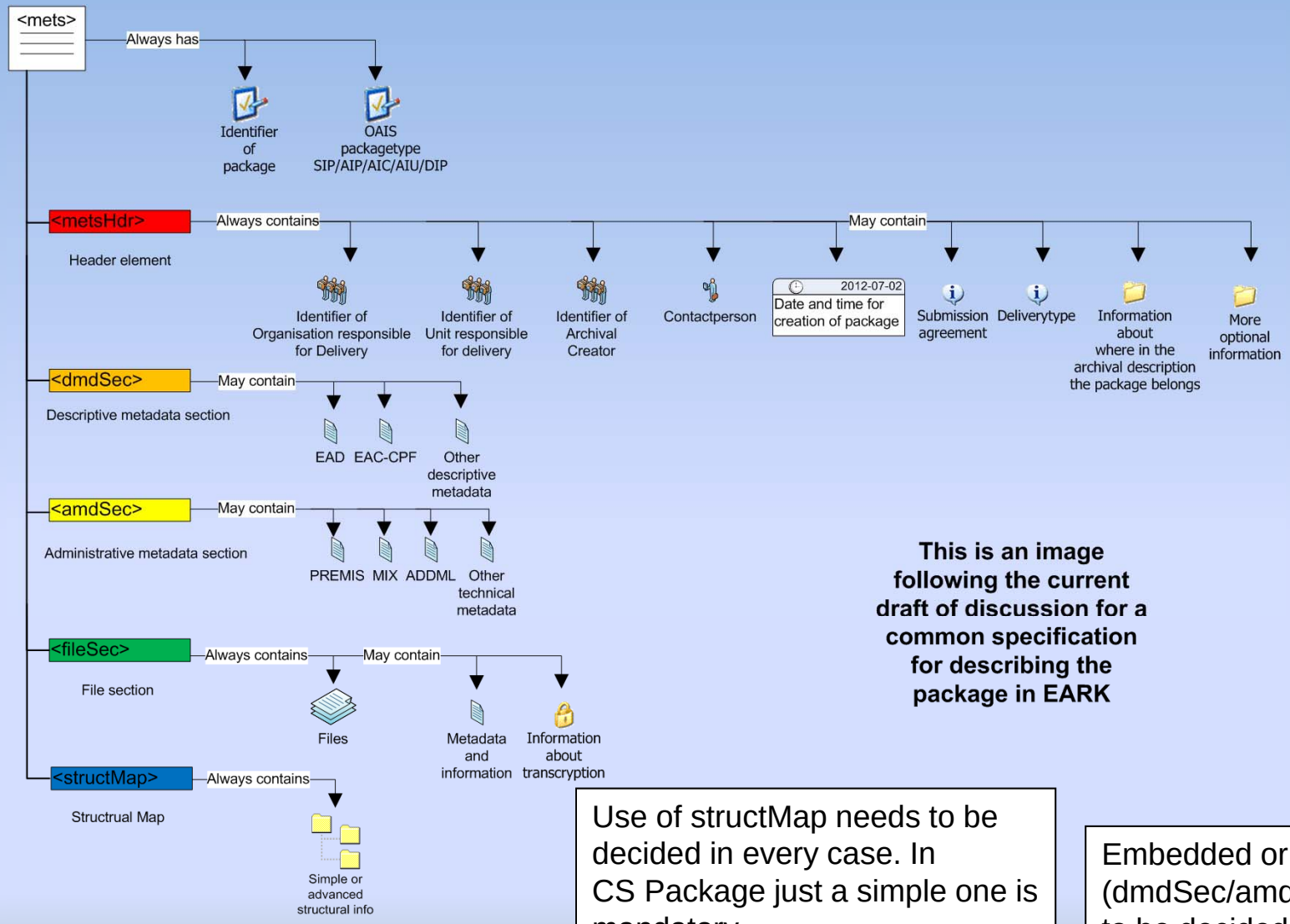
Describing an IP using METS



Document structure IP



In short this information is found in EARK METS

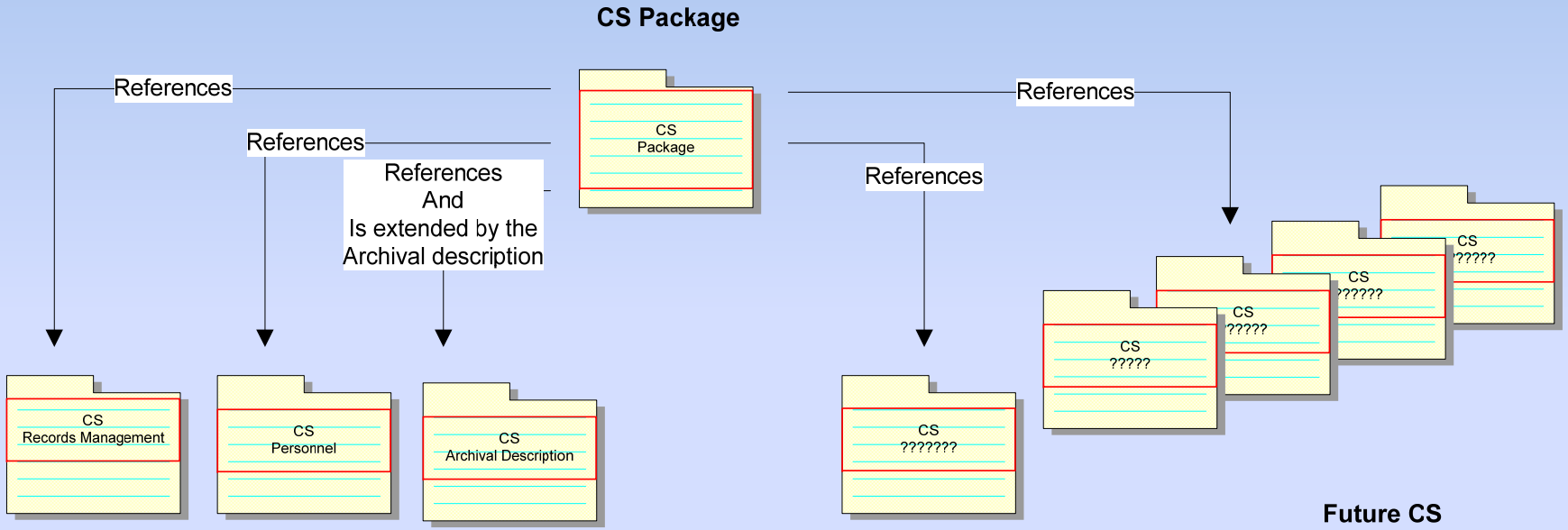


This is an image following the current draft of discussion for a common specification for describing the package in EARK

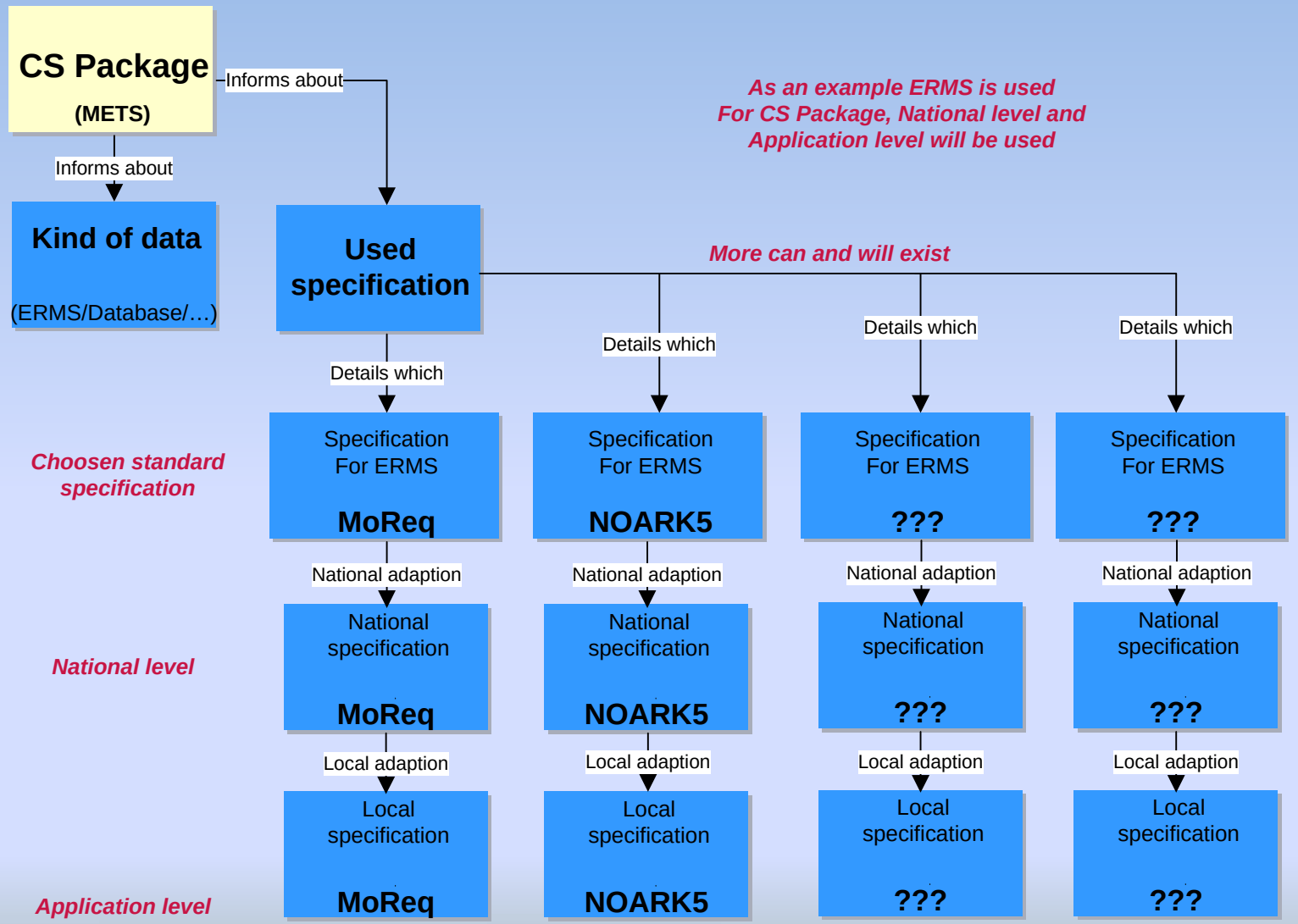
Use of structMap needs to be decided in every case. In CS Package just a simple one is mandatory.

Embedded or referenced data (dmdSec/amdSec/fileSec) needs to be decided in every case.

How to specify all metadata?



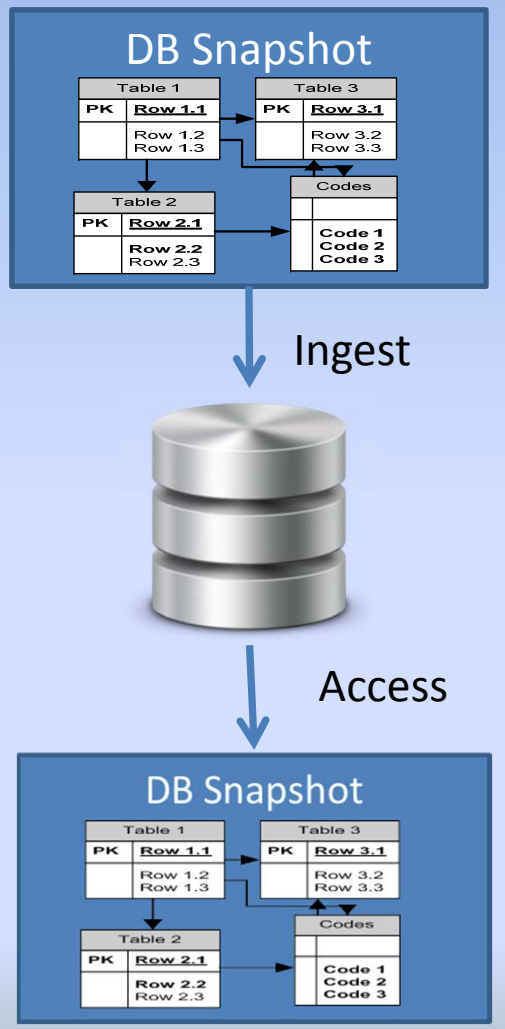
Levels of specification adoptions



Current database archiving practice

„Snapshot“ policy

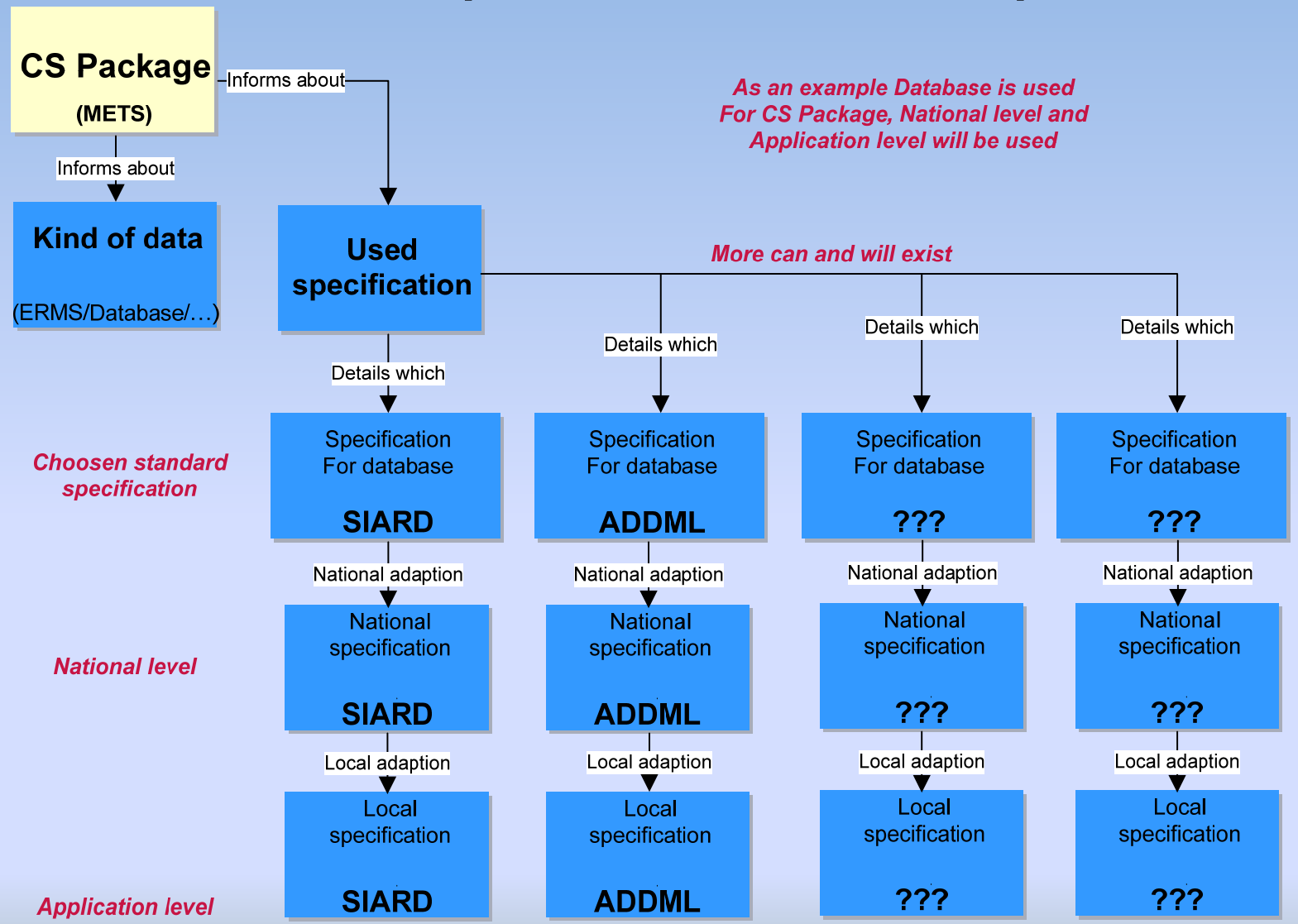
- Ingest: Transform the original relational structure into open formats
- Formats: SIARD, ADDML, DBML
- Access: Users need to find the appropriate snapshot(s), load these into a current DBMS and use pre-defined queries or build their own ones



Problems

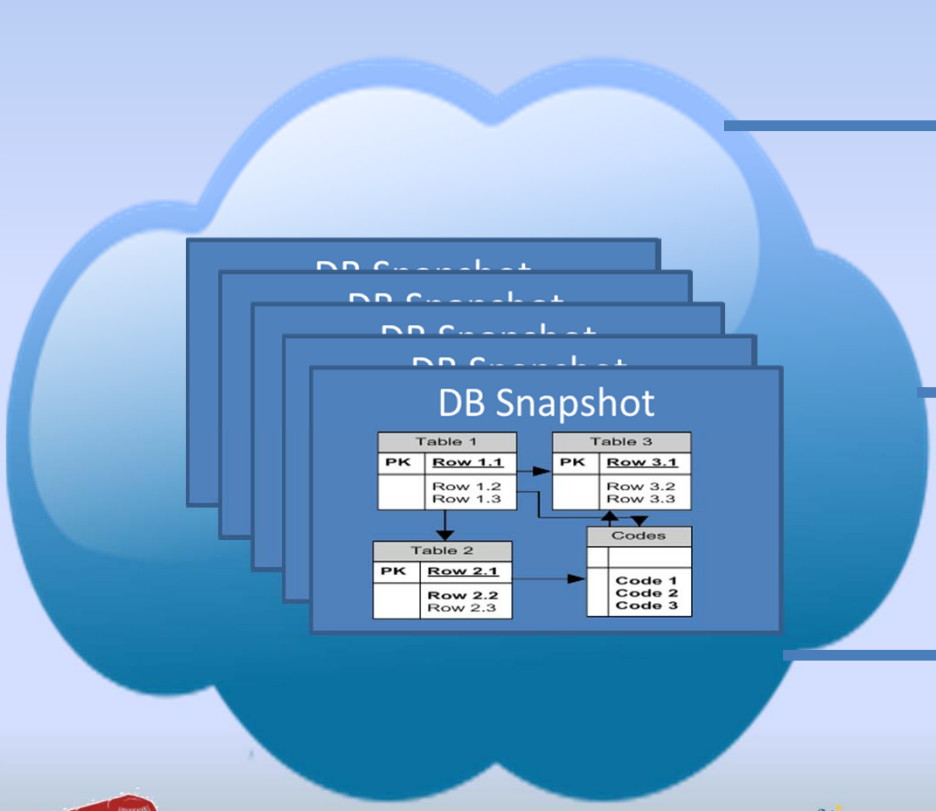
- Finding the appropriate snapshot
  - Most users search for „**data about something**“
    - „Which car had the plate number 111YYY in January 15th 2000“
  - Current practice allows to search for „**the database snapshot** which includes data about something“
    - „Which database includes information about cars in January 15th 2000“
- Scope of the snapshot
  - The scope of data and time period covered in a single snapshot usually do not meet the needs of the user
- Required technical knowledge
  - Relational structures are often highly optimised and hard to grasp
  - Most users do not have the knowledge to build accurate queries for specific access needs
  - The only way is to use pre-defined queries which have been archived along with the data

Levels of specification adoptions



In the ideal world …

- … users do not need to search for databases but data!



Do we need a CS Data Warehouse?

- Yes?
  - A database DIP
- No?
  - A database DIP

And now over to Janet Delve

- Data Warehousing


Q
&
A