Riksarkivet

PREMIS at the Swedish National Archives

Karin Bredenberg 2010-09-19

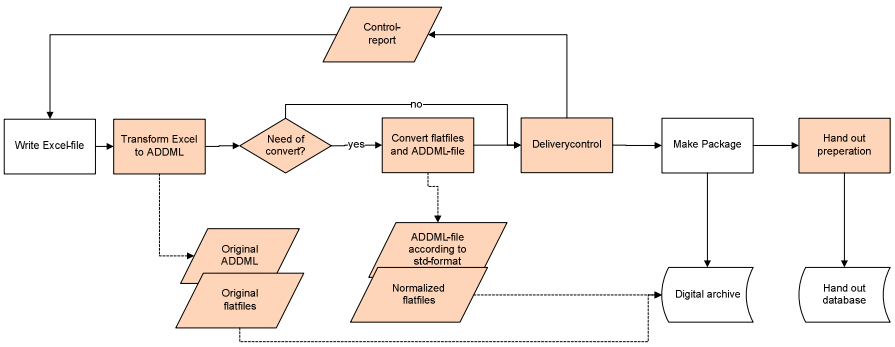
Two different approaches

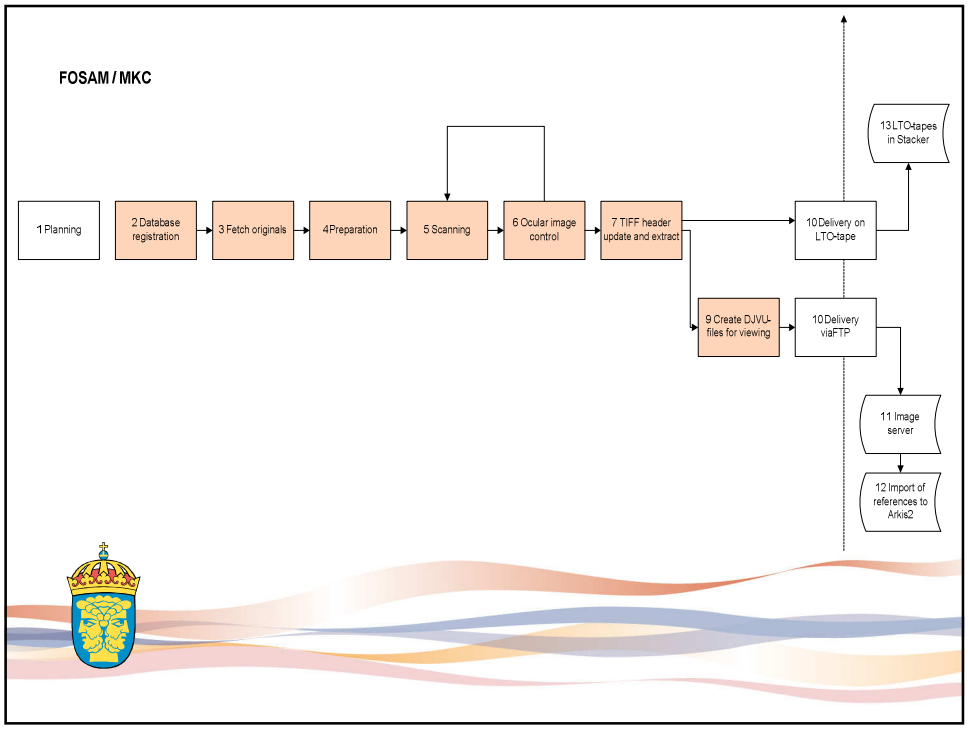
- In our own archival database (in use)
- For our suppliers (in progress)

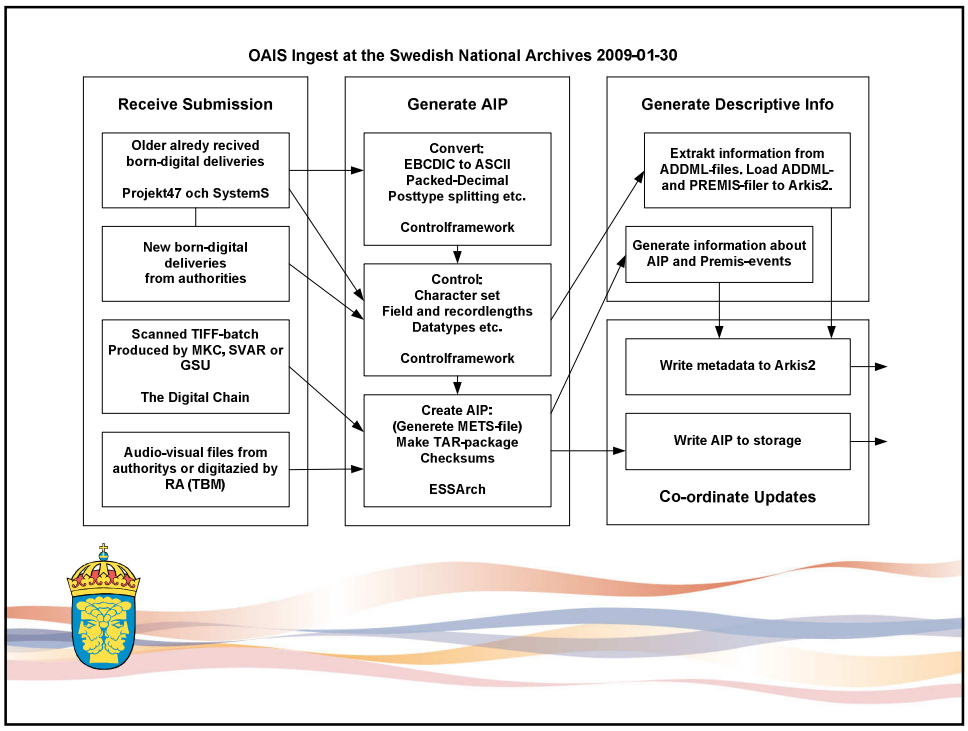
In our database

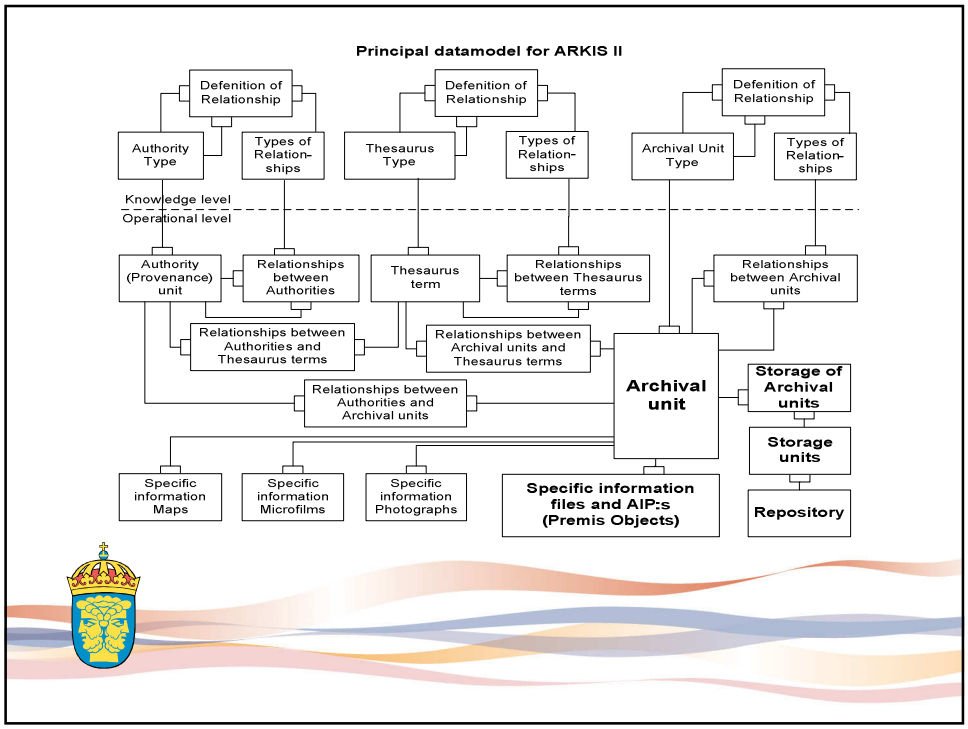
- Incorporated in the archival database ARKIS2
  -A Swedish archival management system
  -Own development
  -All metadata about archives regardless of storage media. (paper as well as digital)
- Connected with the archival object of type born digital
- Used for image files from our scanning projects
- Used as one output from our Controlframework (under development)

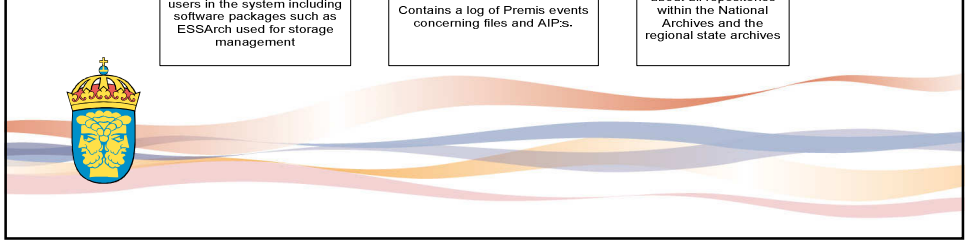
Controlframework for born-digital information




FOSAM / MKC
1 Planning
2 Database registration
3 Fetch originals
4 Preparation
5 Scanning
6 Ocular image control
7 TIFF header update and extract
10 Delivery on LTO-tape
13 LTO-tapes in Stacker
9 Create DJVU-files for viewing
10 Delivery via FTP
11 Image server
12 Import of references to Arkis2

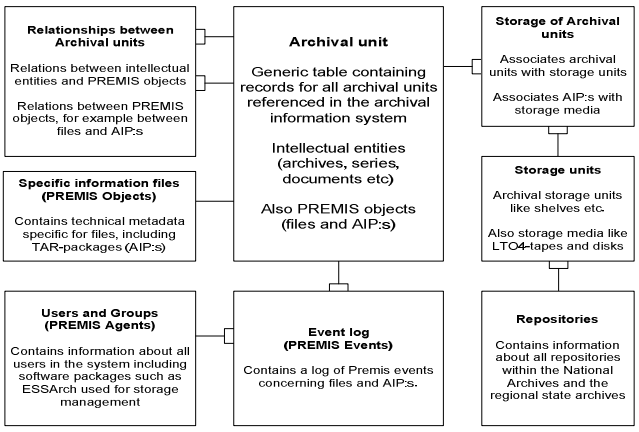

OAIS Ingest at the Swedish National Archives 2009-01-30
Receive Submission
Older alredy recived born-digital deliveries
Projekt47 och SystemS
New born-digital deliveries from authorities
Scanned TIFF-batch Produced by MKC, SVAR or GSU
The Digital Chain
Audio-visual files from authoritys or digitazied by RA (TBM)
Generate AIP
Convert:
EBCDIC to ASCII
Packed-Decimal
Posttype splitting etc.
Controlframework
Control:
Character set
Field and recordlengths
Datatypes etc.
Controlframework
Create AIP:
(Generete METS-file)
Make TAR-package
Checksums
ESSArch
Generate Descriptive Info
Extrakt information from ADDML-files. Load ADDML- and PREMIS-filer to Arkis2.
Generate information about AIP and Premis-events
Write metadata to Arkis2
Write AIP to storage
Co-ordinate Updates



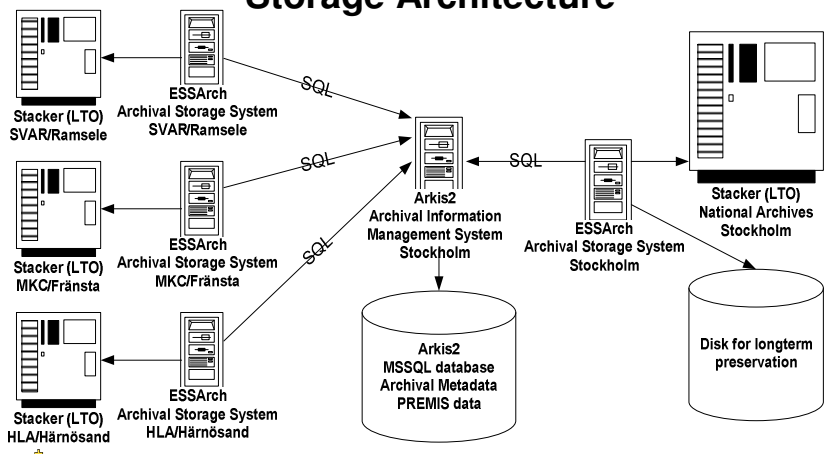


PREMIS implementation

PREMIS Implementation in Arkis II 2009-01-30



Storage Architecture

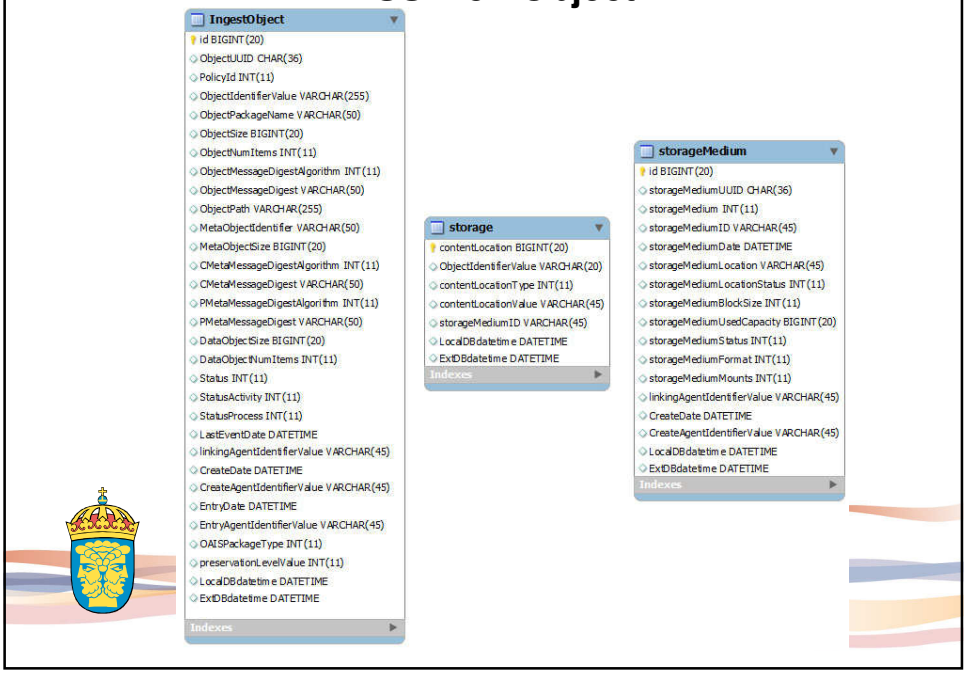


ESSArch

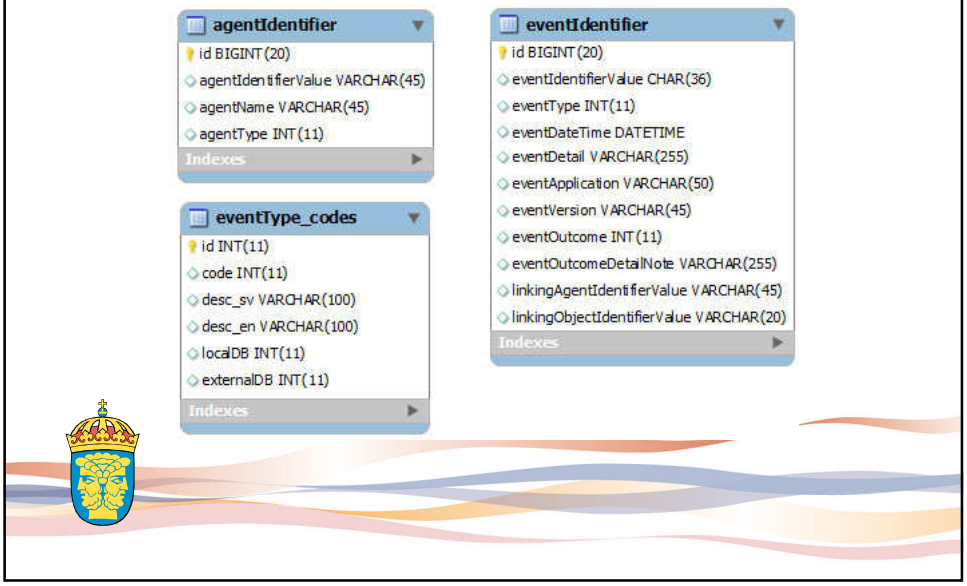
- **Archival Storage system developed by ESSolutions (www.essolutions.se) for the National Archives of Sweden.**
- ESSArch is a back-end system for archival storage according to the OAIS-model. No public interface. Intended for integration with an archival information system (Arkis) or a library system.
- Package, storing and reading AIP:s for longtime preservation. (TAR-format).
- Generates AIP metadata according to METS. The METS-file can contain embedded metadata of PREMIS, ADDML, MIX and XHTML. One AIP consists of one TAR-file with a Content-METS and is followed by a Package-METS.
- Stores AIP:s in one or more bitwise-identical copies on optional storages media. Today LTO-tape and disk.
- Automatically rule based media migration. At NA in current version no automatic format migration.
- Generates, controls and stores checksums. Checksums on both file and package level.
- Logging of all AIP-events according to PREMIS. Also when storages media is handled.
- Stores preservation metadata in a local MySQL-database modeled according to PREMIS 2.0. The local database contains information about AIP:s, storage media (tape and disk) and event logs.
- Updates Arkis through SQL. Information about AIP:s, storage media and events is continuously written to Arkis.
- Physical handling of media (LTO-tapes, harddrives etc.) is handled through a PC-program (RABAR). This program gets updated by a bar-code scanner. The application communicates with ESSArch through a web-interface. Delivery, reception, placement and removal of storage media is administrated through the bar-code scanner.
- Rules for storing on storage media is handled through profiles. A profile can regulate filling of a tape, bufferrate of a tape and so on.

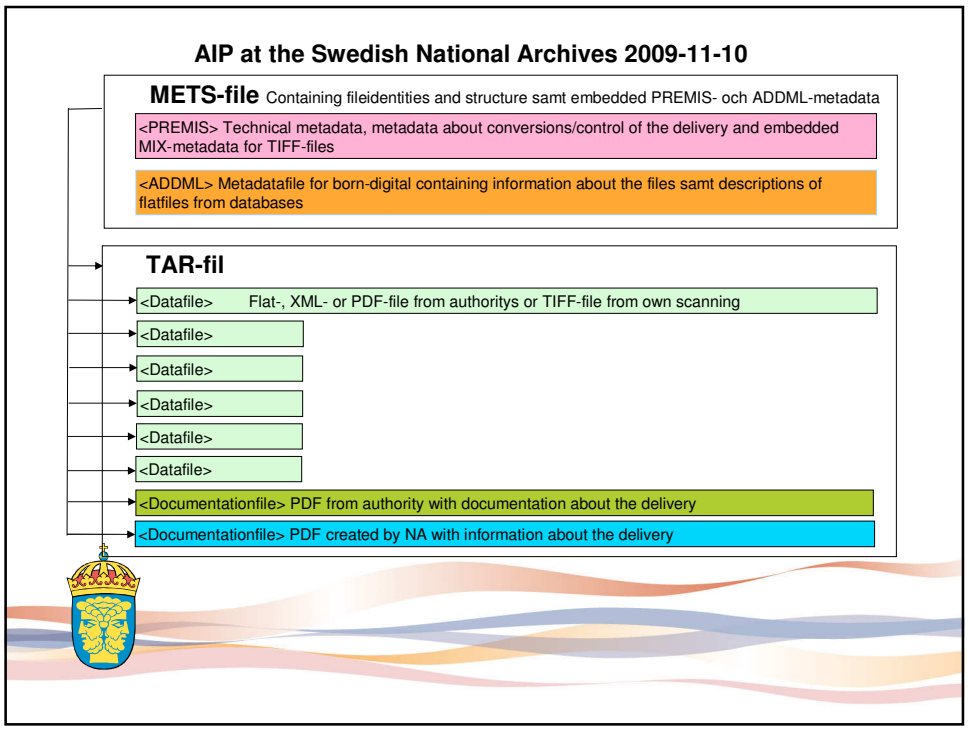
  Based on Open Source, Linux, Apache, MySQL and Python.

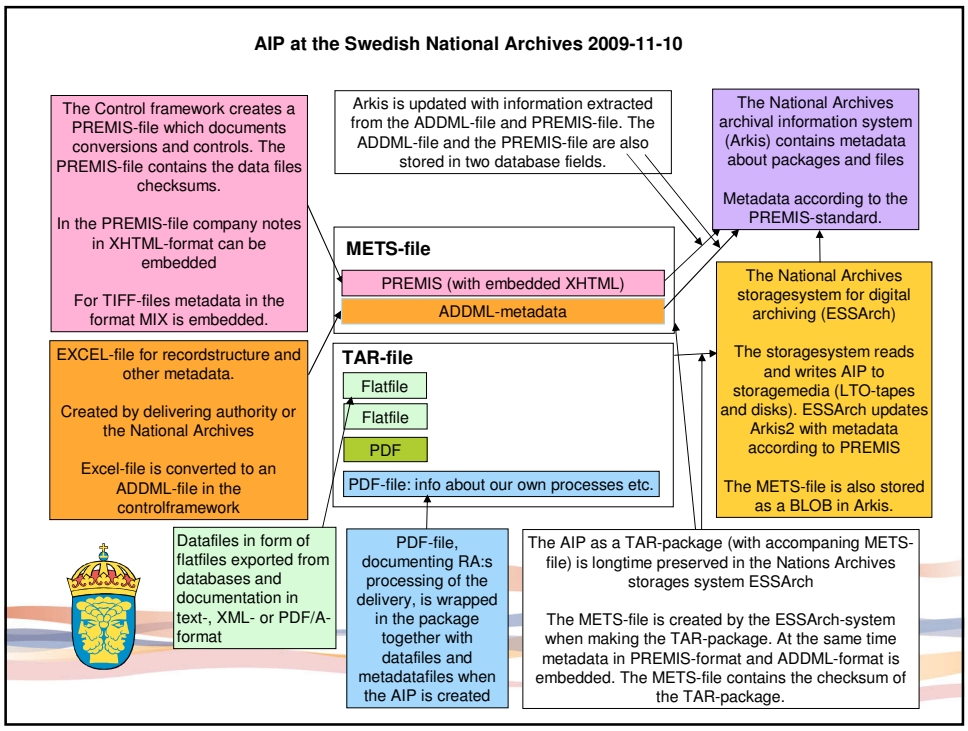
ESSArch Object



ESSArch Event

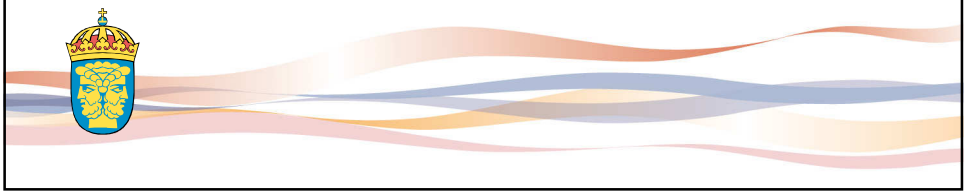



AIP at the Swedish National Archives 2009-11-10
METS-file Containing fileidentities and structure samt embedded PREMIS- och ADDML-metadata
<PREMIS> Technical metadata, metadata about conversions/control of the delivery and embedded MIX-metadata for TIFF-files
<ADDML> Metadatafile for born-digital containing information about the files samt descriptions of flatfiles from databases
TAR-fil
<Datafile> Flat-, XML- or PDF-file from authoritys or TIFF-file from own scanning
<Datafile>
<Datafile>
<Datafile>
<Datafile>
<Datafile>
<Documentationfile> PDF from authority with documentation about the delivery
<Documentationfile> PDF created by NA with information about the delivery


AIP at the Swedish National Archives 2009-11-10
The Control framework creates a PREMIS-file which documents conversions and controls. The PREMIS-file contains the data files checksums.
In the PREMIS-file company notes in XHTML-format can be embedded
For TIFF-files metadata in the format MIX is embedded.
Arkis is updated with information extracted from the ADDML-file and PREMIS-file. The ADDML-file and the PREMIS-file are also stored in two database fields.
The National Archives archival information system (Arkis) contains metadata about packages and files
Metadata according to the PREMIS-standard.
METS-file
PREMIS (with embedded XHTML)
ADDML-metadata
The National Archives storagesystem for digital archiving (ESSArch)
The storagesystem reads and writes AIP to storagemedia (LTO-tapes and disks). ESSArch updates Arkis2 with metadata according to PREMIS
The METS-file is also stored as a BLOB in Arkis.
EXCEL-file for recordstructure and other metadata.
Created by delivering authority or the National Archives
Excel-file is converted to an ADDML-file in the controlframework
TAR-file
Flatfile
Flatfile
PDF
PDF-file: info about our own processes etc.
Datafiles in form of flatfiles exported from databases and documentation in text-, XML- or PDF/A-format
PDF-file, documenting RA:s processing of the delivery, is wrapped in the package together with datafiles and metadatafiles when the AIP is created
The AIP as a TAR-package (with accompaning METS-file) is longtime preserved in the Nations Archives storages system ESSArch
The METS-file is created by the ESSArch-system when making the TAR-package. At the same time metadata in PREMIS-format and ADDML-format is embedded. The METS-file contains the checksum of the TAR-package.

For our suppliers

- A way of letting the suppliers describe the object and the events that have occurred before we receive it.
- XML included (embedded or linked) in the AIP.
- A predefined set of elements that we require.
- Eventually a full PREMIS-file



Suppliers cont.

- Level of **accuracy** depends on decisions of destruction of the data history (e.g. is data about creation, upgrading and so on going to be saved?)
- METS to be used as package information.
  PREMIS inside METS or own file depends on deliverer