Riksarkivet

The Swedish National Archives digital preservation

Mats Berggren, IT-department, 2018-11-29

Swedish National Archives digital preservation

- Born-digital information
- Digitization of documents
- Digital preservation at the National Archives
- The use of standards

Riksarkivet

Swedish National Archives digital preservation

- Born-digital information
- Digitization of documents
- Digital preservation at the National Archives
- The use of standards

Riksarkivet

Recieving born-digital data from agencies

- No fixed delivery time, data files recieved can be new and old
- Deliveries are negotiated between the agencies and the National Archives. Funding are transferred from the agencies to the National Archives
- When agencies are closed down the archives are transferred to the National Archives
- Register laws
- Currently no common record management standard in Sweden

Riksarkivet

Regulations for agencies

- The National Archives issues regulations for digital preservation in the Swedish agencies
  - RA-FS 2009:1, RA-FS 2009:2

- Archive file formats
  - Text files (ISO 8859-1, Unicode)
  - HTML
  - XML (also GML and SGML)
  - PDF (PDF/A-1)
  - JPEG, TIFF and PNG

Common deliveries of ”born-digital”-material

- Databases, data exported as textfiles or XML-files
- Web-pages, Agency web sites are archival data
- Record management systems, database and PDF-documents
- Collections of documents
- Government committes, many small deliveries

Swedish National Archives digital preservation

- Born-digital information
- Digitization of documents
- Digital preservation at the National Archives
- The use of standards

Riksarkivet

Digitization of documents

- Scanning of documents, church records etc, MKC Fränsta
- Microfilm scanning, SVAR Ramsele
- Microfilm scanning by FamilySearch in Salt Lake City, USA. Delivery to SVAR Ramsele. Church records and judicial records

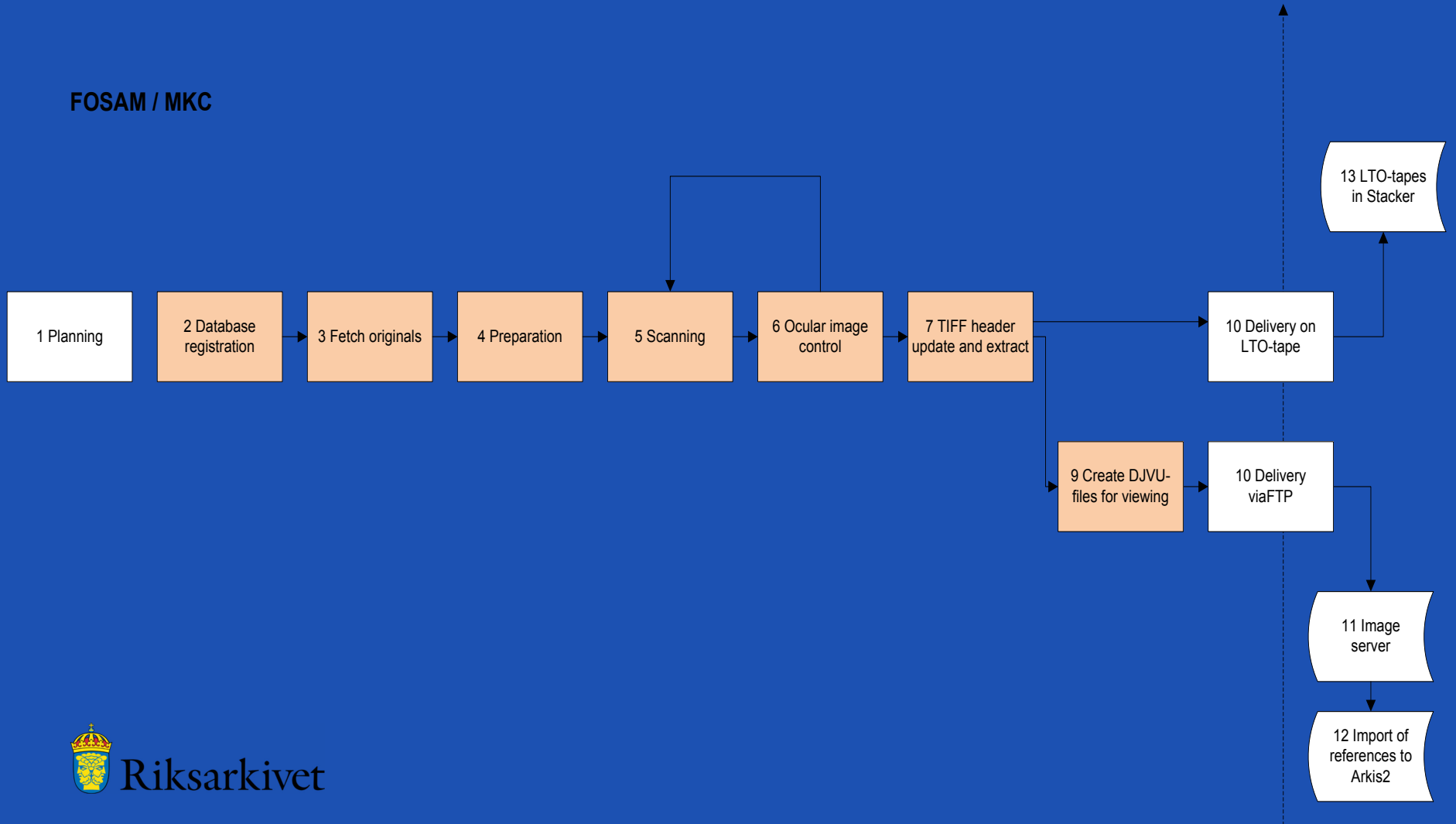
Image formats

- Most scanning projects within the National Archives produce raw TIFF-files of these three types:
  - TIFF/IT (TIFF 6.0), Grayscale, BitsPerSample=8, 300dpi
  - TIFF/IT (TIFF 6.0), Group4 B/W, BitsPerSample=1, 400 dpi
  - TIFF/IT (TIFF 6.0), Colour RGB, BitsPerSample=8x3, 300 dpi

- DJVU, Used for presentation and public access. Converted from TIFF. Proprietary format

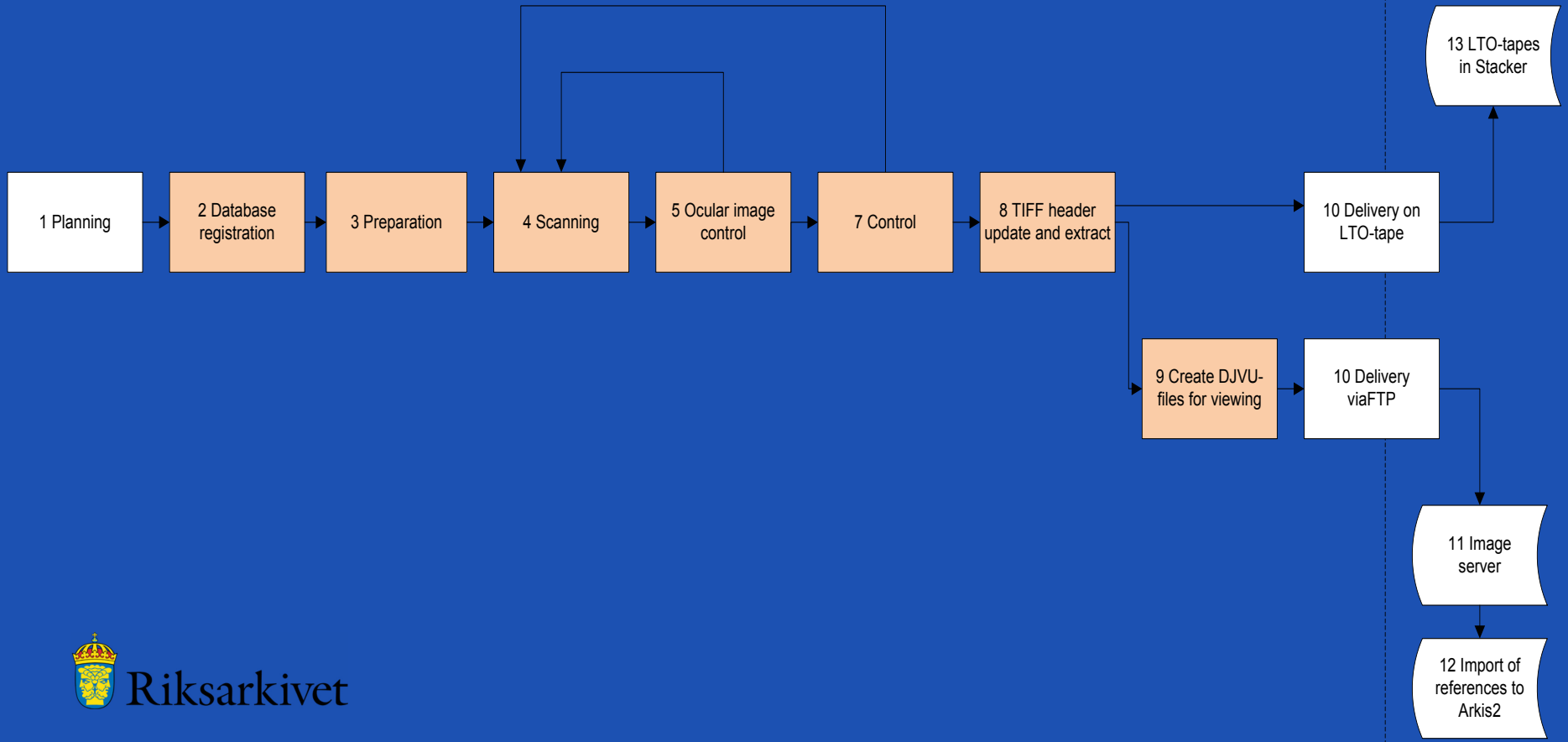
- JPEG, Used by a few projects. Accepted as delivery format from agencies

FOSAM / MKC



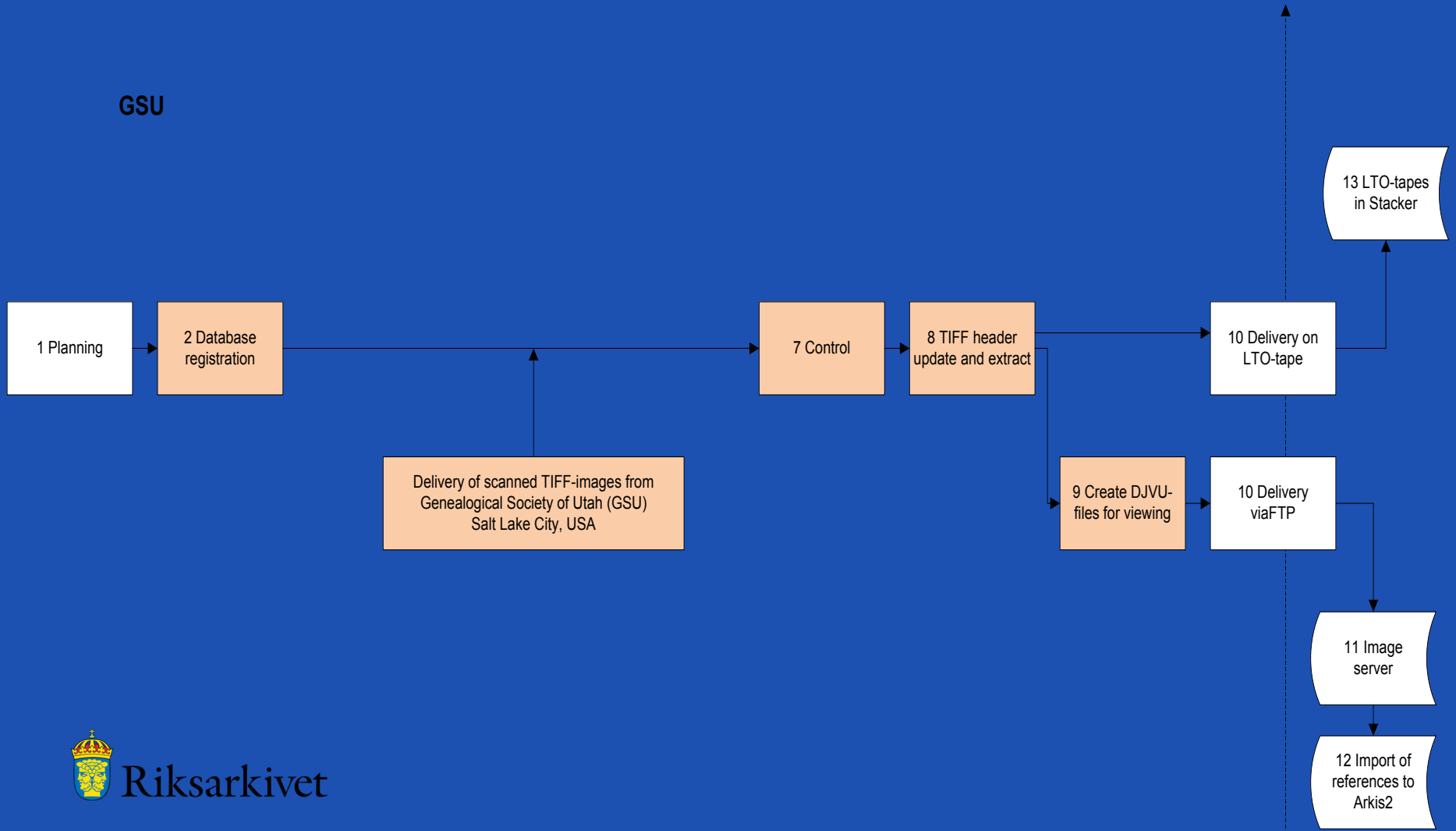
Riksarkivet

Microfilm scanning



Riksarkivet

GSU



Riksarkivet

Digitization of audiovisual media

- Project DIANA:
  - Digitization of audiovisual media, audio and video
  - Digitization done in house by the National Archives
  - Digitization also done by the Royal Library for the National Archives
  - Project started 2015, digitization started 2017

Riksarkivet

Audiovisual formats

- Formats for long term storage:
  - Audio: WAV
  - Video: Matroska / FFV1

- Presentation formats:
  - Audio: MP3
  - Video: MPEG-4

Swedish National Archives digital preservation

- Born-digital information
- Digitization of documents
- Digital preservation at the National Archives
- The use of standards

Riksarkivet

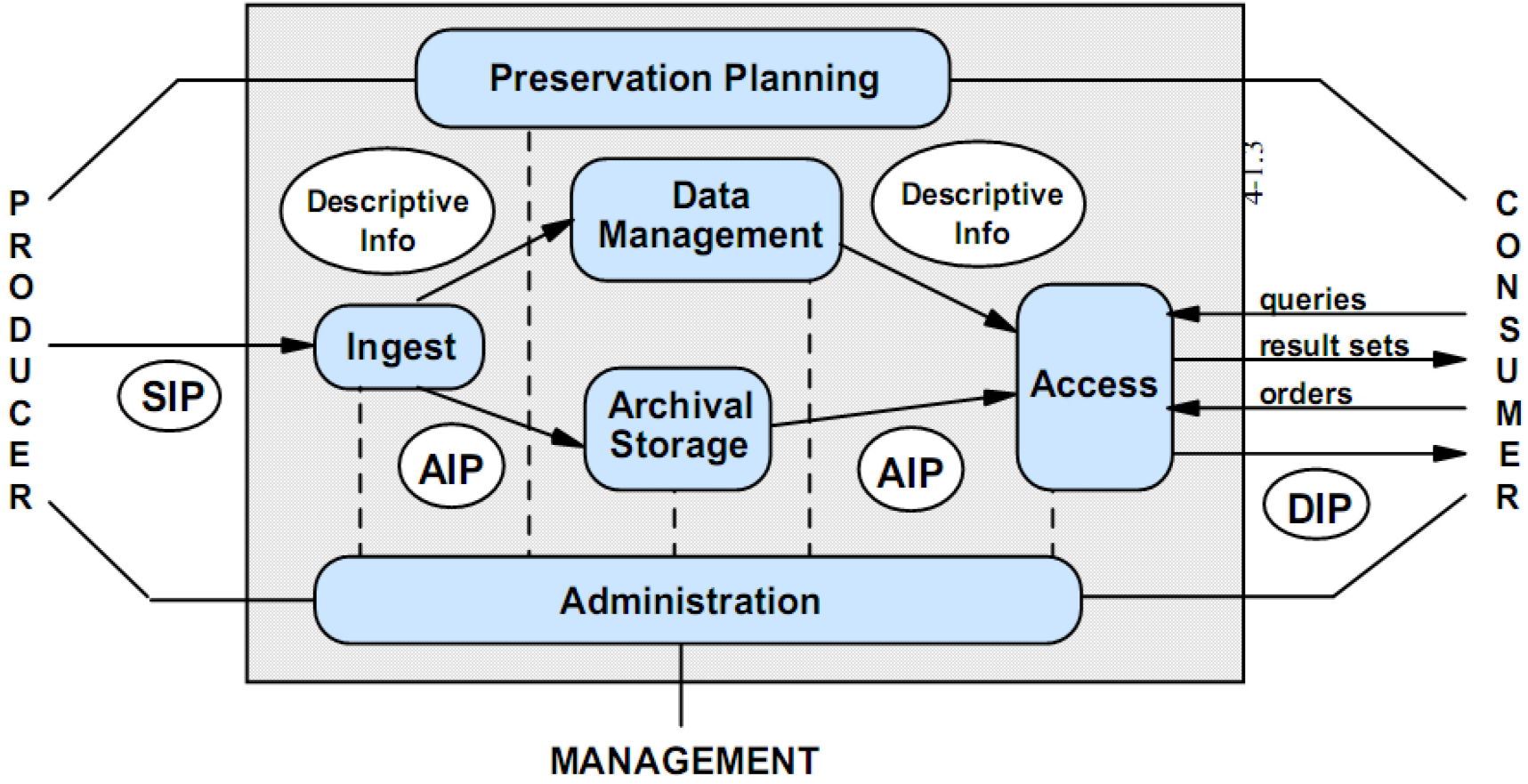
Digital preservation at the National Archives

- History:
  - Archival deliveries of digital data since the 1970:s
  - Large scale digitization of documents since 2003
  - A Hierarchical Storage System (HSM) installed 2004
  - A new storage platform becomes necessary 2007
  - A new platform RADAR is developed based on the OAIS-model
  - RADAR (archiving digital images) since 2009
  - RADAR (archiving “Born-Digital” from agencies) since 2013

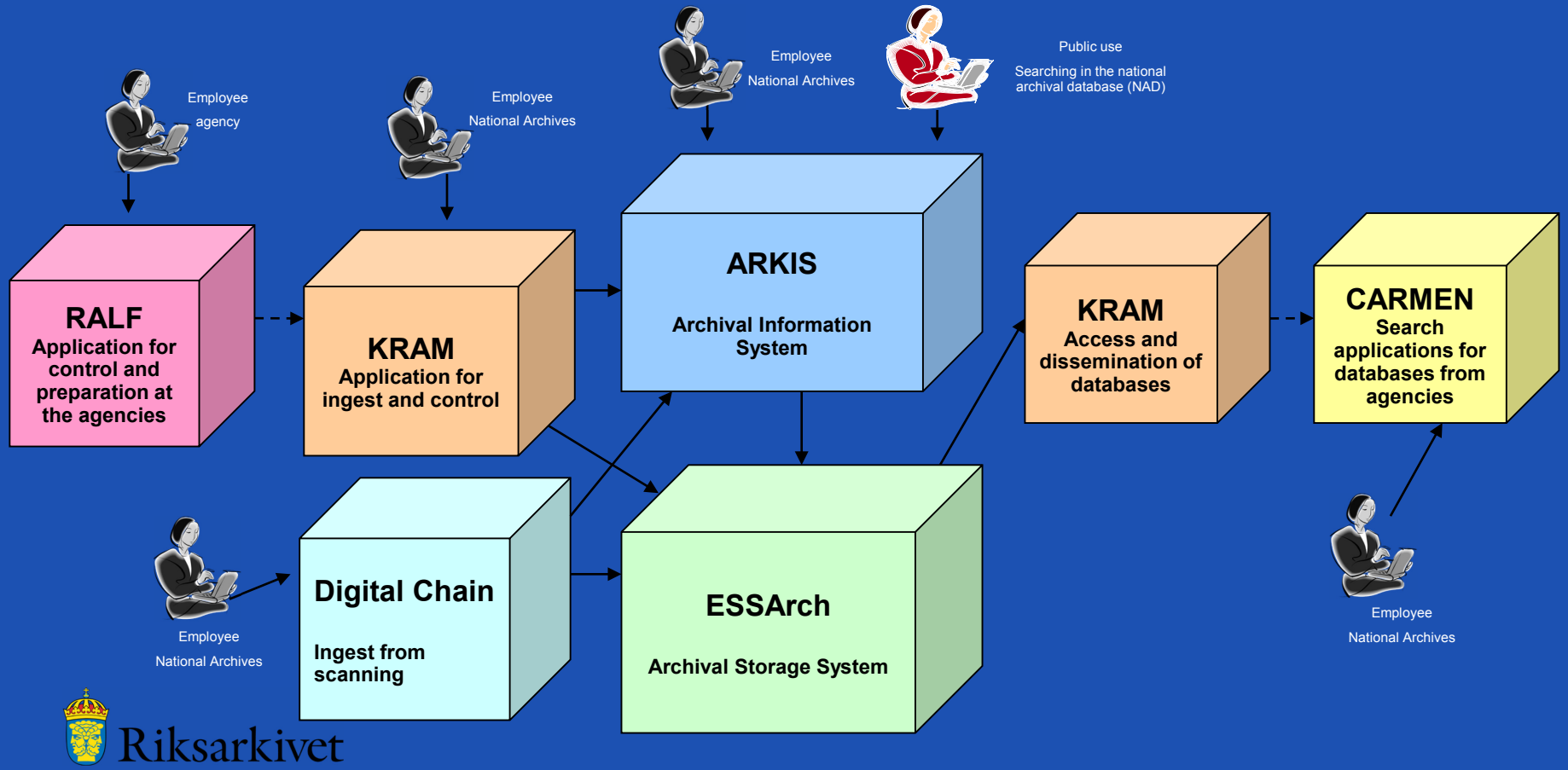
A platform for digital preservation (RADAR)

- What is RADAR:
  - Digital preservation of both “Born-Digital” data and digital images
  - Several copies in geographically separated locations
  - Provenance and descriptive metadata (ARKIS/NAD)
  - Technical metadata and preservation metadata (ARKIS)
  - Standardized metadata formats (METS, PREMIS etc)
  - Specially developed system for archival storage (ESSArch)
  - Can be extended with new modules and tools
  - Media migration (Not automated)
  - Scheduled media validation (Not automated)
  - Format migration (Not automated)

OAIS model


Preservation Planning
Descriptive Info
Data Management
Descriptive Info
PRODUCER
Ingest
SIP
AIP
Archival Storage
AIP
Access
queries
result sets
orders
DIP
CONSUMER
Administration
MANAGEMENT
4-1.3

The Swedish National Archives platform for digital preservation (RADAR)



RADAR parts

- **RALF** – The National Archives tool for preparation of archival transfers. Used by agencies. Can do basic controls and creates a submission package (SIP)
- **KRAM** – Control and validation framework. An application that controls and validates SIP:s from agencies. KRAM kan also be used to convert data from older transfers. KRAM is also used to load files exported from agency databases into a SQL-database
- **Digital chain** – The National Archives digitization of documents. Masterfiles in TIFF-format is packed in AIP:s and stored for long term preservation in RADAR
- **ARKIS** – The National Archives archival information system. Contains archival descriptions and metadata about all archival objects, including digital objects
- **ESSArch** – The National Archives ”storage management system”. Manages the physical storage on tape (LTO4) and disks. Packs AIP:s in TAR-format. Performs checksum-controls. Logs all ingest- and dissemination-events. ESSArch is an Open Source application and is also used by the National Archives in Norway
- **CARMEN** – Search applications for databases (about 30) delivered from agencies

Digital information at the National Archives

- Born-digital files from agencies: about 8 TB
  - Currently in RADAR: 1972 AIP:s (about 6.1 TB)
- Audio-video files and multimedia: approximately 100 TB (so far)
- Digitized paper volumes (one AIP per volume): 524144
- Digitized images (TIFF-format): 2.9 PB (In one copy)
- Images total: 208.2 million
- Images published on Internet: 65.7 million
- DJVU-files (presentation format): 40 TB
- Total storage: 5.8 PB (Two copies)

Swedish National Archives digital preservation

- Born-digital information
- Digitization of documents
- Digital preservation at the National Archives
- The use of standards

Riksarkivet

Archival standards

- ISAD(G) and ISAAR(CPF)
  - The Archival information system ARKIS is modelled after these standards
- EAD and EAC-CPF
  - These formats are used as exchange formats for archival description information in Sweden
  - Supported by several commercial archival information systems
  - Import and export functions in ARKIS
  - A new Swedish EAD and EAC-CPF adaptation (FGS)
- OAIS
  - Widely adopted in Sweden not only by the Swedish National Archives
  - Several commercial E-Archive system claim to be OAIS-compliant

Standards for preservation metadata

- **METS** (Metadata Encoding & Transmission Standard) - Structure for encoding descriptive, administrative, and structural metadata (DLF/LOC) **(2004)**
- **PREMIS** (Preservation Metadata) - A data dictionary and supporting XML schemas for core preservation metadata needed to support the long-term preservation of digital materials (OCLC/LOC) **(2005)**
- **MIX** (NISO Metadata for Images in XML) - XML schema for encoding technical data elements required to manage digital image collections (ANSI/NISO) **(2006)**
- **EBUCore** – XML-format for metadata for audio files and video files. Developed and supported by the European Broadcasting Union (EBU)

Other formats

- ***ADDML*** (Archival Data Description Markup Language) – XML-format used by the National Archives of Norway and Sweden, XML-format for describing flat files exported from databases **(2001, 2009)**

Thank you!
Tack så mycket!

mats.berggren@riksarkivet.se

Riksarkivet