Mass digitisation supervised by the archives:

Advantages and Ventures in Finland

Ville Kajanne
Head of Unit
28/06/2023

THE NATIONAL ARCHIVES OF FINLAND

Part 1.

Goals and accomplishments 2017 - 2022

Introduction

Focus of this presentation:

✓General description of the initial approach and some results of the mass digitisation (MD) in Finland

✓Future outline of digitisation in strategic, administrative and managerial levels

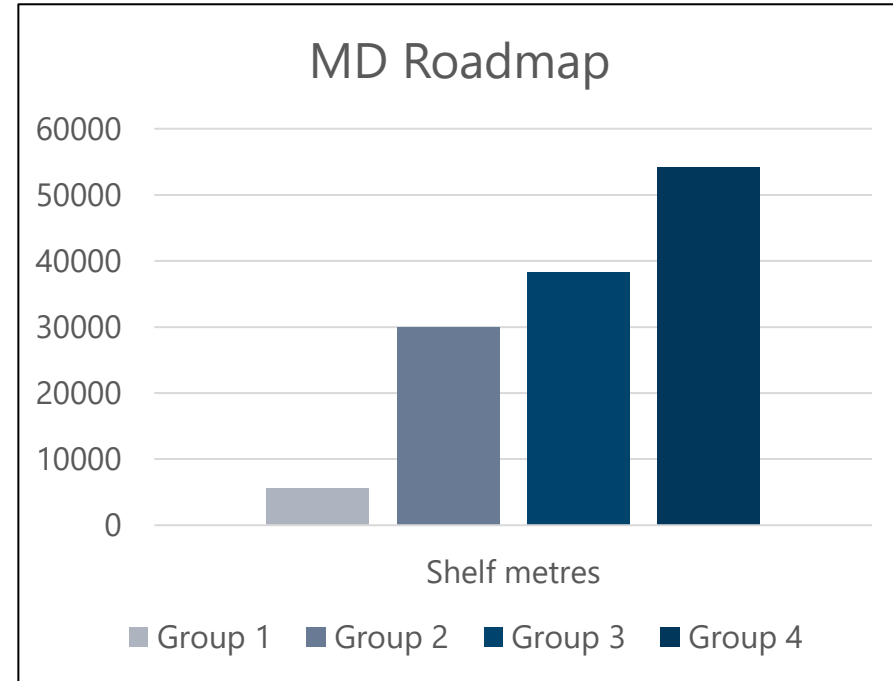✓Some present challenges and goals

The approach 2017

➤ NAF digitisation initiated at early 2000 – mainly introducing access files/image copies at digital platform known as "Digital Archive"

➤ The increasing pressure to accept born-electric records, keeping up with authenticity requirements while struggling with less and less available physical storage space

➤ The Finnish government's 2017 brave visions of digitisation as a tool for better services and cutting costs

➤ Goverment's Letter of Intent 2017

 ✓ Permanently preserved records must be archived digitally

 ✓ NAF co-ordinates the digitalisation process and work

 ✓ All records held by the agencies must be digitised by the 2030

➤ Out-of-date Archival law (1994) with adjustments at 2016:

 ✓ Disposal of analogue records is granted as long as the permanent preservation has been guaranteed

➤ In line and with support with Data Management law 2021:

 ✓ Permanently preserved records must be available digitally from 2023 onwards

THE NATIONAL
ARCHIVES OF FINLAND

MD Roadmap

- Based on broad enquiry in 2017, supplemented gradually

- App. 128 shelf km

- App. 100 organisations (federal offices)

- The Board of Roadmap: led by the NAF, representation by ministries/governmental sectors

- The criteria:
    1) Suitability for digitisation (quality and speed)
    2) Improving usability by digitisation
    3) Other benefits gained by digitisation (reducing storage facilities, improving cost-effectiveness etc.)
    4) Exclusions (data protection, intrinsic value -> app. 80-100 shelf km)

- Mutually recognized benefits from the governmental point of biew, good co-operation, "playing as a team"

THE NATIONAL
ARCHIVES OF FINLAND

MD "Philosophy"

Paper	→	Image	→	Data	→	Information
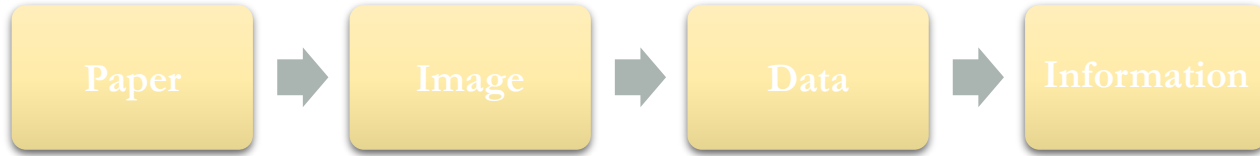
- Its is not about creating images of analoguely existing records
- The data content of the record will not be disposed, only the storage format will change
- The conversion of analogue records into digital format is a process that must be documented
  - Digitising is a part of records' life-cycle
- Instead mass-digitised records can be used as born-digital material and alongside of them (preservation and access platforms & the use of machine learning methods):
  - Full-text search & inquiry
  - Ability to analyze and process data
  - Automatic description (contextual metadata generation)

THE NATIONAL
ARCHIVES OF FINLAND

MD Production framework 2018 - 2022

➢ Broad administrative legitimation of MD

➢ Organisation and principles:

  ✓ Production managed and run by the NAF, based on MD Roadmap

  ✓ After having been digitised analogue records shall be disposed

  ✓ Records with cultural-historical value (intrinsic value) cannot be mass-digitised

➢ Costs (funding) mainly carried out by the NAF

  ✓ App. 1,5 – 2 m € / a.

• To summarize:

  • It has served interests of both federal agencies as well as the NAF to digitalise records

  • Easier access and better cost-effectiveness to agencies, better quality and carefully monitored digitisation process of records

THE NATIONAL
ARCHIVES OF FINLAND

5,5

Km. of records

THE NATIONAL ARCHIVES OF FINLAND

84
Terabytes of master file data

80
Million master files

100 000
Master files daily

Production

- MD Team (Mikkeli & Helsinki) of 20-25 people (slightly diminished from 2022)

- Two units, one set of guidelines, shared mutual virtual environment, one goal

- Activated
  - Helsinki 2019
  - Mikkeli 2021

- 2022 was the first year of the production without major development or expansion

- In-house production solutions and the digitisation environment in general have remained mainly same as they were designed for the pilot stage (2019)

- So far, main effort has been given e.g. to the digitisation process, quality validation, data security, recruiting & individual competence

- 2022 – 2023, machine learning has been developed alongside the digitisation in co-operation with the separate DALAI project

Year	Stage
2017	Planning
2018	Proof of Concept - Pilot
2019	Pilot
2020	Production
2021	Expansion
2022	Full production
2023	New organisation

The Hardware

- 16 Kodak document (i5250, i5650) scanners & flatbeds (10 at Mikkeli)

- 1 IBML Image Trac 6400 high volume scanner

- 2 Contex large format scanner

- 2 Zeutschel book scanners

→ Qualifications for a special material exist, even though not expected

10

Part 2.

Digitisation
Visions and Strategies
2023 - 2035

THE NATIONAL
ARCHIVES OF FINLAND

MD Finance

➤ New government (2023), new funding

➤ There are risks of losing continuity and professionalism

➤ Possibly split funding between partners -> MD production no more "calling the shots" on its own

➤ If other agencies participate, they will have their "veto"

➤ MD must have its production lines prepared for diverse records with certain variety in terms of physical condition etc.

MD Costs (M €)

- Personnel
- ICT & Maintenance
- Aqcuisitions
- Other
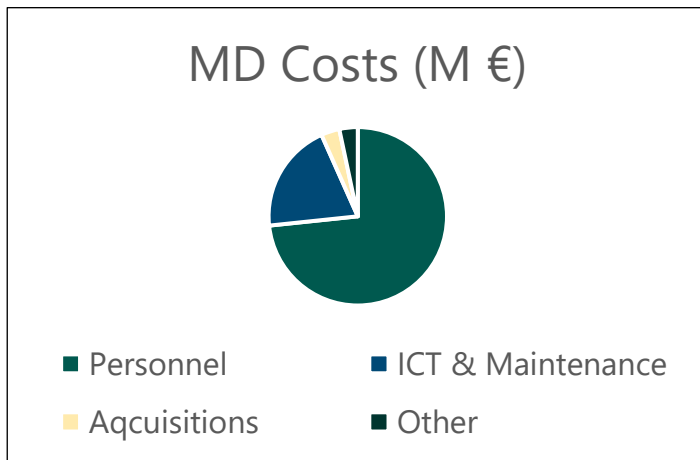
What will be digitised up to 2035?

➢ NAF Digitisation Strategy 2035 confirmed this spring

➢ Choises will be made by the Board of Digitisation (NAF-led new steering committee)

➢ The criteria of prioritizing records and the order of digitisation will be renewed and possibly specified

➢ The scope of the digitisation will be the content, not the quantity of production:
  ✓ What customers want and who they are?
  ✓ User statistics & customer data
  ✓ On demand digitisation
  ✓ Queries

➢ Of course there will be other criteria as well (physical condition etc.) – funding as a novelty?

➢ Digitisation Strategy Target Points:
  1) Customer-oriented digitization improves the effectiveness of digital data and availability
  2) Digitisation must have a clear and efficient process and implementation (production)
  3) High-quality, straightforward and transparent digitisation

THE NATIONAL
ARCHIVES OF FINLAND

Into the Making of the Digitisation Strategy

➢ Board of Digitisation will hear representatives of different experts/customer groups (universities, research projects, genealogical societies, governmental agencies etc.)

➢ Strategy will be interwoven with digital services development strategies in the larger governmental level

➢ There will be less physical visits to the reading rooms and more on demand digitisation

➢ Larger production (roadmap-based) will be centralized in a few premises

➢ The quality (permanent preservation) is the essence of the production

➢ New Archival law will outline more clearly boundaries between agencies and NAF (and make customer service easier)

➢ NAF Digitisation is willing to have an international co-operation

Government report on the digital compass sets the course for Finland's digital transformation

Ministry of Economic Affairs and Employment, Ministry of Finance, Ministry of Transport and Communications

20.10.2022 14.15 PRESS RELEASE

The Government session on 20 October discussed the report on Finland's digital compass. Finland's digital compass is a national strategic roadmap extending to 2030 that provides an overview of Finland's digital transformation and provides the direction for national development work.

The digital compass sets national targets for the effective use of digital systems so that Finland can succeed in the ongoing transformation. It strengthens the shared understanding of the benefits, concepts and direction of digitalisation and the data economy.
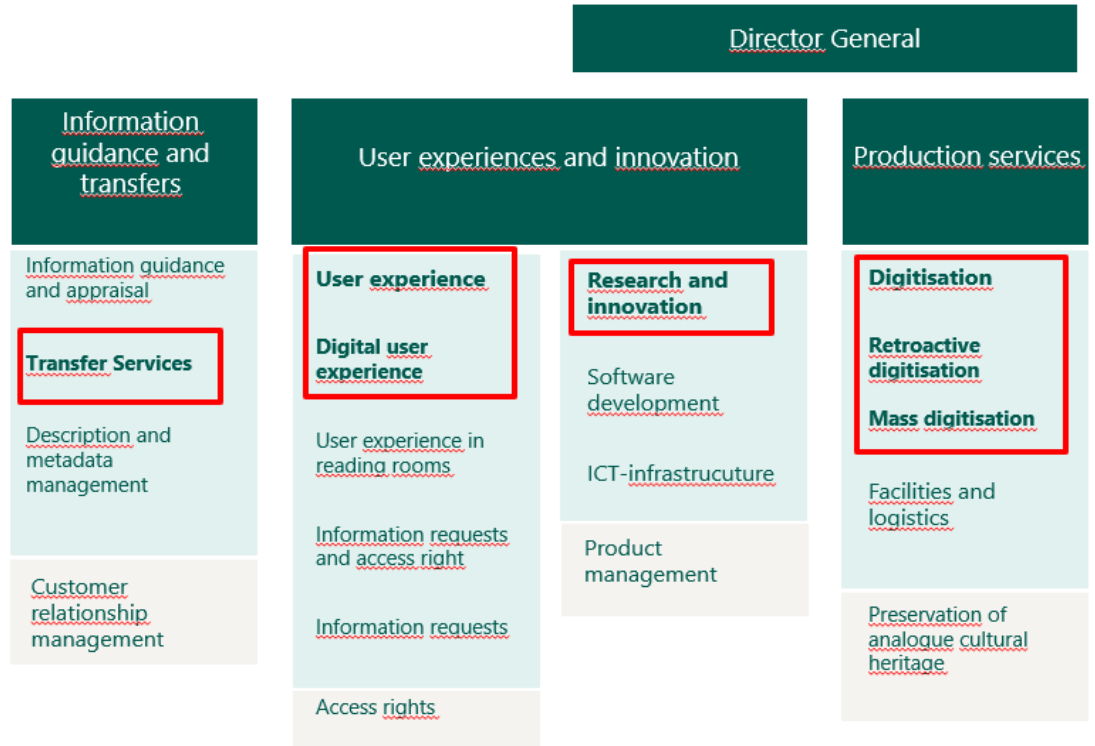
THE NATIONAL ARCHIVES OF FINLAND

Part 3.

Organizing and Developing Digitisation 2023
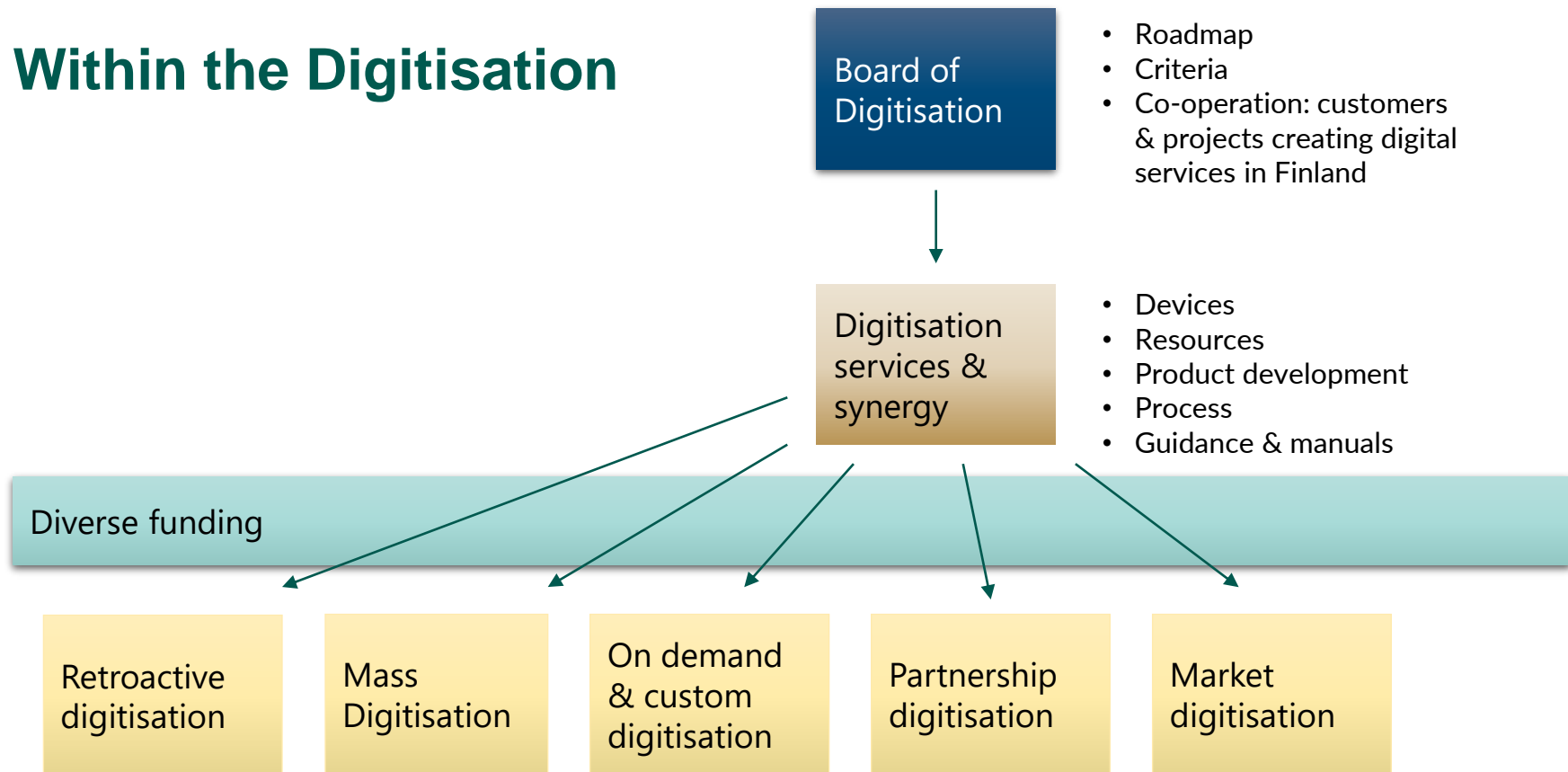
Digitisation and other related units in the NAF

➢ From now on MD will no longer individually make decisions on every step of digitisation process

➢ Instead digitisation teams can fully concentrate and use their resources on issues regarding the quality and efficiency of the digitisation
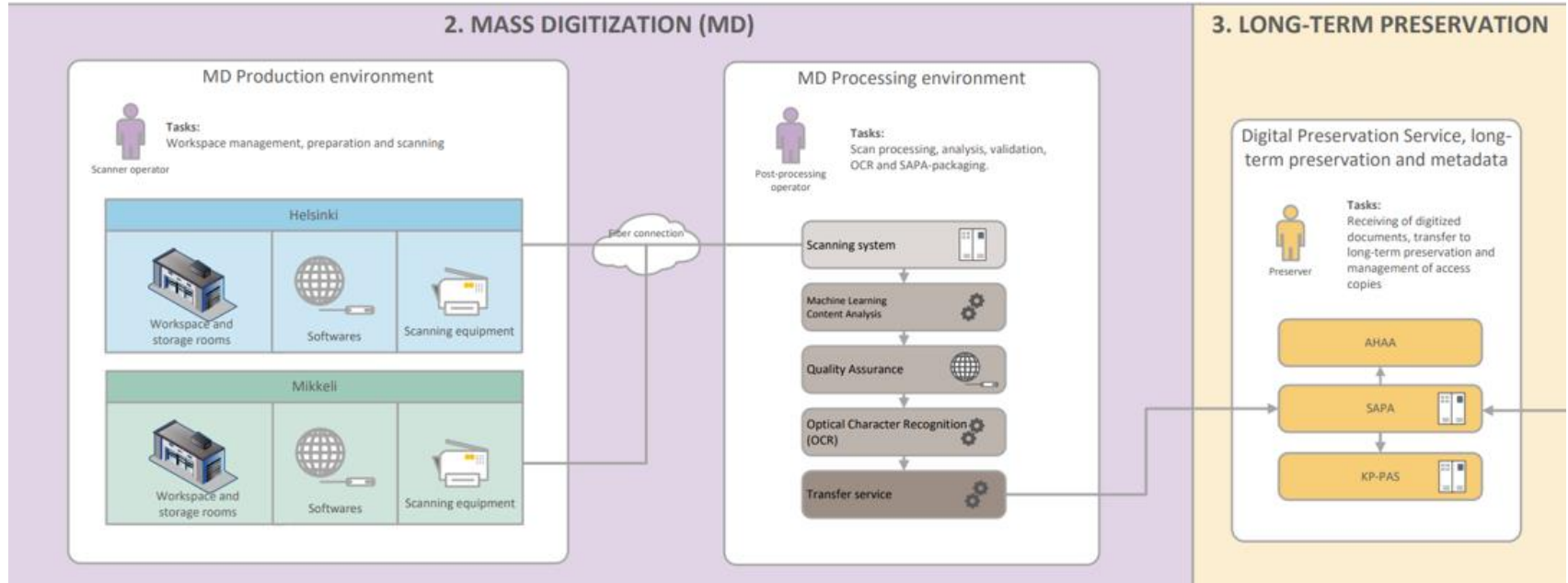
National Archives Organisation

Director General

Information guidance and transfers

Information guidance and appraisal

Transfer Services

Description and metadata management

Customer relationship management

User experiences and innovation

User experience

Digital user experience

User experience in reading rooms

Information requests and access right

Information requests

Access rights

Research and innovation

Software development

ICT-infrastrucuture

Product management

Production services

Digitisation

Retroactive digitisation

Mass digitisation

Facilities and logistics

Preservation of analogue cultural heritage

Within the Digitisation

Board of Digitisation

- Roadmap
- Criteria
- Co-operation: customers & projects creating digital services in Finland

Digitisation services & synergy

- Devices
- Resources
- Product development
- Process
- Guidance & manuals

Diverse funding

Retroactive digitisation

Mass Digitisation

On demand & custom digitisation

Partnership digitisation

Market digitisation

THE NATIONAL ARCHIVES OF FINLAND

MD Environment

MD Digitisation Process

Scan → Analyze → Validate → OCR → Transfer
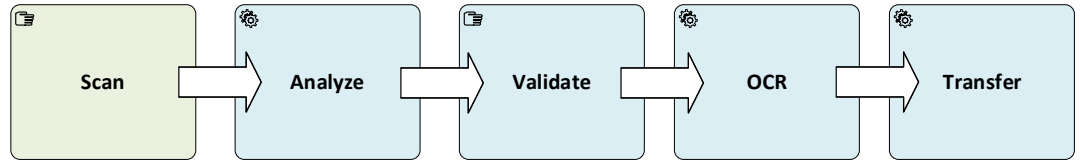
And let's not forget about managing the analogue collections and the storage space.....
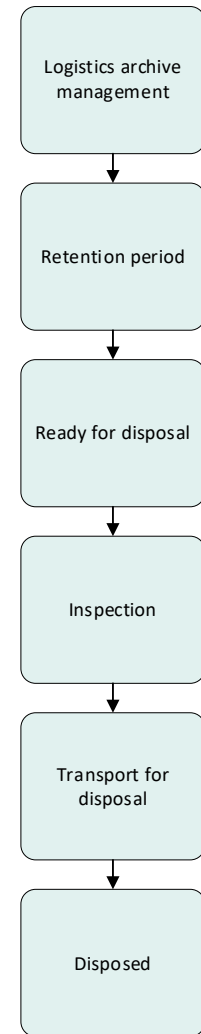
Ongoing projects (MD)

- Required development in solutions to commence with the disposal process
- Scanning software development (mainly more offline working options)
- Validation application development
  - To run all records through image orientation and empty image API's
  - Develop the application to minimize the human labour (use of API's, e.g. pipeline to show only rescanned images and the ones which have already been API-processed)
  - Make the application as user-friendly as possible
- Further ahead:
  - Also "retroactively" digitised analogue records could be disposed? (Less stack space needed)
  - Virtual digitisation environment will become even more virtual? -> (NAF Cloud strategy, piloting 2024 onwards)

Criteria for disposal of analogue material – preconditions

- Mass digitization contracts have been signed

- The metadata in the AHAA cataloging system has been approved and the material is managed by NAF

- The transfer certificate has been delivered to the authority that transferred the material

- Digitization has been approved in validation

- The data of the archival unit includes the entry in SAPA + KP-PAS

- Information about the digital manifestation can be seen in the archival unit in AHAA metadata system

- The archival unit contains the life cycle event entry "digitized and tranferred to the long-term storage" (PAS-AIP-ID)

- The production team has done an ASTIA-inspection (the images have transferred to the display system)
  - The usabilitys of the digitized material is ensured by a sample check

- Retention period after digitization
  - The alert time is triggered by an email to the the team reponsible for authority control of the material
  - Retention period at least 6 months (customizable)

KANSALLISARKISTO

Disposal Process

- Digitisation production will control and see through operational steps of the disposal

- Major development in Logistics application

- Changes in Archive metadata management (AHAA)
  - Disposal of analogue manifestation must absolutely be introduced as a life-cycle description
  - New hierarchy model will rebuild the hierarchy structure to better serve production and disposal as a part of it (disposal will be organized according "transfer unit")

- Multiple process steps to track storage units during disposal

Logistics archive management

↓

Retention period

↓

Ready for disposal

↓

Inspection

↓

Transport for disposal

↓

Disposed

Thank You!

THE NATIONAL
ARCHIVES OF FINLAND