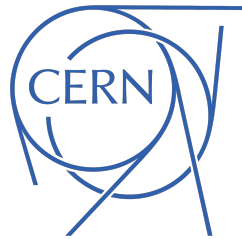




**ARCHIVER**

ARCHIVING AND PRESERVATION FOR RESEARCH ENVIRONMENTS



# The CERN Digital Memory Use Case

Jean-Yves le Meur



# archiver-project.eu in two words

## H2020 Programme - Pre-Commercial Procurement

R&D services for archiving and digital preservation by Industry

Procurer organisations: **CERN**, DESY, PIC, EMBL-EBI

3.4 Million euros budget from Jan 2019 to June 2022

4 stages: Bids, Design, Prototype, Pilot

Started with 5 consortia, reduced to 3 in Dec 2020 to continue into the prototype phases:

Libnova, T-System, Arkivum



<https://zenodo.org/record/7691976#.ZEjVdEjP3u0>

# Digital Memory: context

Digital Memory at CERN started with the preservation of the 20th century analog data

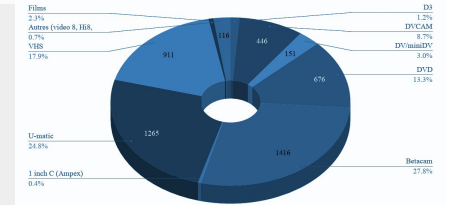


350'000 access files : up to 3'000'000 files with all the formats, labels, replicas and md5

Stored on CERN CLOUD (EOS)



120 films and 5'000 videotapes  
13 different carrier types  
(about 10% loss)



Audio recordings of committees  
7'500 reels and tapes  
(1'700 Kg)



300'000 Colour photos, slides and negatives

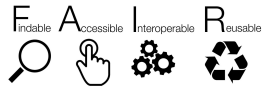
some very degraded  
→ <http://volmeur.web.cern.ch>



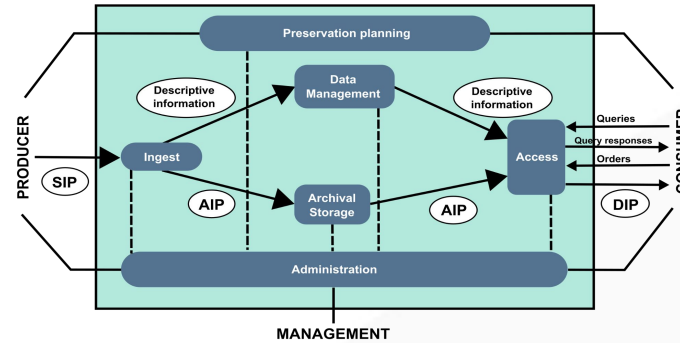


# Digital Memory objectives

Digital Memory is addressing the issue of the long-term preservation of Digitized and Digital-born data



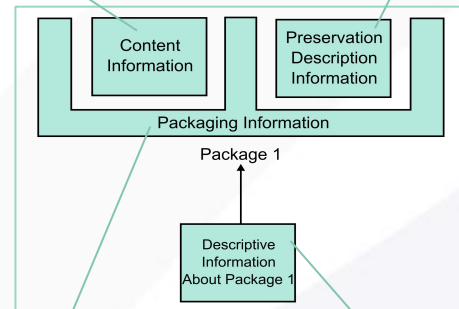
Tools and platform to process records from Information Systems into Preservation Bags (SIPs, AIPs, DIPs) with FAIR-compliance



The **OAIS** reference model rules how preservation should be applied - ISO/Seal to be a Trusted Digital Repo

E.g. title, author, date, etc

E.g. checksum, digital signature



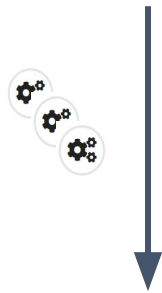
Preserved Bag aka **AIP**

E.g. directory structure, filenames, tape marks

E.g. access control, finding aids

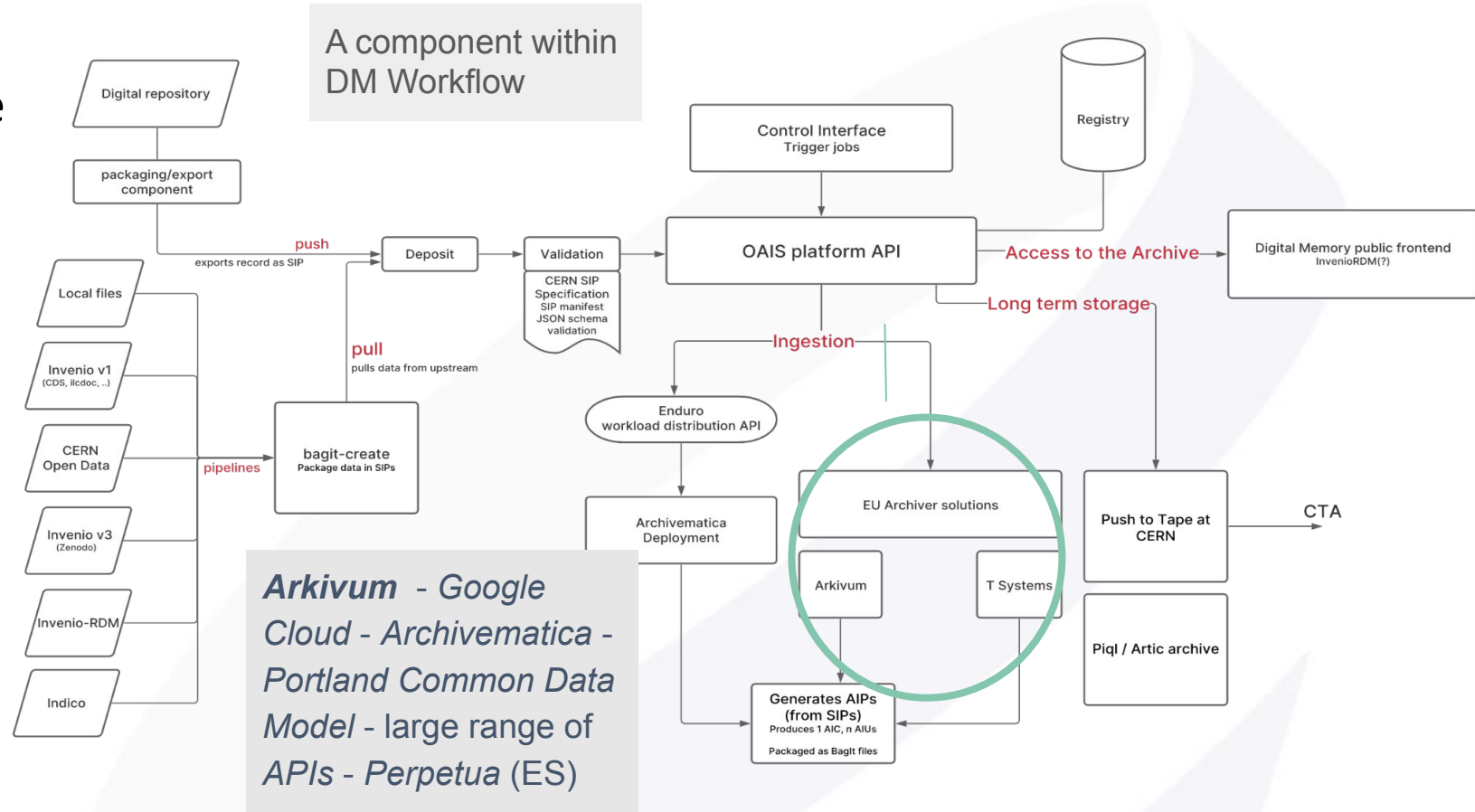
# Digital Memory: processes

Digital Memory is trying Archiver.eu to outsource the conversion of original data into preserved data



3 consortia tested to run as a “factory of preservation Bags”

:

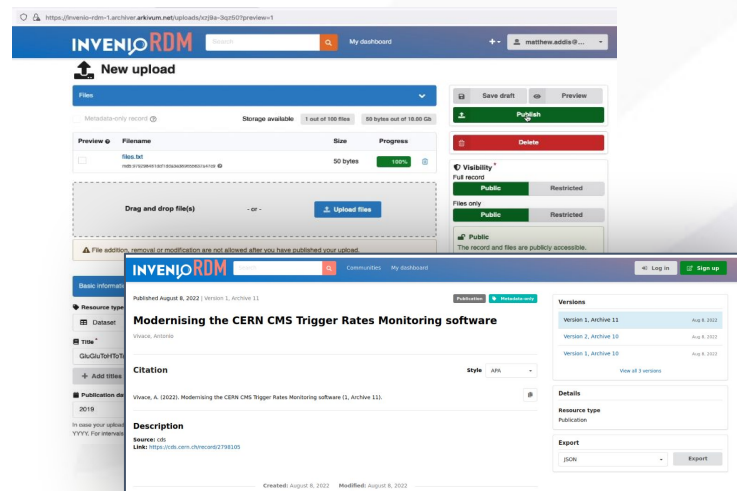
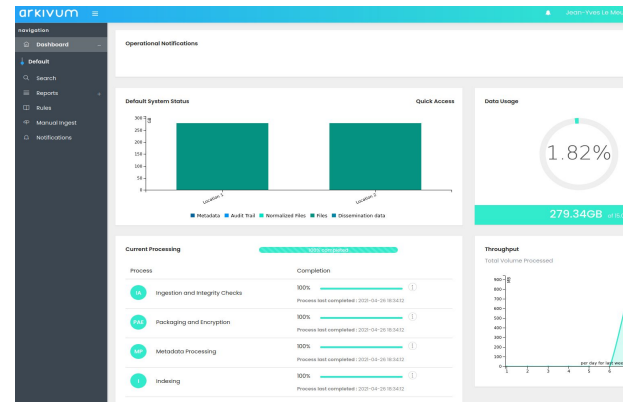


# Digital Memory: results

Digital Memory use case has validated a scalable and integrated outsourced Bag Factory



Applied to rescue the ILC document server with a dedicated pipeline - 4'845 ingests in 48 hours with 40 instances



## Features

Metadata handling ;  
*Logical structure aggregation ;*  
 Versioning ;  
 Automation

## Scalability

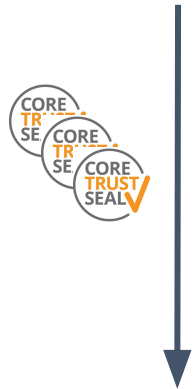
Ingestion of data with manifests ;  
 Large sets in one run (500TB/day);  
 Export of data in bags via G-buckets,  
 including checksums and full new  
 collections ;  
*Archivematica ;*

## Integration

Datapools ;  
*InvenioRDM ;*  
 DataCite schema ;  
 Xrootd export

# Digital Memory: future

How Digital Memory could benefit from this project ?



Procuring CERN Archival Information Packages services to process a few 'closed' collections

Home > CERN Moving Images > CERN Digitized Videos (Public) > Search Results: 1960-1970





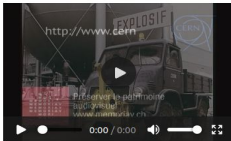
CERN Digitized Videos (Public)

1960->1970  Search Tips  
Advanced Search

Search collections:  
CERN Digitized Videos (Public) \*\*\* add another collection \*\*\*

Sort by:  desc. -or rank by- 10 results split by collection Output format: HTML brief

CERN Digitized Videos (Public)

1.  **Les fondations de la BEBC**  
Les Fondations de la BEBC  
/ @ CERN  
Keywords: BEBC, Construction, Foundation, Drilling, Shaft Date: 1970
2.  **Project CERN (Dexion)**  
Project CERN: a dexion concept for the seventies  
/ 1970  
Language: English, English  
Date: 1970
3.  **Construction ISR 2**  
Intersecting Storage Rings Civil Engineering  
/ 1968-04 / 1968  
Keywords: ISR, Unique, ISR, Meyrin Language: French  
Date: 1968
4.  **ISR Archives**  
/ 17 January 1969 / 1969  
Keywords: ISR, Unique, ISR, Meyrin Date: 1969
5.  **In the heart of CERN in 1967**  
Keywords: History Language: English  
Date: 1967

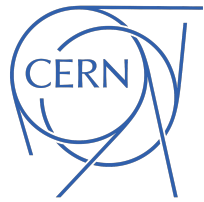
Legacy multimedia (60TB)

Old Web Lectures (57 TB)

IT DHO Archive between 1986 and 2021 (40 distinct file formats)

Creation of Preserved Bags (AIP)

- Uniform metadata, including classification and embargos
- Normalisation to long-term formats
- Preservation checksums allocated
- Linked Data
- Registry to InvenioRDM
- Preserved bags to CTA



# Conclusion

“Ignoring the problems raised by preserving information in digital forms would lead inevitably to the loss of this information” Consultative Committee for Space Data Systems

EU Commission investing in the Preservation technology

Archiver.eu partners with diverse interests

CERN use cases focus on preservation services rather than data storage

Digital Memory use case turned out to have a potential follow-up

“CERN is not just another laboratory. It is an institution that has been entrusted with a noble mission which it must fulfil not just for tomorrow but for the eternal history of human thought” Albert Picot, 3d session of the CERN Council, 10 june 1965