

CERN's Digital Memory

Tim Smith on behalf of:

Digital Memory Project Leader: Jean-Yves le Meur

CDS Team: Jose Benito Gonzalez Lopez, Nico Tarocco

An Eternal Archive

- Albert Picot, 3rd Session of CERN Council, Geneva, 10 June 1955:
 - *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.***
- Archiving Policy at CERN: [↪CDS](#)
 - §4: The archives of a scientific organization comprise any records that **document the results of scientific endeavour and the process of scientific work**, seconded by the necessary administrative and technical support.
 - §10: Pending the adoption of specific guidelines for preserving electronic documents, the archiving of a hard copy of important documents is essential and must be made compulsory in order to ensure that all such documents are archived. **Nevertheless, their electronic form should not be destroyed.**
 - §12: As a place of research, the Organization endeavours, through the CERN Archives, to **preserve the memory of its activities and of its important scientific and cultural role.**

Administrative Basis

- OC3: Rules applicable to archival material and archiving at CERN ([↔CDS](#))
 - §6: Special archival collections at CERN include:
 - Audio and video recordings
 - Photographs and films
 - Press articles
 - Interviews
 - This type of archival material is, at present, the **responsibility of the units concerned** and is subject to special agreements between the originators or the persons responsible for these units and the CERN Archivist
- OC6: CERN Scientific Documents ([↔CDS](#))
 - §16.1: Wherever possible, **each document should be made available in electronic form** (including any figures and diagrams) on a publicly-available server

Photos

5,325 records found 1 - 100 ▶▶ jump to record: 1



Open ["https://cds.cern.ch/record/2644258"](https://cds.cern.ch/record/2644258) in a new tab

CDS Videos

Search CDS

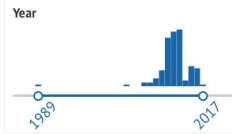


Upload

tim.smith@cern.ch

Found 120 results

Sort: Newest



Category

CERN	115
OPEN	5

Language

English	93
Silent	14
Polish	10
French	2
German	1
Italian	1

Type

VIDEO	95
-------	----



Welcome to the Potato Powered Cosmos



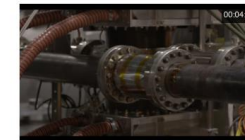
Stock footage ATLAS-ALFA experiment November 2016



Arm 2 of the LHCf detector in the surface barrack at LHC IP1



Stock footage MoEadat experiment September 2016



Stock footage ATLAS-ALFA experiment September 2016



Stock footage Totem experiment in the LHC tunnel



CERN experiment points to a cloudier pre-industrial climate [...]



CERN experiment points to a cloudier pre-industrial climate [...]



Cloud Results HD

Council Documents

- Digitized in 2009/2010

- C, CC (7000)
- FC (12,000)
- SPSC (2,500)

The screenshot shows the CERN Document Server interface. At the top, it says 'CERN Accelerating science' and 'Signed in as: tjs (CERN)'. Below that, the page title is 'CERN Document Server' with navigation tabs for 'Search', 'Submit', 'Help', 'Personalize', and 'Administration'. The breadcrumb trail is 'Home > Articles & Preprints > Committee Documents > CERN Council bodies > Council and Committee of Council > Search Results: year:1952'. The main heading is 'Council and Committee of Council'. A search box contains 'year:1952' and a 'Search' button. Below the search box, there are options for 'Search Tips' and 'Advanced Search', and an 'Add to Search' button. The 'Search collections:' section shows 'Council and Committee of Council' selected. The 'Sort by:' section has 'chronological order' selected. The 'Display results:' section has '10 results' selected. The 'Output format:' section has 'HTML brief' selected. A yellow banner indicates '24 records found' and 'Search took 1.55 seconds'. The first result is '1. Agenda - Council of Representatives for the creation of an European Centre for Nuclear Research - United Nations Educational, Scientific and Cultural Organization', with links for 'English: PDF' and 'French: PDF'.

- Restrictions (embargoes):

- Automatic for new documents
- Archived have between 5 and 30 year embargoes, and automated release

AUDIO

5'183 cassettes



3'648 tapes



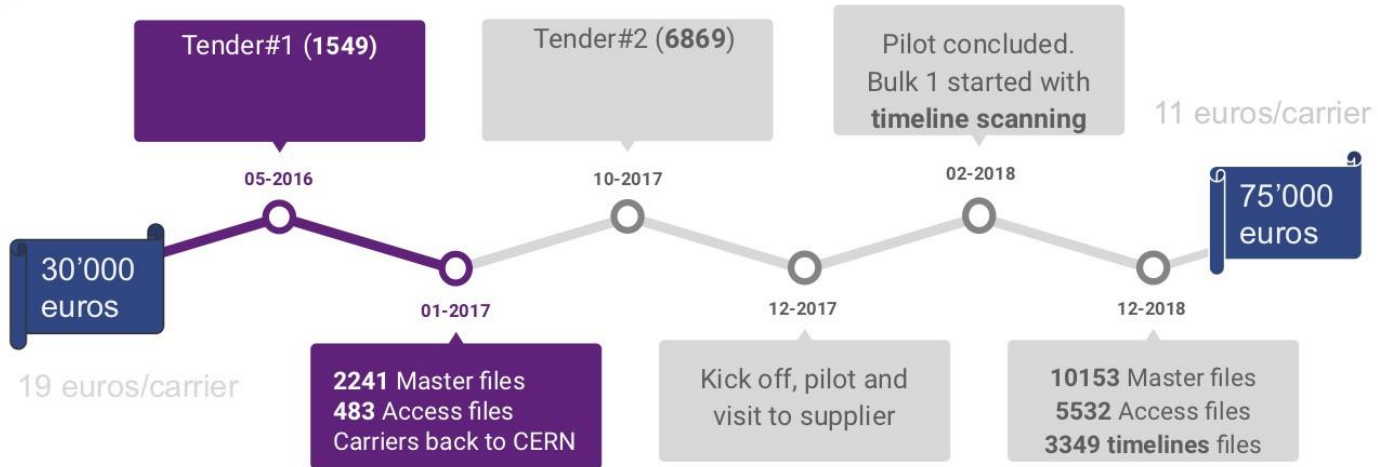
1'740 Kg



Digital Memory

AUDIO

Sponsored by CERN Council



In collaboration with the CERN Archive, Purchasing and Translation Services

AUDIO



Ingested ~30'500 files into 3'000 records -
with 3'300 scanned timelines

- The digitally-born audio recordings ?
(post-2006)
- Run speech to text ?
- Open the access
 - → in ~100 years ?



Surprise!

Treated: **5112** cassettes, **3289** tapes



# of tapes requiring cleaning	28
# of sticky tapes requiring baking	80
# of 'no signal' (bande vierge)	677
# with only 1 face recorded	688
# of duplicates	125
# of degraded audio signal	95
# of failures (sticky, blocked, blank)	68
# of failed continuity between tracks	1111

VIDEO

- Documentaries and Seminars recordings
- Complex initial situation:
 - Material scattered everywhere on CERN site
 - Very partial catalog, disconnected from the support material
- Inventory in collaboration with the Library & VideoLab
 - About 5'000 videotapes on 13 different carriers: from D3 to HDCAMs or 1 inch C tapes!
 - About 120 films
 - Systematic QR-coding of "everything"
 - Carrier analysis to either update existing records or create new ones
- Bidding: 121'000 euros (~23.7 €/carrier)
 - INA & Memoriam expertise

CERN Document Server

Video Inventory

1,847 records found

Quantum chromodynamics: 1994 ed. / Martiniello, Guido
L. Hultonen on bound states of quarks and gluons. 2. Quark models, confinement and asymptotic freedom. [...] Geneva: CERNA, 1994 - Transparencies : 2 DVD video : 2 VHS video. CERNA Academic Training Lecture. 302 Fulltext: PDF; Transparencies, pt.1: T1; Transparencies, pt.2: T2;
IN - Academic Training Lectures, CERNA, Geneva, Switzerland, 1 Sep 1994 - 30 Jun 1995 - CERNA library copies

Beyond the Standard Model: 1994 ed. / Dimopoulos, Spyros K
The LEP measurement of the weak mixing angle provides us with the only precise experimental indication for physics beyond the Standard Model. It points in the direction of unified theories with supersymmetric partners accessible to LHC. [...] Geneva: CERNA, 1994 - Transparencies : 1 DVD video : 3 VHS video. CERNA Academic Training Lecture. 302 Fulltext: PDF; Transparencies: T1;
IN - Academic Training Lectures, CERNA, Geneva, Switzerland, 1 Sep 1994 - 30 Jun 1995 - CERNA library copies

```
<record>
<controlfield tag="001">2203501</controlfield>
<controlfield tag="005">20171121121852.0</controlfield>
<datafield tag="245" ind1="1" ind2="1">
<subfield code="a">
6th International Conference on Public Communication of Science
</subfield>
</datafield>
<datafield tag="690" ind1="C" ind2="1">
<subfield code="a">movingImages</subfield>
</datafield>
<datafield tag="852" ind1="1" ind2="1">
<subfield code="8"></subfield>
<subfield code="a">68</subfield>
<subfield code="b">MB500</subfield>
<subfield code="c">CN-A0000334</subfield>
</datafield>
<datafield tag="852" ind1="1" ind2="1">
<subfield code="8"></subfield>
<subfield code="a">68</subfield>
<subfield code="b">MB500</subfield>
<subfield code="c">CN-A0000334</subfield>
</datafield>
<datafield tag="852" ind1="1" ind2="1">
<subfield code="8"></subfield>
<subfield code="a">68</subfield>
<subfield code="b">MB500</subfield>
<subfield code="c">CN-A0000334</subfield>
</datafield>
<datafield tag="852" ind1="1" ind2="1">
<subfield code="8"></subfield>
<subfield code="a">68</subfield>
<subfield code="b">MB500</subfield>
<subfield code="c">CN-A0000334</subfield>
</datafield>
```



VIDEO

- 3 main sets delivered
 - public/restricted/copyrighted
- In-depth study of speech to text transcription software
 - Mitigated conclusions so far
- QA run by Contractor and CERN



Surprise!

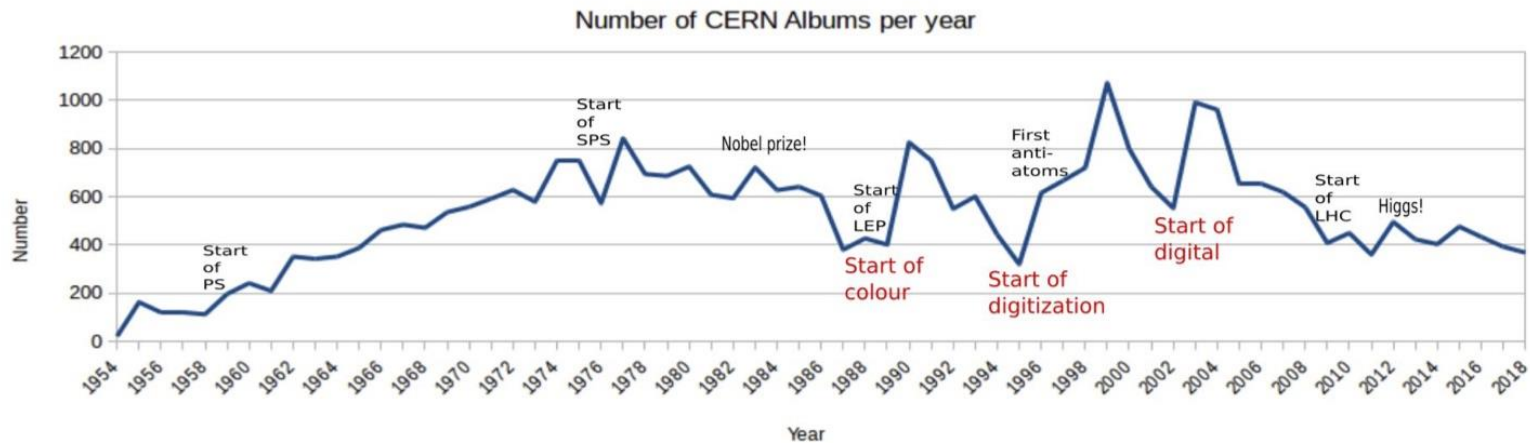
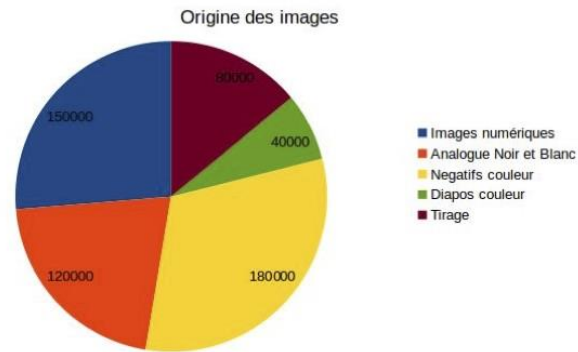
Originals	Preservation Master	Access Master	Access Copy
35mm films	Wrapper:	.mkv	.mov
	Video codec:	FFV1 - 10 bits RGB	Apple ProRes 422 HQ
	Audio codec:	24 bits PCM, 48kHz	24 bits PCM, 48kHz
	Definition / Aspect ratio:	4096x? / Original	4096x? / Original
16mm films	Wrapper:	.mkv	.mov
	Video codec:	FFV1 - 10 bits RGB	Apple ProRes 422 HQ
	Audio codec:	24 bits PCM, 48kHz	24 bits PCM, 48kHz
	Definition / Aspect ratio:	2048x? / Original	2048x? / Original
Analogue and digital SD video	Wrapper:	.mkv	.mov
	Video codec:	FFV1 - 10 bits YCbCr	Apple ProRes 422 LT - SD profile
	Audio codec:	24 bits PCM, 48kHz	24 bits PCM, 48kHz
	Definition / Aspect ratio:	7x576 / Original	7x576 / Original
All		Maintain the original recording standard, interlacing, number of audio channels and auxiliary information such as original timecode	Maintain the original recording standard, interlacing, number of audio channels and auxiliary information such as original timecode
			Progressive scan at 25 frames per second. Maintain the original number of audio channels and auxiliary information such as original timecode

	A	B	C	D	E	F	G	H	I
	Fichier	URL	Titre inventorié	Année inventoriée	Remarks at digitization time	Validité du titre ?	Validité du son ?	Validité de l'image ?	Plusieurs vidéos / support ?
1									
2	ac_C1V1_423111.mp4	http://cds.cern.ch/record/423111	LHC Committee (LHCC) held on 5-6 November 1992 - ATLAS Collaboration.	1992.	Low RF, noisy picture Videotape dubbing error	TITRE CORRESPOND A VIDEO	SON OK	IMAGES OK	NON
3	ac_C1V1_1047966-1.mp4	http://cds.cern.ch/record/1047966	At the edge of Glasgow's desert.	1992.	Low RF, noisy picture	TITRE CORRESPOND A VIDEO	SON OK	IMAGES OK	NON
4	ac_C1V1_2196947.mp4	http://cds.cern.ch/record/2196947	Element Genesis. Solving the mystery. Traces de collisions UA-1 et/ou UA-2 - Filmées sur télescopes.	2001.	Low RF, noisy picture Video tracking issue Tape not in PAL format	TITRE CORRESPOND A VIDEO	SON OK	AUTRE DEFAULT	NON
5	ac_C1V1_2213869.mp4	http://cds.cern.ch/record/2213869	OPAL Safety.	CERN.	Low RF, noisy picture Video tracking issue	TITRE CORRESPOND A VIDEO	SON OK	IMAGES OK	NON
6	ac_C1V1_2258188.mp4	http://cds.cern.ch/record/2258188	Edoardo Amaldi: Miscellaneous videos.	1952	Low RF, noisy picture Video tracking issue	TITRE CORRESPOND MAIS TRONQUE	SON OK	IMAGES OK	NON
7	ac_C1V1_2259523.mp4	http://cds.cern.ch/record/2259523	CHORUS J. Dupont INFN (Naples).	Geneva - CERN - 1993 (7).	Video tracking issue	TITRE CORRESPOND A VIDEO	SON OK	AUTRE DEFAULT	NON
8	ac_C1V1_2266633.mp4	http://cds.cern.ch/record/2266633	Foraky (Société Anonyme de forage et de forçage Belgique).	Geneva - CERN - 1984.	Video tracking issue Low RF, noisy picture Videotape dubbing error	TITRE CORRESPOND A VIDEO	SON OK	IMAGES OK	NON
9	ac_C1V1_2270605.mp4	http://cds.cern.ch/record/2270605							

PHOTO

34'000 albums since 1954

Licenses: CERN or CC-BY-4.0



PHOTO

- Black & White serie scanned in 2014
- PhotoLab store organized in albums
- Approximate count of:
 - 24x36 negatives: 178'618
 - 24x36 slides: 38'844
 - Medium & Large formats: 81'803
- Keys:
 - TIFF 48 bits (RGB) & 4800 ppi (for 24x36 size images)
 - File naming pattern to enable album identification
- Bidding: 62'000 € for slides/ML & 47'000 € for negatives (~0.35 €/image)



PhotoLab Archives






Search 19,066 records for:

[Search Tips](#) [Advanced Search](#)

search also PhotoLab Archives Images [Add to Search](#)

PhotoLab Archives images

Please note: the Photolab Archive images are currently being scanned, uploaded and automatically matched with old album records. matching is being inspected manually to check for errors. You can use the links below to see various sets of images:





[Surprise!](#)

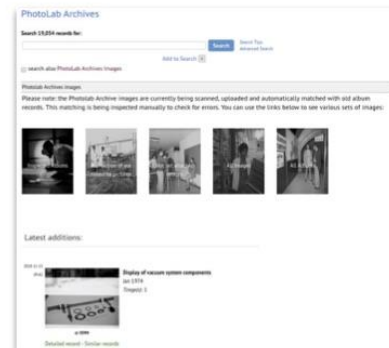
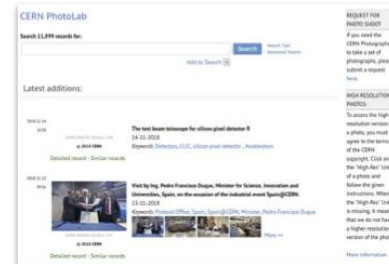
The Metadata Challenge

- Record representation ?
 - La flexibilité de MARC21
- Best structure ?
 - Each media with its own record and file
 - Each album with a record and media files
- Trials to enrich metadata
 - From the source: validate and extract information from the shooting request
 - Back to the tree: CERN retirees from Alumni network can caption all the images !
 - With the help of technologies: reconnaissance faciale/object recognition, tagging and propagation by similarity, crowd-sourcing ?

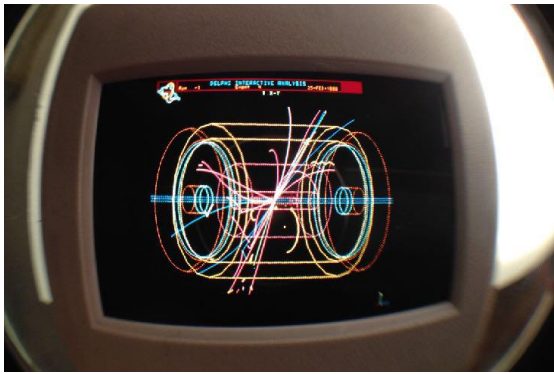
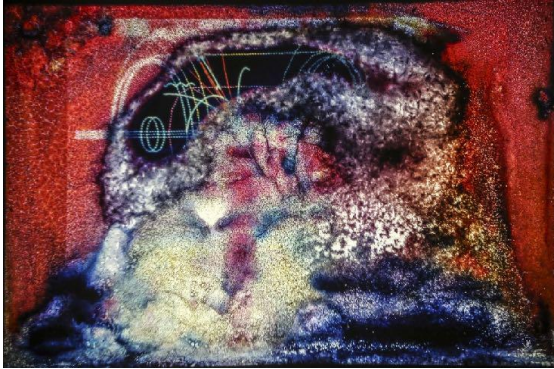


The analog/digital Merging Challenge

- Digital media VS Digitized media
- Which Master format for préservation ?
- Which structure for records ?
- Which interfaces should be shared ?
- The endless use cases !
 - Duplicata: some analog negatives scanned randomly in the 90's
 - Duplicata: albums with same images in B&W and colours
 - Consistency: connection between images & albums either missing or wrong
 - Consistency: do we apply GDPR on 1950's images
 - Support and preserve new exotic formats (ex: panoramic photos and co) ?



Art Discovered



<http://cern.ch/volmeur>

Questions ?

