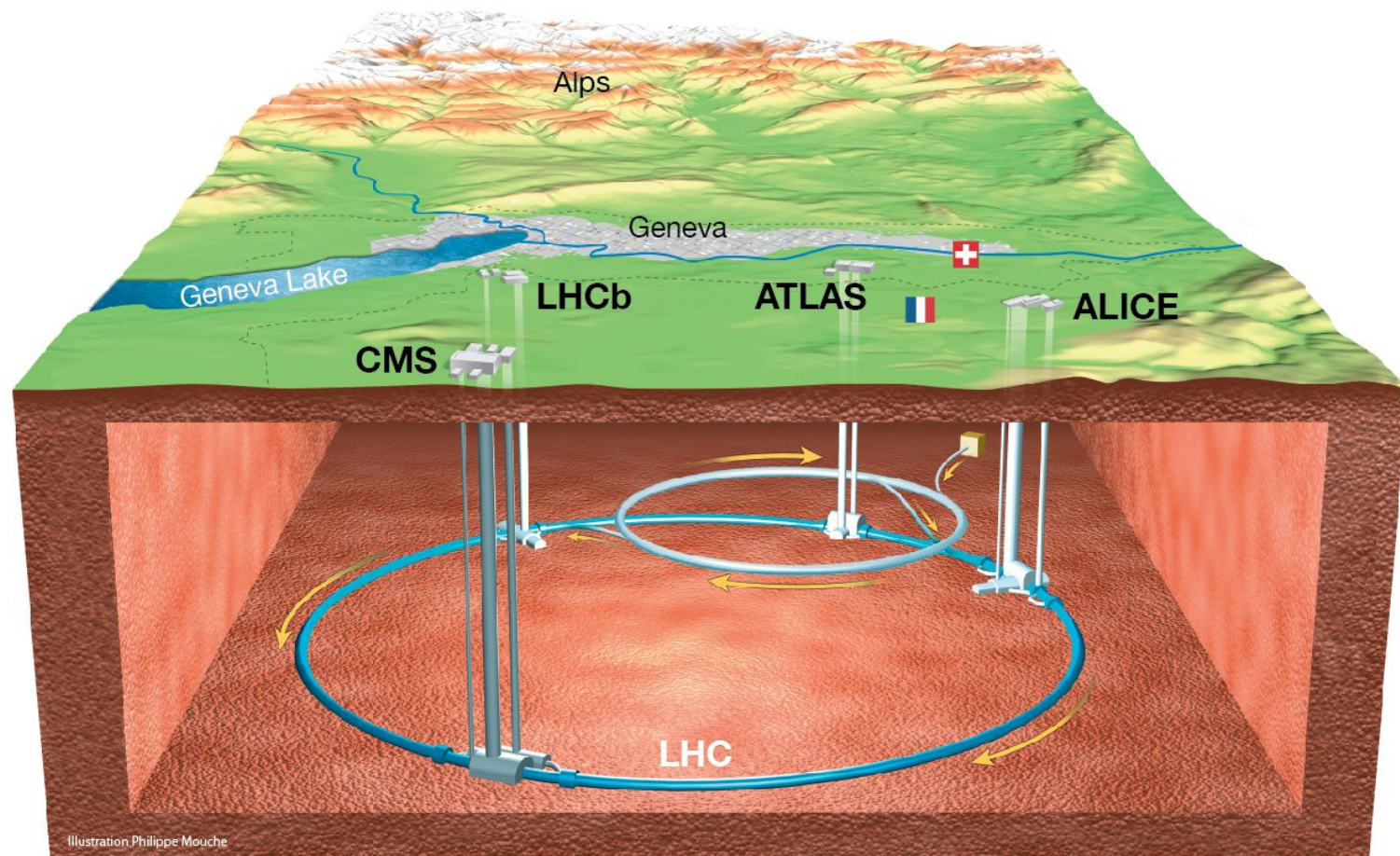
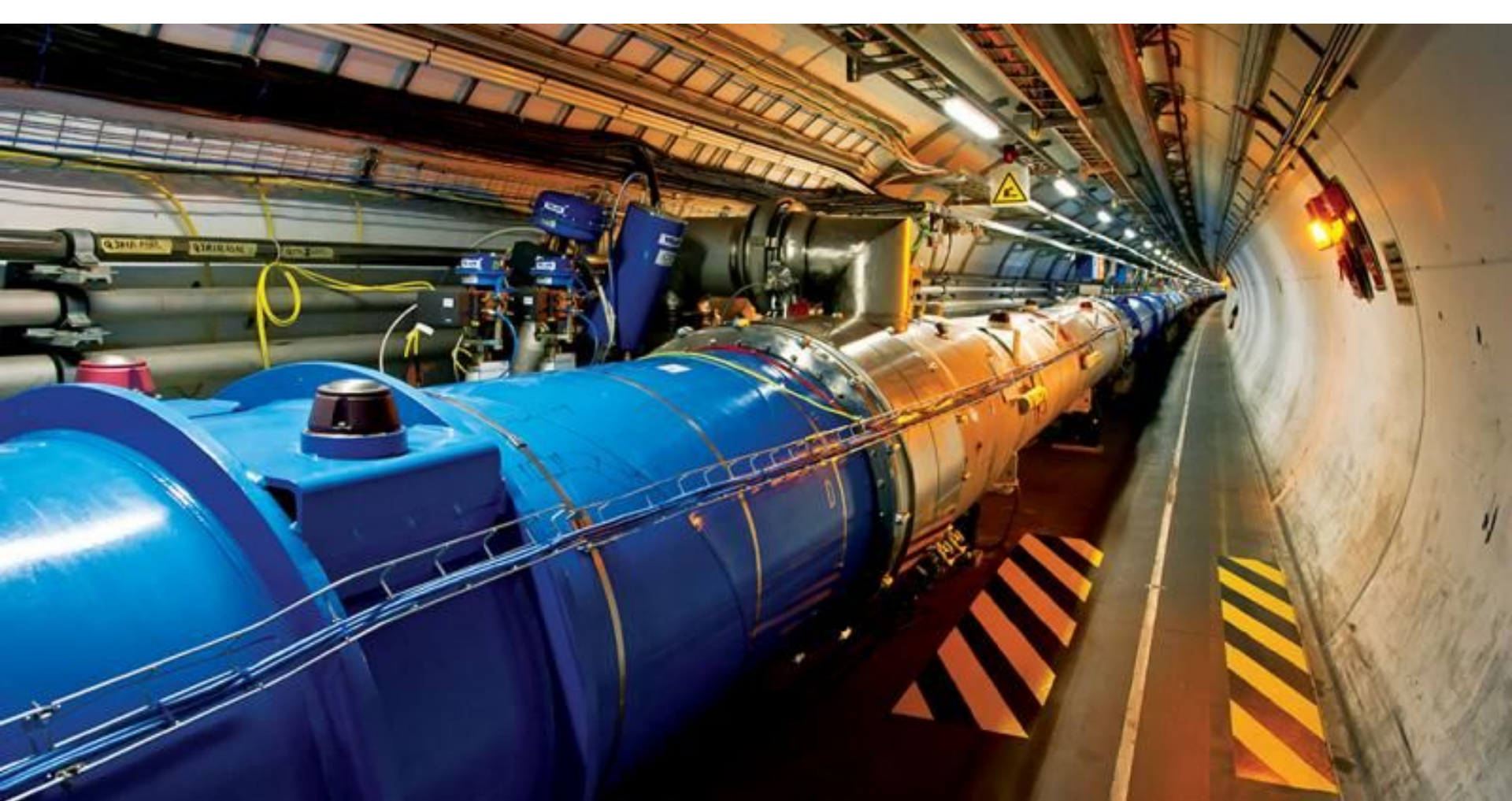


# When Patrimony Data meets Scientific Data

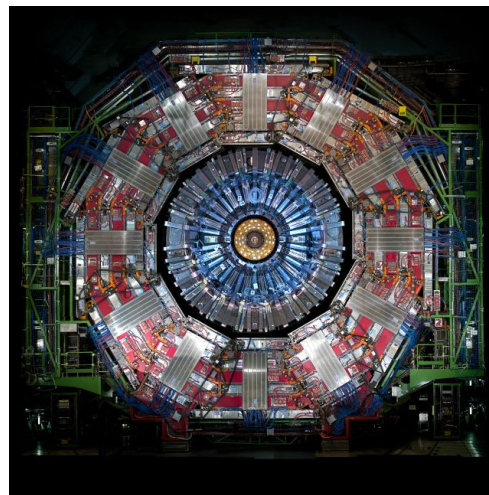
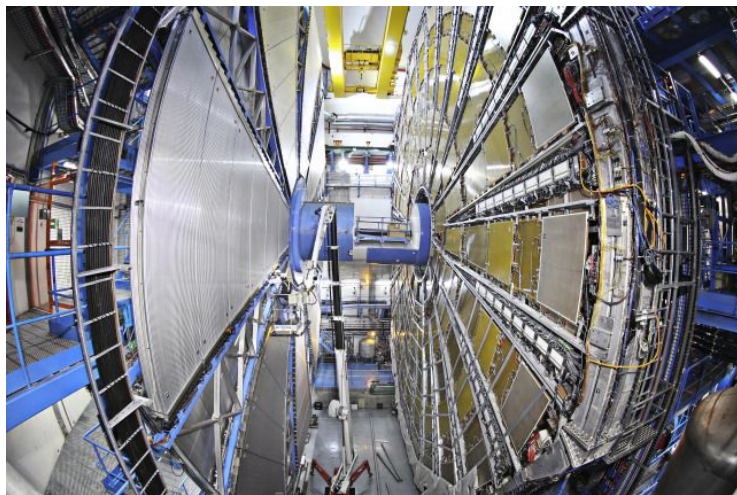
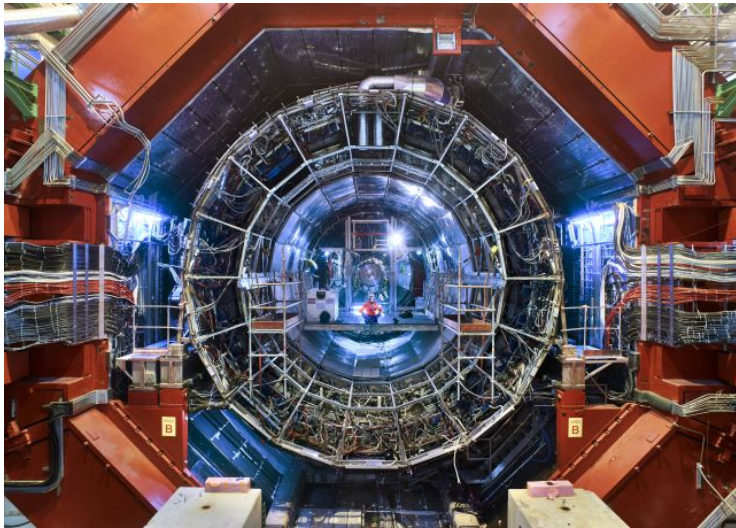
CERN's Digital  
Memory

JY Le Meur



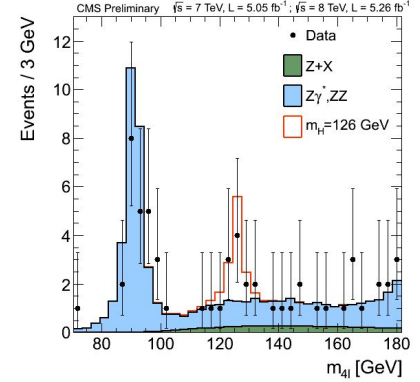
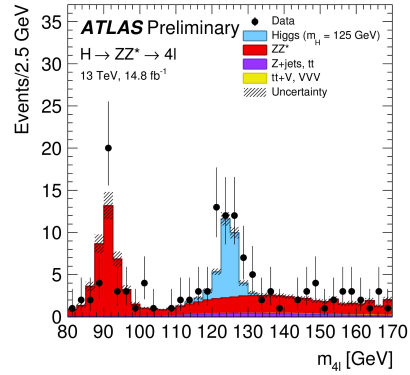
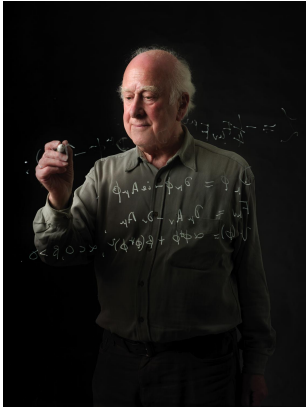




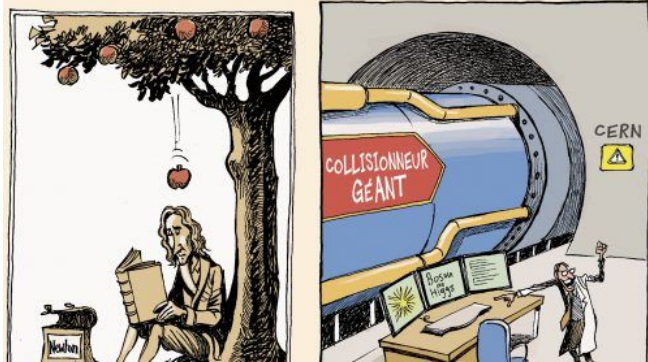


[Live stats](#)





## HIGGS BOSON



@Chapatte 2011



## WWW



## The World Wide Web

The WorldWideWeb (W3) is the universe of network-accessible information, an embodiment of human knowledge. It is an initiative started at [CERN](#), now with many participants. It has a body of software, and a set of protocols and conventions. W3 uses hypertext and multimedia techniques to make the web easy for anyone to roam, browse, and contribute to.

The [W3 Consortium](#) now ensures the continued interoperability which is W3 though its rapid evolution. This is run by [MIT](#) in the context of a [joint initiative](#) involving CERN and MIT. Everything there is to know about W3 is linked directly or indirectly to this document.

### What's out there?

Pointers to the world's online information, [subjects](#), [W3 servers](#), etc.

### WWW Software Products

What there is and how to get it: [clients](#), [servers](#), [gateways](#), [libwww](#) and [tools](#).

Discussion

[News groups](#), [WWW mail addresses](#), [how to contact the WWW Team](#), [W3 interactive talk](#)

Conferences

[first](#), [second](#), and [third](#) WWW conferences.

[Bibliography](#)

Paper documentation on W3 and references. Also: [manuals](#).

[Project background](#)

[Summary](#), [illustrated talk](#), [People](#), [History](#), why the [W's are green](#)

[How can I help?](#)

If you would like to support or add to the web..

[Getting CERN code](#)

Getting the code by [anonymous FTP](#), etc.

[Frequently Asked Questions](#)

If you can't find it above or below ([old small list](#)).

## Technical details

[How to provide data](#)

How can I make my own data available on the web?

[Developments](#)

Ideas for new projects, things people are doing.

### Specs, Tutorials, etc:

[HTTP](#)

All available information on Hypertext Transfer Protocol (HTTP) and related protocols.

[HTML format](#)

The markup language used for some documents and for search hit-lists.

[Graphics](#)

Pointers to information, specs, FAQ etc.

[Access Authorization](#)

Document protection in W3.

[Common Gateway Interface](#)

The interface between httpd servers (CERN's and NCSA's at least) and user-written gateways to other database systems, etc.

[Addressing \(URLs\)](#)

The syntax of W3 document addresses.

[WWW'94 Conference Proceedings](#)

A searchable index of all the papers from the Conference are now available on the Web

### Discussion:

[Design Issues](#)

Discussions of decisions to be made when designing or selecting a hypertext/IR system. See also [related products](#).

[Working notes](#)

Work in progress at the drawing board, notes of meetings etc. Additions welcome: send URLs

[News](#)

Some internet/usenet newsgroups of possible interest to the WorldWideWeb project.

### Other:

[Coding standards](#)

A basic style guide for W3 code contributors. If you write code, read this!

[Using CVS in WWW](#)

A guide to setting up the code management system for for those hacking the code.

[Test data](#)

A collection of data for testing

[Tim Bl](#)



## First International Conference on the World-Wide Web

**May 25-26-27 1994, CERN, Geneva (Switzerland)**

The Conference is over... This page remains of interest as a report and as pointer to the next ones.

The Third one will be:

[WWW Spring '95](#)

Darstadt, Germany, April 10-14 1995, organised by the

[Fraunhofer Gesellschaft](#).

The second conference was held in the US:

[WWW Fall '94: Mosaic and the Web](#)

Chicago, October 17-19, organised by [NCSA](#) and CERN.

The three days in Geneva were attended by 380 [participants](#) from all over the world.

My thanks go to many [people](#), and also to many [institutes and companies](#), especially to the co-organisers from the [CUJ](#) (Centre Universitaire d'Informatique), Oscar Nierstrasz and Bertrand Ibrahim.

There were 49 formal presentations, 11 [workshops](#) and many discussions going on in parallel. The [proceedings](#) will be published in two forms, and until then, the [preliminary proceedings](#) are available.

The Best of WWW [Awards](#) were handed out at the Conference Dinner.

The conference closed on Friday afternoon with a general [panel](#) discussing issues for the future.

You can submit [trip reports](#) (but I do not guarantee publication!)

### User Society

Furthermore, we are considering creating a WWW User Society, which will serve the purpose of grouping the individual users, institutes and existing interest groups. It will run the conferences and suggestions on purpose and structure, as well as offers of help with this Society, by e-mail to [www-usersoc@www.cern.ch](mailto:www-usersoc@www.cern.ch).

### Past History...

You can still access the [old information](#) temporarily.

[RC 02 June 94](#)



# FileSystem Phase-Out in 2018

→ 39'000 files dated between 1990 and 1997

"WWW", "Administration", "hypertext", "History", "ftp",  
"www0", "People", etc.

## Donation Option

All files of the AFS w3org filesystem between 1990 and 1997

```
https://www.cbl.lu.se/users/af30000/
[Oct 7 1994] AccessAuthorization
[Dec 12 1993] AccessControllist.html
[Dec 12 1993] AriL.html
[Sep 7 1993] ATTENTION_DO_NOT_EDIT_THESE_FILES
[Dec 12 1993] BasicDiscussion.html
[Dec 12 1993] Basic.html
[Dec 12 1993] Browser.html
[Dec 12 1993] CERNServer.html
[Dec 12 1993] CERNServerNutsShell.html
[Dec 12 1993] FingerHole.html
[Dec 12 1993] GroupFile.html
[Oct 7 1994] icons
[Oct 7 1994] 32x32
[Oct 8 1993] bomb.gif
[Oct 8 1993] bsd_daemon.gif
[Oct 8 1993] calendar.gif
[Oct 8 1993] caution.gif
[Oct 8 1993] dir.gif
[Oct 10 1993] escherknot.gif
[Oct 8 1993] file.gif
[Oct 8 1993] finland.gif
[Oct 8 1993] flagup.gif
[Oct 8 1993] globe.gif
[Oct 8 1993] hearts.gif
[Oct 8 1993] help.gif
[Oct 8 1993] home.gif
[Oct 8 1993] index.gif
[Oct 8 1993] info.gif
[Oct 8 1993] match.gif
[Oct 8 1993] music.gif
[Oct 8 1993] paint.gif
[Oct 8 1993] phone.gif
[Oct 8 1993] question.gif
[Oct 8 1993] RIP.gif
[Oct 8 1993] smiley.gif
[Oct 10 1993] triangle.gif
[Oct 8 1993] warning.gif
[Oct 8 1993] x.gif
[Oct 7 1994] 48x48
[Oct 8 1993] bomb.gif
[Oct 8 1993] coins.gif
[Oct 8 1993] daemon.gif
[Oct 10 1993] escherknot.gif
```

```
[May 2 1991] Installation.html
[May 2 1991] Installation.html
[May 2 1991] InstallationOS9.html
[May 6 1991] InstallationUnix.html
[May 2 1991] InstallationVMS.html
[Oct 29 1991] Internet.html
[May 2 1991] M6809.html
[Dec 19 1990] M68kStandAlone.html
[Dec 19 1990] Macintosh.html
[May 2 1991] MoreInformation.html
[May 6 1991] MSDOS.html
[Dec 1 1992] OS9.html
[Dec 1 1992] OS9.html
[Mar 20 1991] .places
[May 24 1991] Platforms.html
[Dec 19 1990] remains.sgml
[Oct 25 1991] Routines.html
[May 2 1991] RPCL.html
[Dec 19 1990] rpcuser.sgml
[Dec 19 1990] RS232.html
[Jan 31 1995] RunTimeErrors
[Mar 11 1991] AccessViolation.txt
[Mar 11 1991] BadRPCAddress.txt
[Mar 11 1991] DeviceInactive.txt
[Mar 11 1991] ExceededQuota_2.txt
[Mar 11 1991] ExceededQuota.txt
[Mar 11 1991] InsufficientPrivilege_2.txt
[Mar 11 1991] InsufficientPrivilege.txt
[Mar 11 1991] NetworkPartnerExited.txt
[Mar 11 1991] NoIOChannelsAvailable.txt
[Mar 11 1991] NoSuchDevice.txt
[Mar 11 1991] OutputBufferFull.txt
[Mar 11 1991] PathLost.txt
[Mar 11 1991] .places
[Dec 19 1990] SystemTypes.html
[Mar 4 1991] Types.html
```

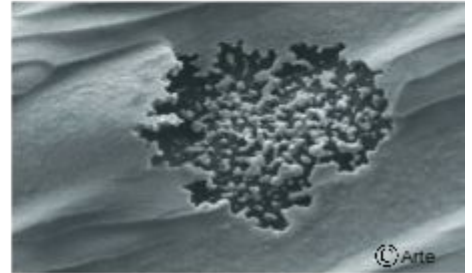
*"CERN is not just another laboratory. It is an institution that has been entrusted with a noble mission which it must fulfil not just for tomorrow but for the eternal history of human thought."*  
(Albert Picot, 3rd Session of CERN Council, Geneva, 10 June 1955)

**1955**



- Bit rot and redundancy failure
- Obsolescence of readers, formats, OS, HWs
- Lost in transitions
- Missing context
- Economic failure
- Corruption, mistake or attack
- Dissipation

...





**ORGANISATION EUROPÉENNE POUR LA RECHERCHE NUCLÉAIRE  
EUROPEAN ORGANIZATION FOR NUCLEAR RESEARCH**

Laboratoire Européen pour la Physique des Particules  
European Laboratory for Particle Physics

**OPERATIONAL CIRCULAR N° 3**

**Issued by Personnel Division**

This operational circular was examined by the Standing Concertation Committee at its meeting of 25 June 1997.

<i>Applicable to :</i>	Members of the Personnel External users of CERN Archives
------------------------	---

<i>Person responsible for the matter concerned</i>	: Director-General
--	--------------------

<i>Date</i>	: October 1997
-------------	----------------



**RULES APPLICABLE TO ARCHIVAL MATERIAL  
AND ARCHIVING AT CERN**

**1979**

**PREAMBLE**

CERN is responsible for its documents and files and their preservation. It must formulate its own archiving policy and rules, taking into account legal obligations deriving from ordinary law. The aim of this document is to set out rules and procedures for archiving and access to archival material at CERN.



# ARCHIVING POLICY AT CERN

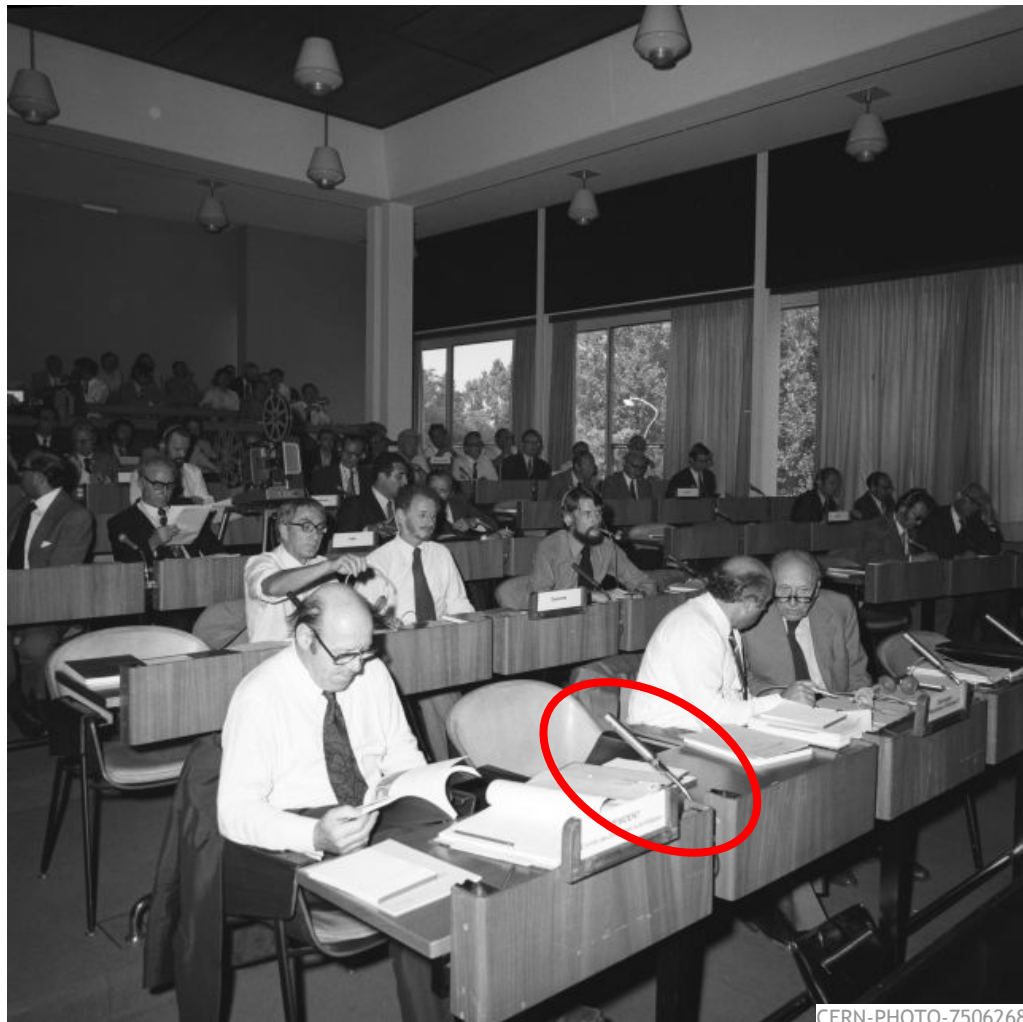
## PREAMBLE

Science does not exist in a social vacuum; therefore cultural, economic and political developments also find their expression in the records of scientific activity.

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.
11. There exist collections of files to which special archiving obligations apply.

**1997**

# **Digitizing the “Big Data” of the 20th Century**





# AUDIO

## 5'183 cassettes



## 3'648 tapes



## 1'740 Kg

**CERN Document Server**

Search Submit Help Personalize

Home > Archives > CERN Archives > CERN Council and its committees > Audio recordings of meetings (Archives)

### Audio recordings of meetings (Archives)

Search 2,387 records for:

Search Add to Search

Latest additions:

- 2018-11-16 **CERN-AUDIO-ARCH-1992-2002**  
15:37 **Audio Archive** : Audio Archive - 1992-2002  
CERN, Geneva, Audio. From 1992-01-01 to 2002-12-31  
Detailed record - Similar records
- 2018-11-16 **CERN-AUDIO-ARCH-1982-1991**  
14:16

About audio recordings  
Sound recordings of the meetings of the CERN Council are made for the sole purpose of assisting the official minute-writers in their work. Once approved by the Council, the minutes constitute the sole formal record of its meetings. The CERN Council's audio archive is therefore not publicly accessible.

# 15 TB

504 records found ◀◀ 498 - 504 jump to record: 498

- Conférence sur la Recherche de l'espace (publique)** / Prof Nassi(?)(en présence de M Adan  
CERN-AUDIO1960-001.- . - 83' English © 1964 CERN  
▶
- More...  
Detailed record - Similar records
- 499. **Les Grands Accélérateurs** / R Gourian  
CERN-AUDIO1964-006.- . - 60' French © 1964 CERN  
▶  
Detailed record - Similar records
- 500. **Allocation du D.G.- 10me anniversaire** / Prof Weisskopf et ?  
CERN-AUDIO1964-005.- 1964. - 23' eng/fre © 1962 CERN  
▶  
Detailed record - Similar records
- 501. **Conférence de Presse** / Butler  
CERN-AUDIO1964-004.- 1964. - 16' English © 1960 CERN  
▶  
Detailed record - Similar records



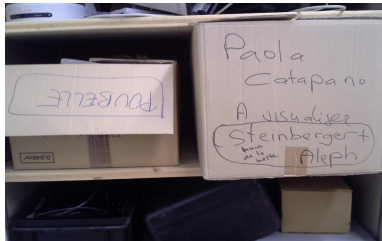
Ingested ~30'500 files (master and access) into 3'000 records - with 3'300 scanned timelines

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

# 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

# The initial situation, offices and corridor storage





**VIDEO**

	Bat-510	Bat-500	Bat-60	Bat-698	Total
<b>35 mm</b>	0	0	0	4	4
<b>16 mm</b>	61	51	0	0	112
<b>8 mm</b>	0	0	0	5	5
<b>D3</b>	0	40	0	22	62
<b>Digital Betacam</b>	0	5	0	0	5
<b>DVCAM</b>	45	9	25	367	446
<b>DV/miniDV</b>	9	0	0	142	151
<b>DVD</b>	56	440	0	180	676
<b>Betacam SP</b>	188	385	600	195	1368
<b>Betacam</b>	23	0	0	10	33
<b>1 inch C</b>	0	0	20	1	21
<b>U-matic S/SP</b>	273	120	565	108	1066
<b>U-matic</b>	128	0	0	71	199
<b>VHS</b>	189	388	180	154	911
<b>Others</b>	27	0	1	16	44
<b>Total</b>	999	1438	1391	1275	5109

# Numbers



# Destination file formats

Originals		Preservation Master	Access Master	Access Copy
35mm films	Wrapper:	.mkv	.mov	.mp4
	Video codec:	FFV1 - 10 bits RGB	Apple ProRes 422 LT	H.264 @ 5Mbps
	Audio codec:	24 bits PCM, 48kHz	24 bits PCM, 48kHz	16 bits AAC, 44.1kHz, 256kbps
	Definition / Aspect ratio:	4096x? / Original	1920x1080 / Pillar-letterbox	1920x1080 / Pillar-letterbox
16mm films	Wrapper:	.mkv	.mov	.mp4
	Video codec:	FFV1 - 10 bits RGB	Apple ProRes 422 LT	H.264 @ 5Mbps
	Audio codec:	24 bits PCM, 48kHz	24 bits PCM, 48kHz	16 bits AAC, 44.1kHz, 256kbps
	Definition / Aspect ratio:	2048x? / Original	1920x1080 / Pillar-letterbox	1920x1080 / Pillar-letterbox
Analogue and digital SD video	Wrapper:	.mkv	.mov	.mp4
	Video codec:	FFV1 - 10 bits YCbCr	Apple ProRes 422 LT - SD profile	H.264 @ 1Mbps
	Audio codec:	24 bits PCM, 48kHz	24 bits PCM, 48kHz	16 bits AAC, 44.1kHz, 256kbps
	Definition / Aspect ratio:	?x576 / Original	?x576 / Original	640x360 / Pillar-letterbox



- Documentaries and Seminars recordings, and all footages
- Inventory in collaboration with the Library & VideoLab
  - Systematic QR-coding of “everything”
  - Carrier analysis to either update existing records or create new ones
- QC by both contractor and CERN: with OpenRefine app
- Bidding: 121'000 euros (~23.7 €/carrier)
  - INA & Memoriv expertise



CERN Document Server

Home > CERN Departments > Information Technology (IT) > CDS Documents > CDS Internals > Moving Images > Video Inventory > Search Results: cern

Search Submit Help Personalize

cern Search Search Tips Advanced Search

Add to Search

Search collections: Video Inventory \*\*\* add another collection \*\*\*

Sort by: Display results: Output format: latest first desc. - or rank by: 10 results single list HTML brief

Video Inventory 1,867 records found 1858 - 1867 Jump to record: 1858 Search took 1.25 seconds.

1858. **Quantum chromodynamics ; 1994 ed. / Martinelli, Guido**  
1. Hadrons as bound states of quarks and gluons. 2. Quark models, confinement and asymptotic freedom. [...] Geneva : CERN, 1994 - Transparencies ; 2 DVD video ; 2 VHS video. CERN Academic Training Lecture, 301 **Fulltext:** PDF; **Transparencies, pt.1:** TIF; **Transparencies, pt.2:** TIF;  
In: Academic Training Lectures, CERN, Geneva, Switzerland, 1 Sep 1994 - 30 Jun 1995 - CERN library copies  
Detailed record - Similar records

1859. **Beyond the Standard Model ; 1994 ed. / Dimopoulos, Savas 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 particles accessible to LHC. [...] Geneva : CERN, 1994 - Transparencies ; 1 DVD video ; 1 VHS video. CERN Academic Training Lecture, 302 **Fulltext:** PDF; **Transparencies:** TIF;  
In: Academic Training Lectures, CERN, Geneva, Switzerland, 1 Sep 1994 - 30 Jun 1995 - CERN library copies  
Detailed record - Similar records

```
<record>
  <controlfield tag="001">2293591</controlfield>
  <controlfield tag="005">20171121121652.0</controlfield>
  <datafield tag="245" ind1=" " ind2=" ">
    <subfield code="a">
      6th International Conference on Public Communication of Science
    </subfield>
  </datafield>
  <datafield tag="690" ind1="C" ind2=" ">
    <subfield code="a">movingimages</subfield>
  </datafield>
  <datafield tag="852" ind1=" " ind2=" ">
    <subfield code="8">A</subfield>
    <subfield code="a">60</subfield>
    <subfield code="b">MB500</subfield>
    <subfield code="c">CM-A00000334</subfield>
  </datafield>
  <datafield tag="852" ind1=" " ind2=" ">
    <subfield code="8">B</subfield>
    <subfield code="a">60</subfield>
    <subfield code="b">MB500</subfield>
    <subfield code="c">CM-A00000334</subfield>
  </datafield>
  <datafield tag="852" ind1=" " ind2=" ">
    <subfield code="8">C</subfield>
    <subfield code="a">60</subfield>
    <subfield code="b">MB500</subfield>
    <subfield code="c">CM-A00000334</subfield>
  </datafield>
  <datafield tag="852" ind1=" " ind2=" ">
    <subfield code="8">D</subfield>
    <subfield code="a">60</subfield>
    <subfield code="b">MB500</subfield>
    <subfield code="c">CM-A00000334</subfield>
  </datafield>
</record>
```



VIDEO



Category

OPEN

22

Language

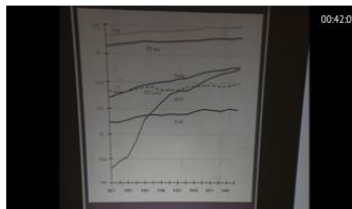
English

22

Type

VIDEO

22



00:42:09

Tim Berners-Lee: The Future of the Web (WWW94)



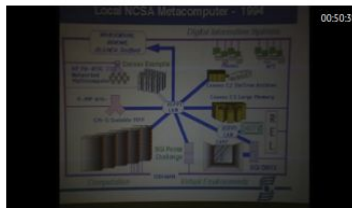
00:51:08

Borre Ludvigsen and Denis Anthony's talks at the WWW94



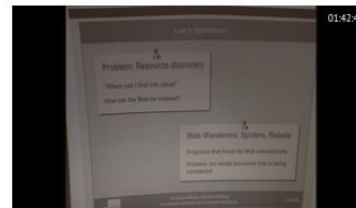
00:50:44

Steffen Meschkat and John Mallery's talks at the WWW94



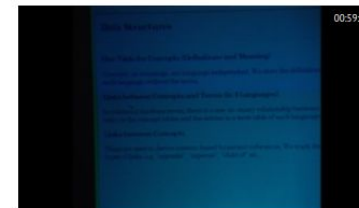
00:50:39

Joseph Hardin's talk on the state of the web (WWW94)



01:42:47

Christian Neuss, Stefanie Höfling, Doug McKee, P.M.E. De Bra [...]



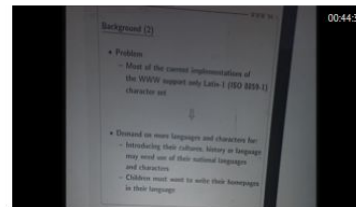
00:59:33

Talks by David Eichmann (et al.) and Carlos A. Varela (et al [...])



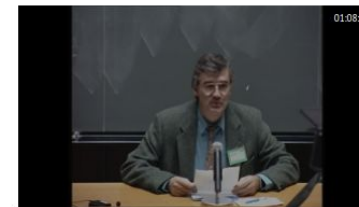
00:52:32

Robert Cailliau: What is WWW and Where is it used now (WWW94 [...])



00:44:32

Talks by Toshihiro Takada, Nick Williams and Tim Wilkinson a [...]



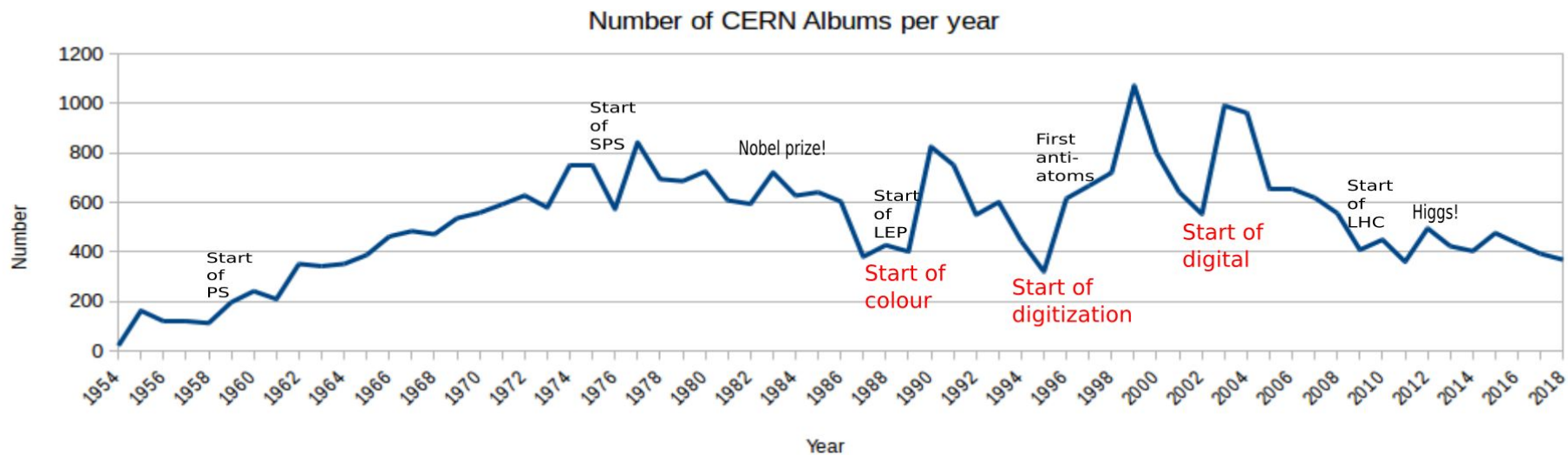
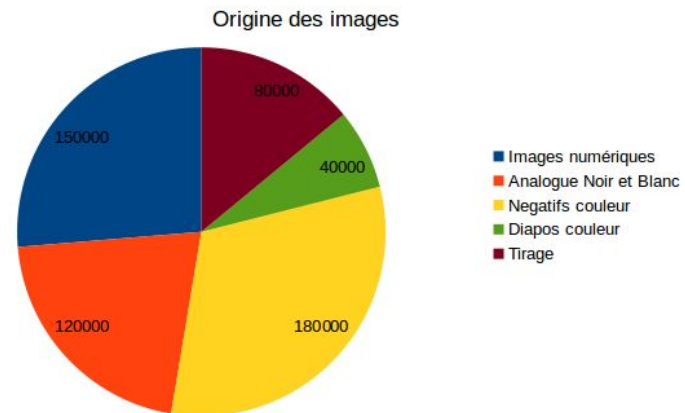
01:08:26

WWW94: Plenary Panel

PHOTO

# 34'000 albums since 1954

Licenses: CERN or  
CC-BY-4.0



PHOTO

- PhotoLab store organized in albums
- B&W: 120'000 treated in 2014
- Approximate count of Colours:
  - 24x36 negatives: 180'000
  - 24x36 slides: 40'000
  - Medium & Large formats: 80'000
- Keys:
  - TIFF 48 bits (RGB) & 4800 ppi (for 24x36 size images)
  - File naming pattern to enable album identification
- Challenges:
  - Metadata enrichment
  - Merge with digital-born content



PhotoLab Archives

Search 19,066 records for:

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

search also PhotoLab Archives Images

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:

[Inspected Albums](#) [Selection of our](#) [Not yet attached images](#) [All Images](#) [All Albums](#)

Photos 142 records found 1 - 100  Search took 1.10 seconds.



PHOTO

## CERN World-Wide Web Days 8-9 March 1995

Over two hundred people, journalists and communication experts from all over Europe came to CERN for the World-Wide Web Days. Day one was devoted to practicalities, with talks from different speakers who all had a hand in the first years of its development.

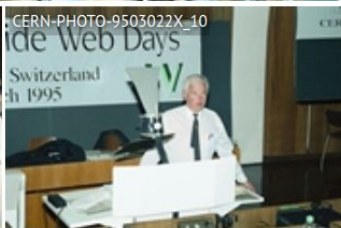
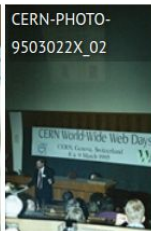
Date: Mar 1995

14 photos

 [Suggest a caption](#)

[Get the link](#)

[Embed Album](#)



# The challenge of preserving photos

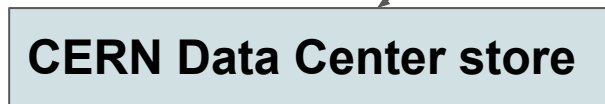
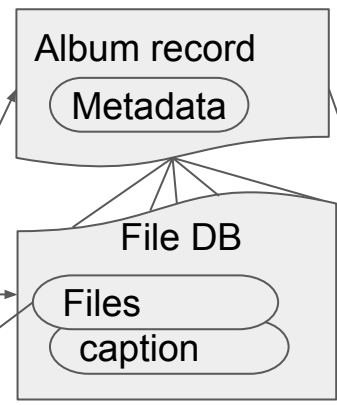
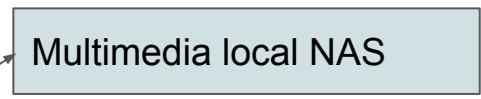
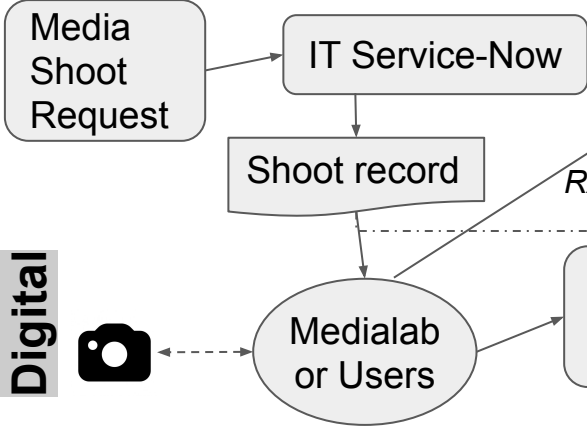
<https://youtu.be/CsmLb4cPZ0Q>



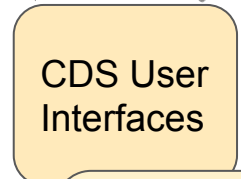
Vint Cerf  
(Google) at  
CERN

Workflows

Digital

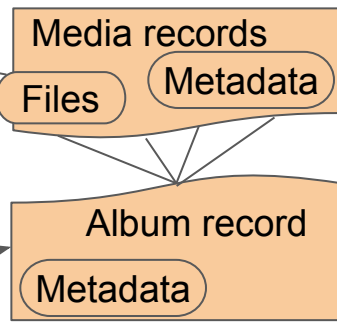
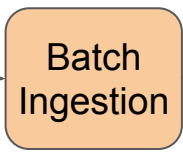
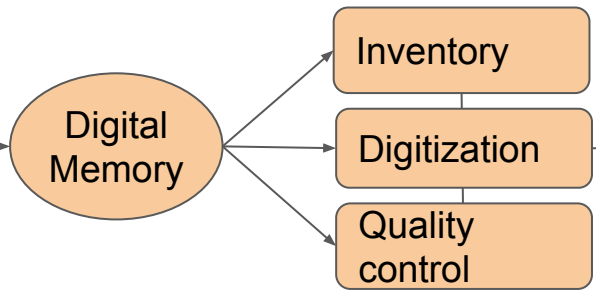


INVENIO



- Descriptive Metadata
- Access Restriction
- Copyright & license
- Authorities

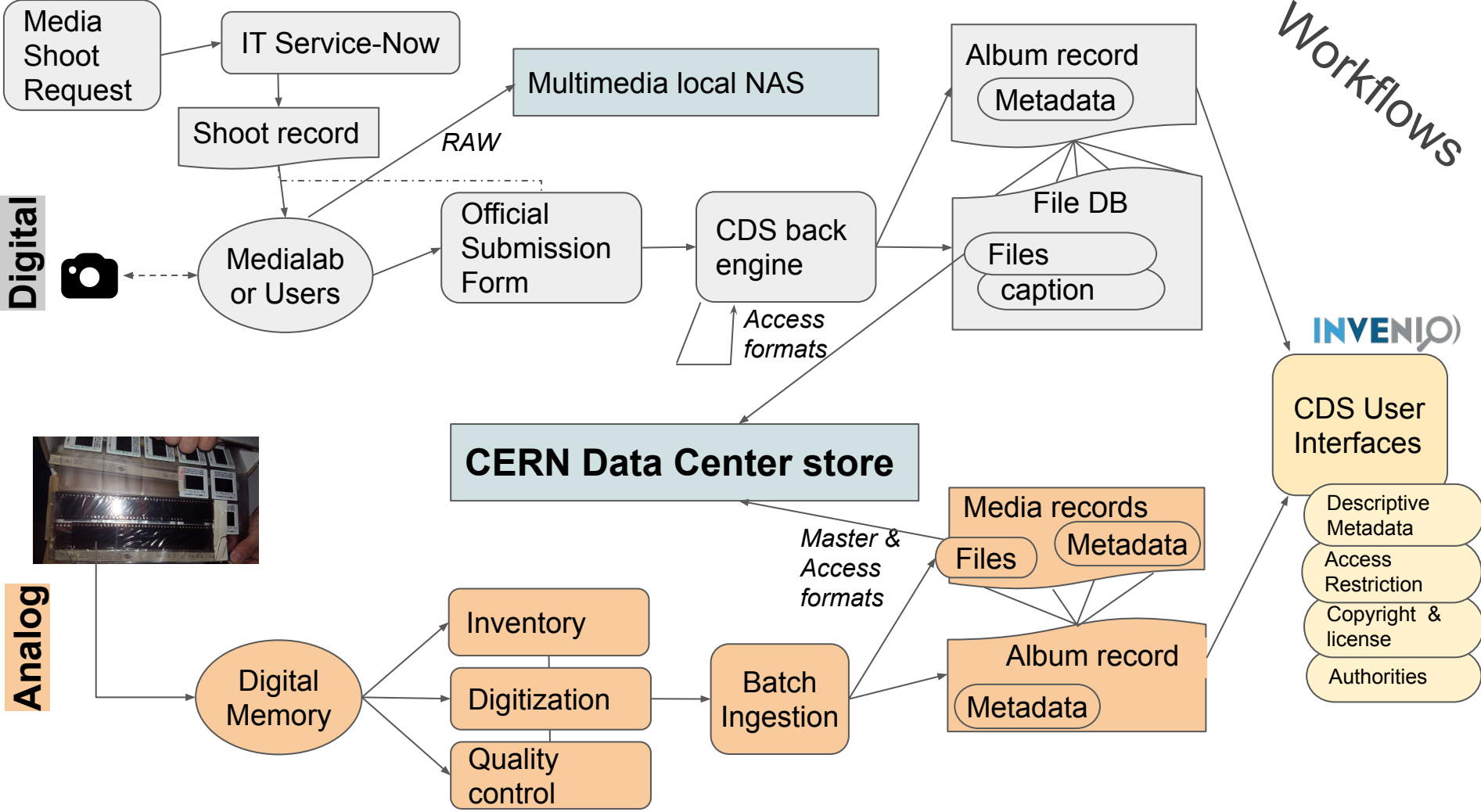
Analog



RAW

Access formats

Master & Access formats











# Status Report of the DPHEP Collaboration: A Global Effort for Sustainable Data Preservation in High Energy Physics

[www.dphep.org](http://www.dphep.org)

Thursday 25 January 2018

## Status Report of the DPHEP Study Group Towards a Global Effort for Sustainable Data Preservation in High Energy Physics

[www.dphep.org](http://www.dphep.org)

### Abstract

Data from high-energy physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. An inter-experimental study on HEP data preservation and long-term analysis was convened as a panel of the International Committee for Future Accelerators (ICFA). The group was formed from large collider-based experiments and investigated the technical and organisational aspects of HEP data preservation. An intermediate report was released in November 2009 addressing the general issues of data preservation in HEP. This paper extends the intermediate report. It provides an analysis of the research on data preservation and a detailed description of the various projects at experimental, laboratory and international levels. In addition, the paper provides a concrete proposal for an international organisation in charge of the data management and policy in high-energy physics.

### Abstract

Data from High Energy Physics (HEP) experiments are collected with significant financial and human effort and are mostly unique. An inter-experimental study on HEP data preservation and long-term analysis was convened as a panel of the International Committee for Future Accelerators (ICFA). The group was formed from large collider-based experiments and investigated the technical and organisational aspects of HEP data preservation. An intermediate report was released in November 2009 addressing the general issues of data preservation in HEP. This paper extends the intermediate report. It provides an analysis of the research on data preservation and a detailed description of the various projects at experimental, laboratory and international levels. In addition, the paper provides a concrete proposal for an international organisation in charge of the data management and policy in high-energy physics.

This status report of the DPHEP collaboration details the project from 2013 – 2015 inclusive.

## CERN and ISO 16363 Metrics

CERN produces and preserves a variety of data that can be broadly grouped into the following categories:

- Scientific data from the various experiments conducted at CERN;
- Papers and publications, many of which derive directly from the above and / or are related to R&D, service deployment and operations etc;
- The “Digital Memory” of the organisation, which includes minutes of meetings, videos, digitised images and so forth.

This list is non-exhaustive and may be amended in future revisions of this document.

This document contains a set of responses to the metrics contained in ISO 16363 – itself based on ISO 14721, the reference model for an Open Archival Information System (OAIS).

Many of the metrics below refer to the host organisation whereas some require detailed discussion that varies according to the type of data being preserved, the associated services, as well as the *designated communities* who might access or re-use this data.



Study Group for Data Preservation and Long Term Analysis in High Energy Physics

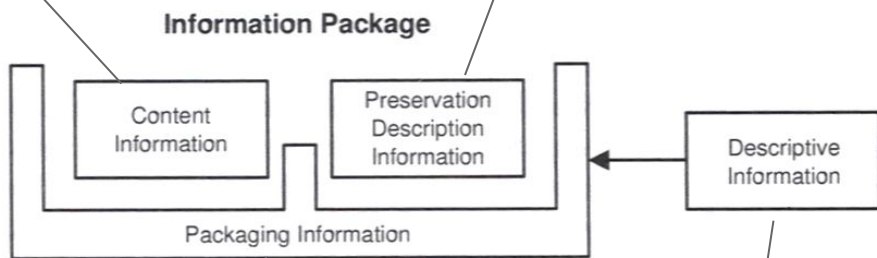


International Collaboration for Data Preservation and Long Term Analysis in High Energy Physics

# An Archival Information Package (AIP)

E.g. title, author, abstract, etc

E.g. checksum, digital signature



E.g. directory structure, filenames, tape marks

E.g. access control, finding aids

→ Must have redundant copies

→ Must be regularly checked

→ Must be supported by preservation plan

→ Must be sustained by an organizational policy

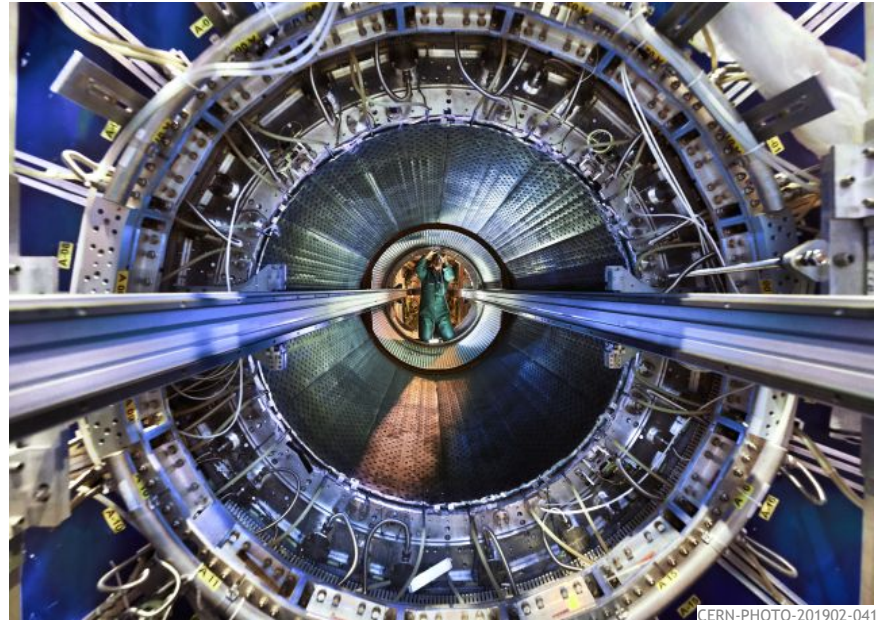
**Unique and  
Expensive**

**Different**

**Massive**

- 100 TB/experiment at LEP
- 1-10 PB/exp at HERA, TEVATRON at BaBaR
- X00 PB/exp at LHC
- >10 EB planned at HL-LHC

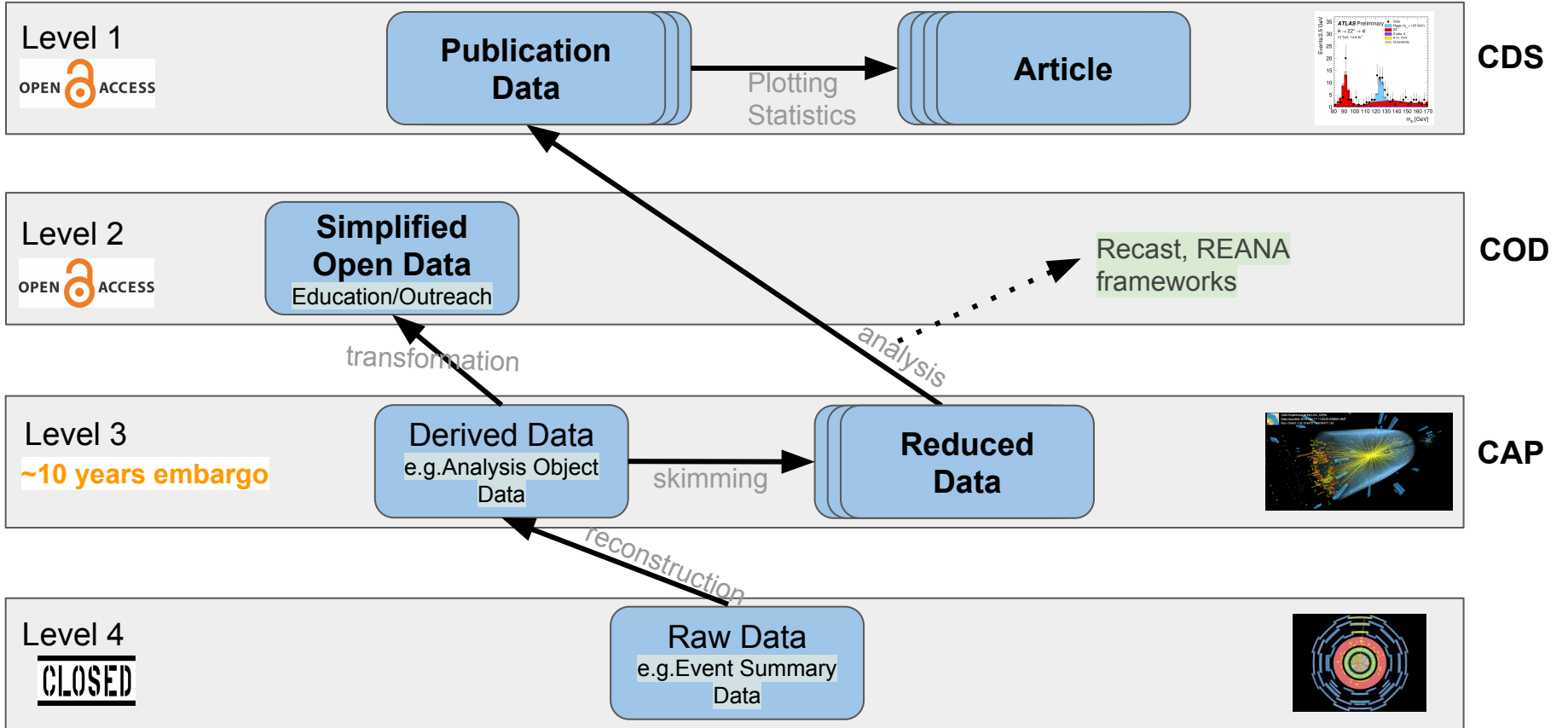
## HEP DATA



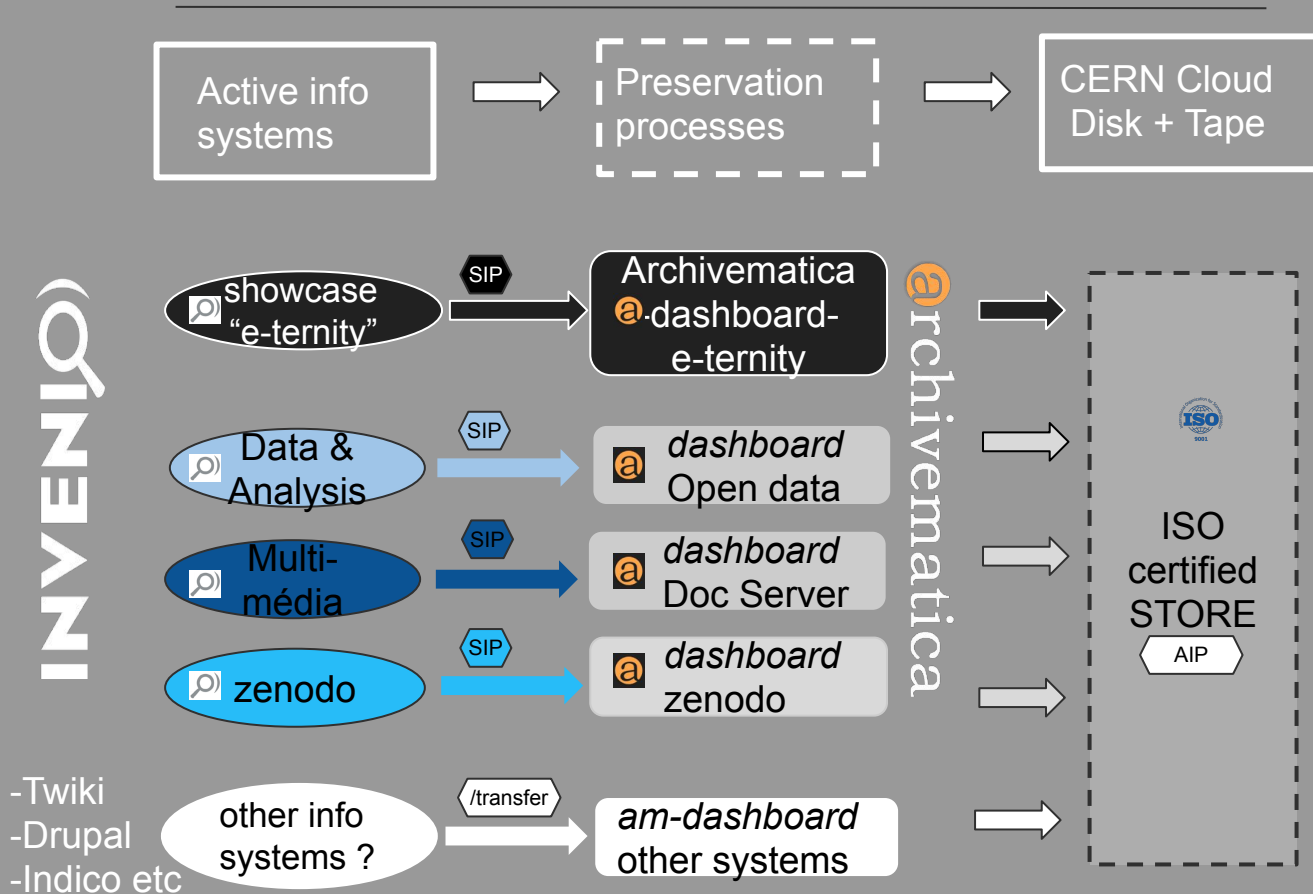
CERN-PHOTO-201902-041

**and complex !**

# Preserve What ? By Who ?



# E-Ternity Dark Archive (proof of concept)





# (1) VIRTUALIZING Archivematica TO RUN ON CERN CLOUD

[github.com/CERN-E-Ternity/archivematica-puppet](https://github.com/CERN-E-Ternity/archivematica-puppet)

# (2) CREATING A SIP STORE WITH Invenio

[Baglt File Packaging Format\(IETF\)](https://tools.ietf.org/html/draft-ietf-baglt-file-packaging-format) [github.com/inveniosoftware/invenio-sipstore](https://github.com/inveniosoftware/invenio-sipstore)

# (3) AUTOMATING THE SIP PROCESSING INTO AN AIP

○ [github.com/inveniosoftware/invenio-archivematica](https://github.com/inveniosoftware/invenio-archivematica)

Invenio-Archivematica

build passing coverage 95% pypi package or version not found

Invenio 3 module to connect Invenio to Archivematica

# Difficulties with revised files and with huge size files

“Weak Archiving”

HT primary dataset in AOD format from RunA of 2011 (/HT/Run2011A-12Oct2013-v1/AOD)

/HT/Run2011A-12Oct2013-v1/AOD, CMS collaboration

Cite as: CMS collaboration (2016). HT primary dataset in AOD format from RunA of 2011 (/HT/Run2011A-12Oct2013-v1/AOD). CERN Open Data Portal. DOI:10.7483/OPENDATA.CMS.28VG.KY9R

Dataset Collision CMS Collision energy 7TeV Accelerator CERN-LHC

## Description

HT primary dataset in AOD format from RunA of 2011. Run period from run number 160404 to 173692.

## Notes

This dataset contains all runs from 2011 RunA. The list of validated runs, which must be applied to all analyses, can be found in

[CMS list of validated runs Cert\\_160404-180252\\_7TeV\\_ReRecoNov08\\_Collisions11\\_JSON.txt](#)

## Characteristics

Dataset: 88881838 events, 3873 files, 14.7 TB in total

Explore more than **1 petabyte** of open data from particle physics!

Start typing...

search examples: [collision datasets](#), [keywords:education](#), [energy:7TeV](#)

### Explore

[datasets](#)  
[software](#)  
[environments](#)  
[documentation](#)

### Focus on

[ATLAS](#)  
[ALICE](#)  
[CMS](#)  
[LHCb](#)  
[OPERA](#)

Search

the jet transverse momenta (HT); or a high MR  
ing in some cases kinematic requirements on the

## HLT trigger paths

The possible HLT trigger paths in this dataset are:

# Conclusion

- Moving from Preservation by chance to Preservation by mission
  - The Digitization of 20th Century Multimedia
  - The Collision with DPHEP
  - The Attempt to Converge into an OAIS Dark Archive
- 
- **Enlightening the digital archive....**

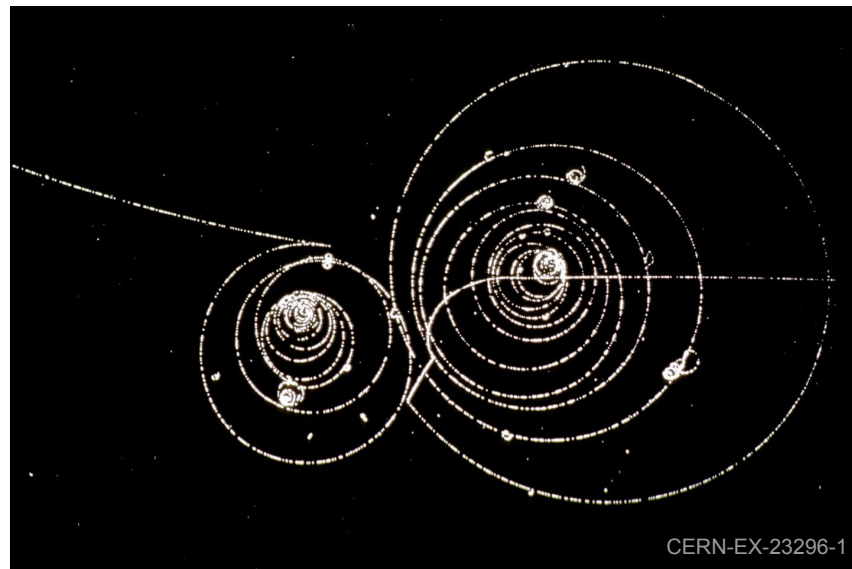
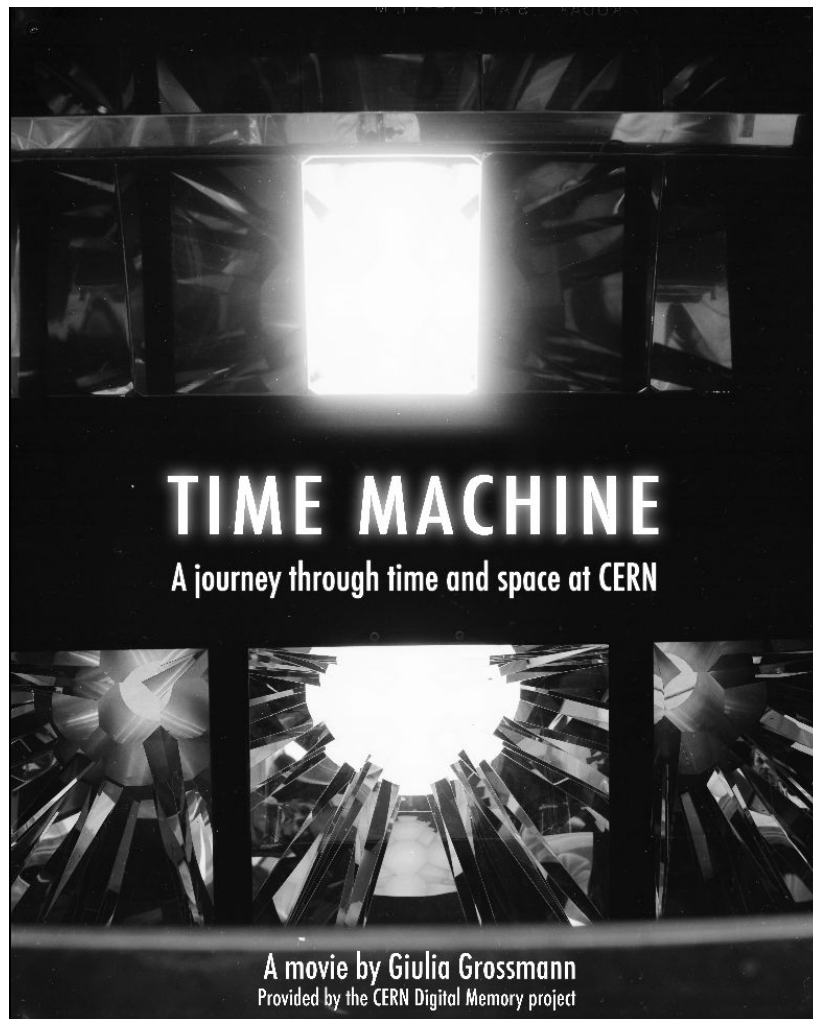




PHOTO DISPLAY: the EPFL Memory Collider

Docu-fiction with CERN  
Historical Videos



# TIME MACHINE

A journey through time and space at CERN

A movie by Giulia Grossmann  
Provided by the CERN Digital Memory project



# An artistic surprise

The [VolMeur](http://volmeur.org) collection of the CERN Digital Memory project

<http://volmeur.org>

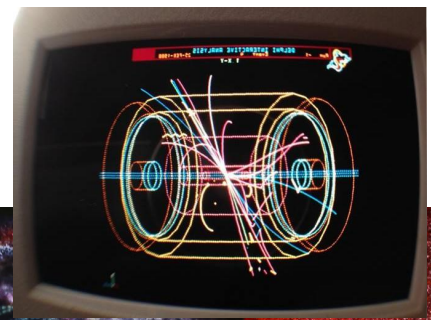
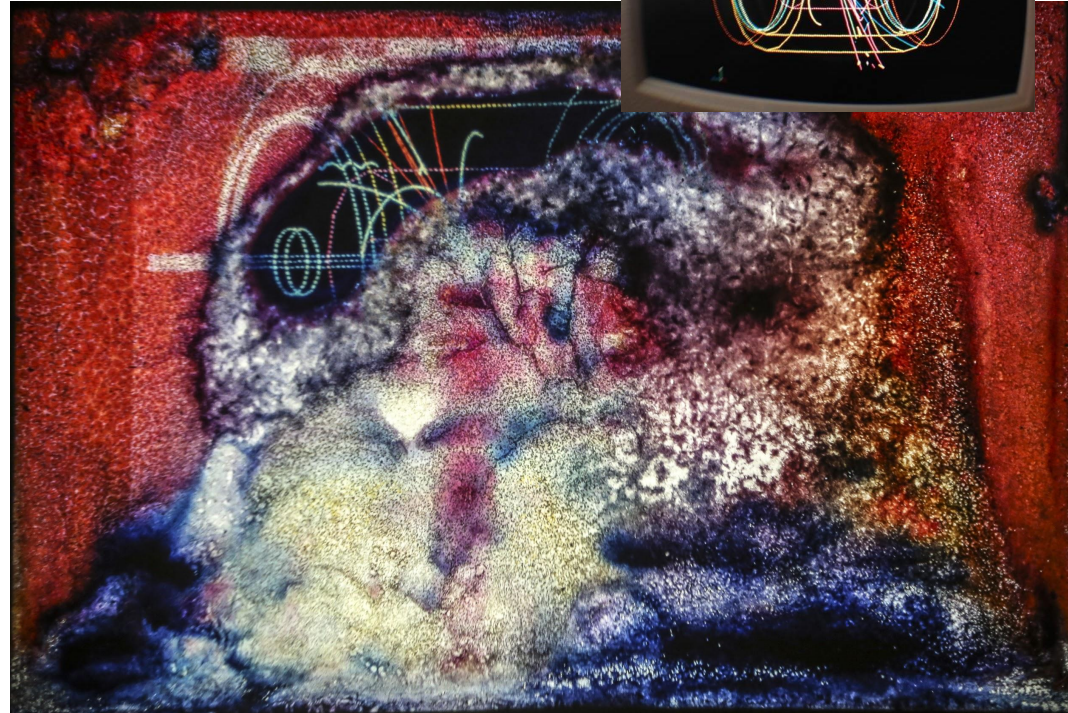


Arts visuels Publié hier à 21:57

Mal conservées, des diapositives du CERN se muent en oeuvres d'art



Des diapos endommagées du CERN devenues des oeuvres d'art L'actu en vidéo / 1 min. / hier à 15:40



Currently Exhibited at New-York MFTA Gallery



[Surprise!](http://volmeur.org)

Thanks for  
your  
attention !

[Live stats](#)