

Archives de France
Centre des Archives Contemporaines

CONSTANCE :
*Twenty years
of data bases archiving*

ERPANET WORKSHOP

Bern April 09-11 2003

Summary

- 1- Institutional, structural and legal frames
- 2- General strategy and policy
- 3- Processes from appraisal to ingestion
- 4- Data base documentation and metadata
- 5- Access and re-use : several policies
- 6- Conclusions

Preliminary question :

(for inquisitive People)

What is CONSTANCE ?

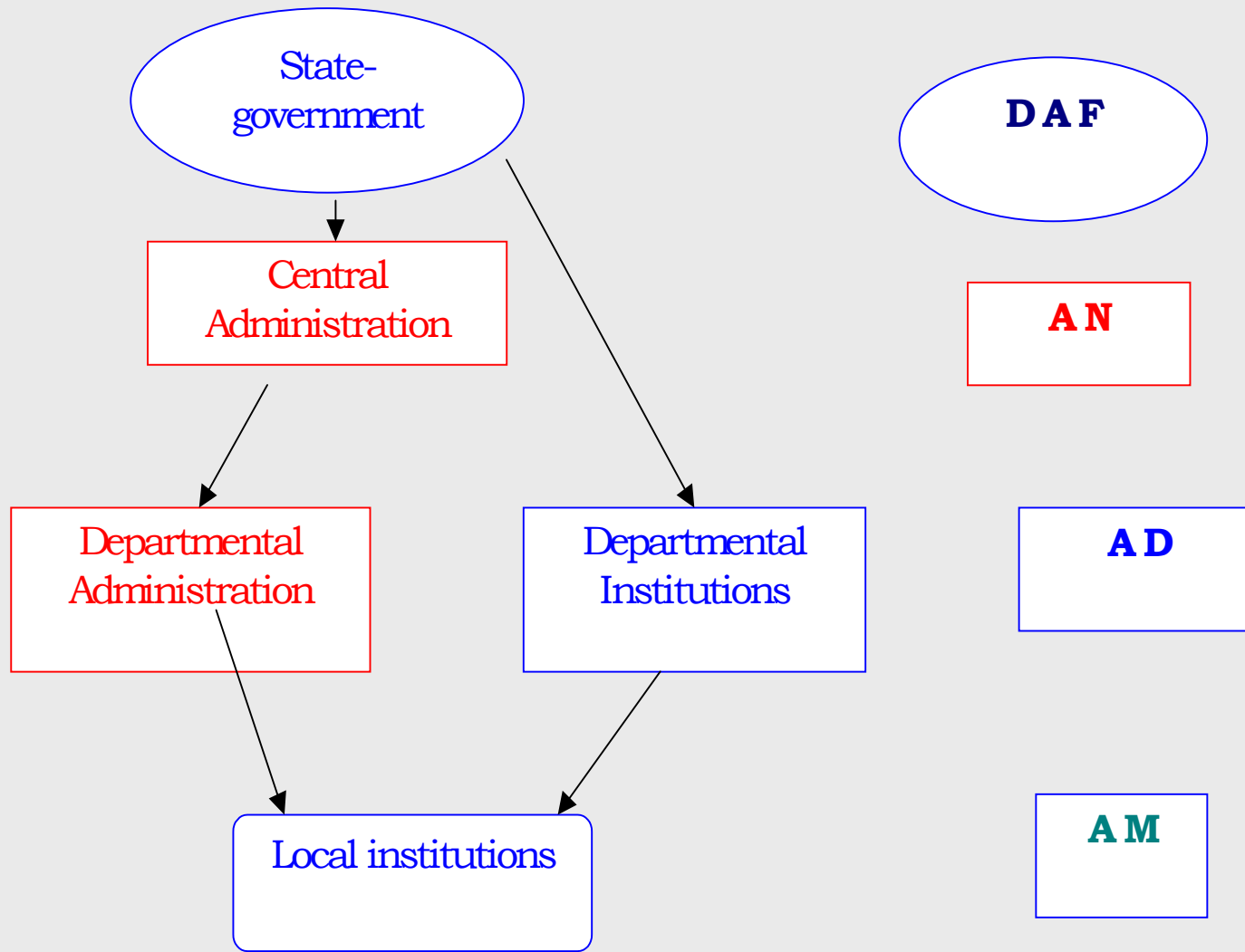
CONSTANCE

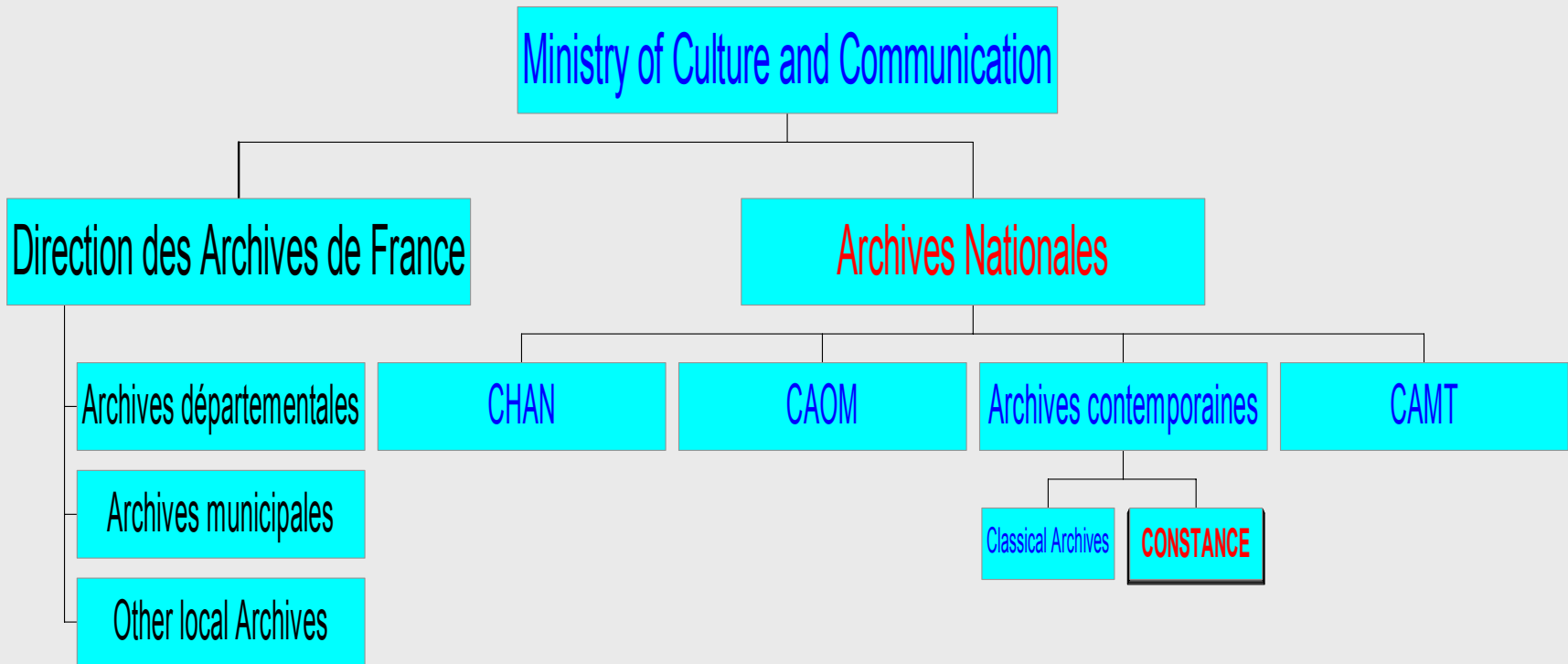
- CONSERVATION et TRAITEMENTS des ARCHIVES
NOUVELLES CONSTITUÉES par l'ELECTRONIQUE
- Preservation and Treatments of New Archives
issued from computer processing (or IT)
- This acronym was formulated in 1979 by an
Archivist.

1- CONSTANCE :

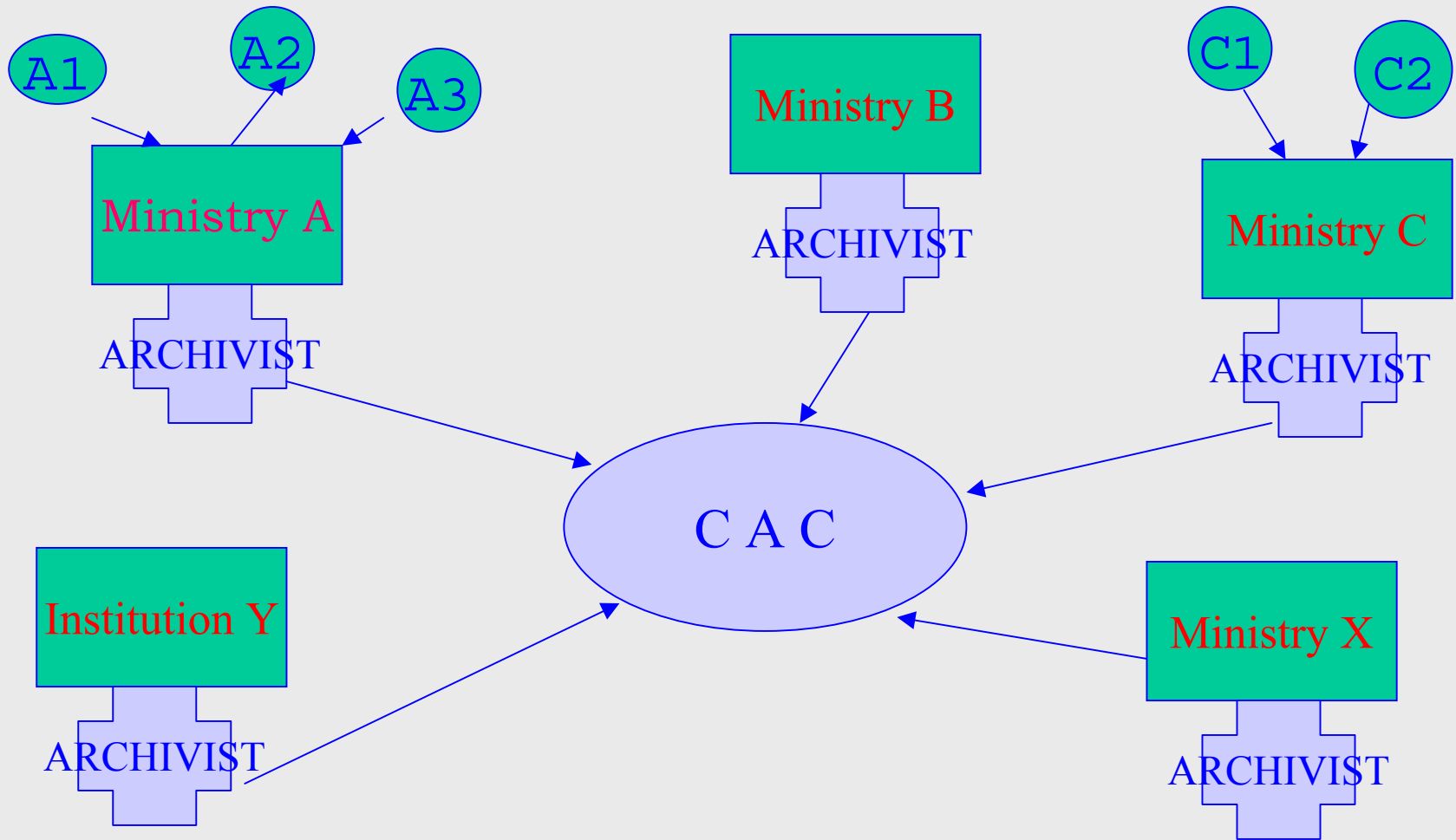
Institutional, structural and legal
frames

Archives and French Institutions





CONTEMPORARY ARCHIVES : the Network



Legal frame : accessibility

- The 1979 law on the archives :
all documents produced whatever their form and support...are an archival material
- communication regulations :
 - all documents : 30 years without access, except on a derogatory basis
 - all statistical nominative files : 100 years without derogation
- Public access : free for every citizen

Legal frame : citizens protection

Since 1978 : CNIL (Commission Nationale de l'Informatique et des Libertés)

- Agency for the respect of the 1978 law on IT and Liberties
- Every nominative file must be declared (private or public)
- Any IT project involving nominative data must be submitted to the agency's opinion
- When needed, the French National Archives should be mentioned as final recipient of the data as agreed by CNIL.

2- CONSTANCE :

GENERAL STRATEGY AND
POLICY

A big Question :

- Why preserve on the very long term so many data bases at such a cost ?
- Because you cannot envision the 20th century society memory without taking into account the prominent place of the management, statistic and scientific data bases, and their intensive and unavoidable use.

Constance : the chronological steps

- 1978-1983 :the birth
- 1983-1986 :the first steps
- 1986-1993 :a too fast growth
- 1993-1995 : the transition
- 1995-1997 :moulting and migration
- 1998-2000 :a new vision
- 2001-2003 :new fields to investigate



Main options taken

- Data bases of historical value, at the national level are to be preserved indefinitely,
- The CAC is in charge of the Constance programme,
- The Constance team will be the operational core,
- A computer centre was set up in 1983 in the CAC,
- The team is composed of archivists and IT specialists,
- The team can deal directly with the producers and users (end-users or data base managers) when needed

Technical options

- Data are extracted from data bases by the producer in flat files
- No software components or elements are preserved
- Data format is the Ascii character mode (one Ascii character in a byte)
- Metadata must be delivered with data files
- Data base producers or managers comply to the National Archives technical recommendations

CONSTANCE : Essential functionalities

5- Technological watch and advice

4-Help and expertise for

3- Ingestion and integration

2- Digital objects

**1-Management of
technological
resources**

Conservation management

management

Appraisal and collect

Awareness and education

3- CONSTANCE :

From appraisal to ingestion

Process from appraisal to ingestion

- Archivists detect a data base
- Appraise it
- Ask IT staff for extraction of data in flat file(s)
- Ask the relevant persons for metadata and documents
Valid the components before sending the bundle to the CAC
- Constance team ingests the documentation and the data files
- Archival updates in various finding aids are made
- Then the follow up cycle for preservation of accessibility and integrity begins.

The main steps (1)

- Archivist detects a data base of historical value in a governmental organisation
- A presentation of the functional, legal, technical aspects is made by the owners and managers of the data base
- An informal protocol of acquisition is agreed between the two parties
- The data base managers extract the relevant data, following the National Archives technical specifications
- The data base managers gather the technical sets of metadata concerning each file provided for long term preservation

The main steps (2)

- The data base managers and owners gather and articulate the general documents related to this data base : organisation, design, purposes and use.
- The archivist in charge appraises the documents and the metadata
- An output of a significant data set is produced and checked against the documentation
- An electronic notice is sent to the CAC for archival management purpose
- File (s), documentation and metadata are sent to the Constance team
- A final and more technical check is made before proceeding with file storing.

Ingestion tasks

- check and upgrade the documentation and metadata
- control the file (s)
- scan different parts of documentation and store them in TIF format in an ERMS
- if in electronic form, store them also in an ERMS
- copy the data files onto the conservation media (DLT tapes for the time being), twice for security.
- Create and store the first set of metadata related to the file(s) integrity and traceability on the long term

4- CONSTANCE :

Data base

Documentation and Metadata

The different sets of documents

- A set of 16 technical metadata for each file
- The lay out (or design) of the file record (s)
- The data and codes dictionary, in alphabetical order
- An output of some records from the file to be transferred
- Sets of input screens or forms, notices and results
- Bibliographical references of publications when relevant
- The report concerning the conception of the project
- Manuals concerning the uses and technical management of the data base
- The legal documents (or references) attesting the validity and rights of the procedure involving the data base, its content and uses

A file metadata set (1)

- 1- Information system name
- 2- File name
- 3 Media references and characteristics
- 4- Copy date of the file onto the media
- 5- Operating system references
- 6- Software name and version supporting the database
- 7- Dates (first and last) of the data in the file
- 8- Records file format (v, f, string..)
- 9- Size in bytes of the records file
- 10- Number of records files
- 11- Number of objects contained in the file

A file metadata set (2)

- 12- number of different variables in the file
- 13- total number of bytes
- 14- sort keys
- 15- names and references of the linked or related files
- 16- date and references of the completion of this inventory form.

5- CONSTANCE :

Data base

Access and re-use

Some figures :

- 1981 : 1 file
- 1983 30 files
- 1986 300 files
- 1994 5000 files
- 2002 6000 files

Several possible policies :

- Deliver a copy of the file and its relevant metadata
- provide in situ some expertise
- provide some tools and computer resources
- provide technical resources and experts' help
- provide results to some "clients" on request

But who pays what ?

Some links :

- *Jean-pierre.teil@culture.gouv.fr*
- *<http://www.archivesdefrance.gouv.fr>*
- *<http://www.archivesnationale.gouv.fr>*
- *<http://www.archivesnationales.gouv.fr/cac/fr/index.html>*