# Archivage de données cliniques au moyen de CDISC ODM

**Pilote mené par Sanofi-aventis et Accovion entre 2004 et 2006**

**Pierre-Yves Lastic**
**Groupe d'Utilisateurs Francophones de CDISC**

**9 octobre 2008 – APHP / Hôpital St-Louis, Paris**



sanofi aventis
Because health matters

# Table des matières

- Origine et définition du projet
- Raisons du choix de CDISC ODM
- Obstacles techniques rencontrés
- Conclusions

*NB: La présentation est en anglais, vu que toute la documentation du projet est dans cette langue*

# Project Background

- Clinical data are collected and stored in a CDBMS together with historic information (audit trail) and structural information (metadata) under the responsibility of Clinical Data Management.
- There are regulatory requirements concerning retention time of clinical data to be fulfilled. Often the data retention time outlasts the lifetime of software systems. The need for application retirement causes the need for application independent archiving of clinical data.
- Archiving of CDBMS data is necessary in order to be able to explain the origin and history of each single data item, and NOT to be able to rerun analyses on the data, regulatory agencies may ask for. For the latter need, SAS archives have to be kept, which have to comply quite different requirements.

# Project Overview

- Develop an Archiving Methodology
  - Archive Format was Proposed (CDISC/ODM)
  - Methodology Proposed
    - Phase 1: Data Extraction and Staging
    - Phase 2: Transformation into CDISC/ODM
    - Phase 3: View the XML files with ODM Viewer
    - Phase 4: Archive with ADELE (Sanofi-aventis archiving system)
  - Proposed to Create a Tool for Users to Extract & Archive
- Test the Methodology in a Pilot
  - Archive at least one complete study
  - Start with Clintrial v3.3

# Requirements (1/2)

- Provide an archiving concept and strategy to be used for long term archiving of Clintrial study data.
- The concept has to be flexible enough to be embedded into a Clintrial retirement project on the one hand or into a study archival concept comprising archiving of Clintrial & SAS data and full study documentation like CRFs, DCFs, TMF and other study documentation and reports.
- The concept has to follow the Sanofi-aventis interpretation of 21 CFR part 11 Section 11.10
- Data have to be archived for a period of time (called the record retention period) characterized in Commission Directive 2003/63/EC.
  - Since there is no fix number of years, but only a minimal time period for record retention given, the archival period has to be envisioned to be at least 25 years in general.
  - For this period data has to stay readable, secured and easily retrievable

# Requirements (2/2)

- Data / information has to be archived, not systems / technology / functionality. Information has to be stored in an industry standard, open systems format, which may and shall be independent from the database, versions of soft- and hardware, it has been kept in originally.
- Since secure archival of data allows system retirement and physical removal from the system as a consequence, procedures have to be **proposed** which verify the readability of the archive data over time by periodic read attempts.
- A proof of concept has to be put in place in order to show that the concept is feasible.
- Choosing a Phase III study has the advantage compared to choosing a Phase I study, that the expected diversity of data is part of the test

# Other archiving considerations (1/2)

- A complete study archival concept has to comprise archiving of Clintrial & SAS data and full study documentation like CRFs, DCFs, TMF and other study documentation and reports.
- All components of study information have to be physically archived together, e.g. all electronic data on one CD, and the link between electronic medium and paper binders with data of the same study has to be obvious.
- All electronic components of a study archive have to be in an open industry-standard format. Clintrial data in CDISC ODM or at least in XML, SAS data in SAS transport format, and documents/reports in pdf format.
- Since secure archival of data allows system retirement and physical removal from the system as a consequence, procedures have to be **put in place** which verify the readability of the archive data over time by periodic read attempts.

# Other archiving considerations (2/2)

- The design for SAS archives may follow the following considerations:
  - In principle, SAS programs are independent of the version and could be archived directly as ASCII-File.
  - Other SAS-Objects normally belong to a version and it could be problematic to use them in newer versions:
  - SAS Datasets could be converted to a SAS-program as ASCII-File that contains statements to reproduce the dataset in the current version.
  - SAS views created with proc sql could be saved also as a SAS-program that will reproduce that view in the current version
  - SAS format catalogs could also be converted to a SAS-program that will reproduce the formats in the current version
  - for SAS macro catalogs only the list of included macros could be saved, as these are in a compiled format and SAS offers no way to reproduce the source code. But the source code of this macro should also be available in the program, that created the entry in the macro catalog.
  - If there are other objects, then they need to be investigated.
  - When archiving specific projects it is also necessary to archive the used global objects, like coding tables.

# Scope Limitations

- Archiving does not have to maintain decommissioned soft- and hardware
- Clinical data archives do not have to contain batch load files for external data (like lab data),
  - but only the converted format of those data after its loading into the database.
- Clinical data archives for historical studies do not have to contain definition or layout of data entry screens.
  - However for future RDC studies with calculations or plausibility checks during DE, data entry screens have to be archived!
- For historical studies there is no necessity to archive CRFs and DCFs as scanned images but only in paper format.
  - However for future studies image archiving is required!
- Clinical data archives do not have to contain system validation documentation of the IT systems it originally has had been stored in.
  - Those documents are stored/archived with IT software lifecycle documentation.
- Clinical data archives contain all information about and around a clinical study, but not the functionality of the CDBMS!
  - The archive contains data, validation rules and queries, but not the ability to rerun a validation rule and recreate a query.

# Why CDISC ODM ?

- Three alternatives for Long Term Clinical Data Archival were chosen
  - ASCII
  - SAS
  - CDISC ODM

# ASCII

| Pros | Cons |
| --- | --- |
| • Most common format<br>• Can be read from any platform with any editor<br>• Stays unchanged or downwardly compatible “forever” | • Has no internal structure<br>• There is no way of interpretation of data by any viewer |

# SAS

| Pros | Cons |
| --- | --- |
| • SAS archives for study data exist anyway<br>• SAS offers a whole suite of tools for retrieving and analyzing data | • is not an open format. SAS software licenses are needed to read data |

# CDISC ODM

| Pros | Cons |
| --- | --- |
| • Is going to become the industry standard for clinical data archives<br>• License free viewer exists<br>• Is optimally structured for keeping data, metadata, codelists etc.<br>• Data and format is stored together (XML) | • Higher effort to transform data, metadata etc. in ODM format<br>• Very storage space consuming |

# Pilot Overview

- Objectives:
  - Implement a solution and process to allow the closure of Clintrial V3 databases, taking into account the status of the studies hosted in these databases
- The Pilot has been divided in 2 parts :
  - Migration of the on-going studies
    - Large Oncology Study
    - Three other studies
  - Archiving the databases
- The Pilot was stopped in December 2006 due to other, higher priorities

# Pilot Organization

- Part of the work was subcontracted to Accovion
  - Accovion delivery:
    - Study set-up and Edit checks for the 3 studies
    - SAS/SQL users programs adapted
    - Program of treatment of queries with specific status adapted
    - Program for the mapping adapted
    - Script to transfer logging data from CT3 to CT4 adapted
  - sanofi-aventis part:
    - Validation of each delivery
    - OQ of tools needed for each study
    - Distribution of codelist/protocol/dictionary
    - Installation of packages
    - Migration of data

# Content of Clinical Study Archive (from Clintrial) (1/2)

- Clinical data.
  - All content from _DATA tables of all study panels
- Audit History.
  - All content from _AUDIT tables of all study panels
- Content of _UPDATE tables.
  - Although _UPDATE tables of study panels are expected to be empty for a closed study, if they do contain data by chance, it has to be archived.
- Database design specifications. A complete description of Metadata including
  - • Table descriptions
  - • Item definitions per table
  - • Associated codelists per item
  - • Associated Thesauruses per item
- However there is no possibility of version control.
  - Metadata descriptions are taken in the version as of creation of archive.

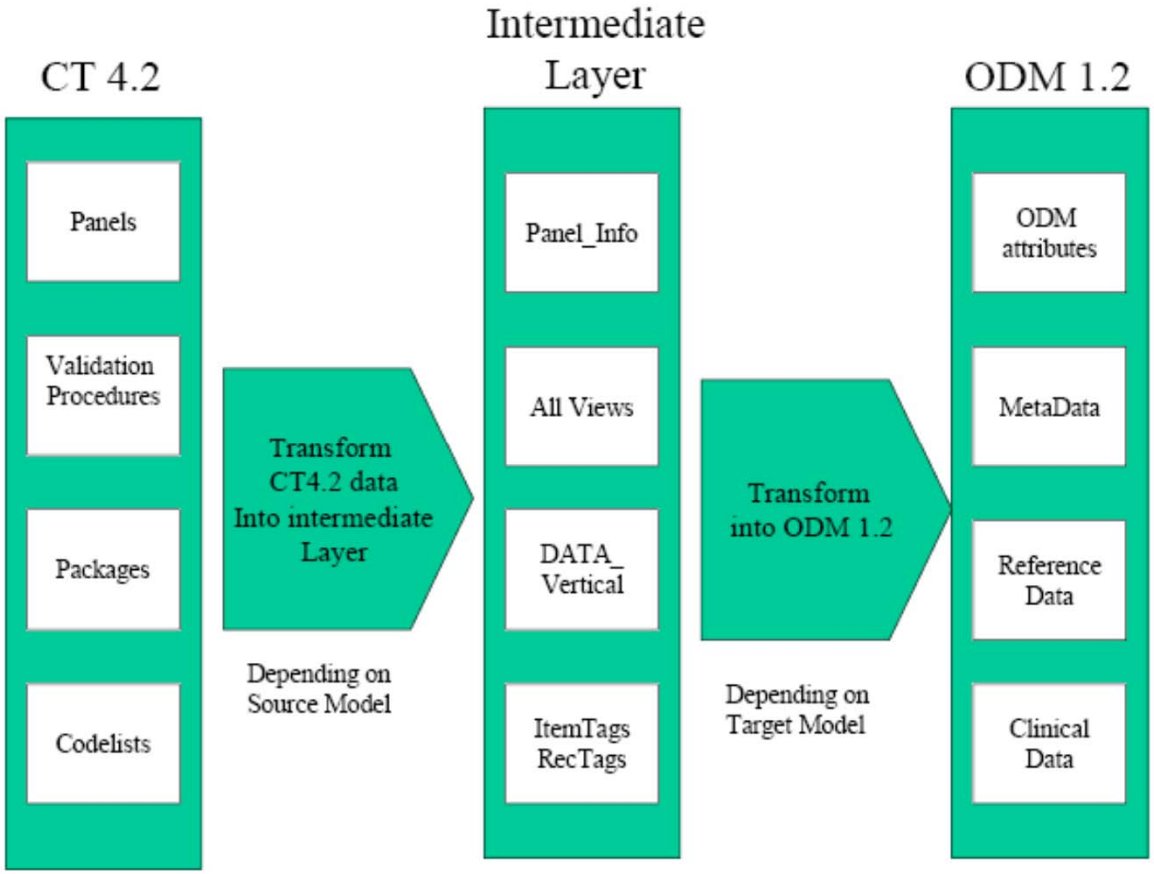
# Content of Clinical Study Archive (from Clintrial) (2/2)

- Validation rules per panel.
  - The PL/SQL code of all validation rules will be listed inclusive called procedures and functions. However there is no possibility of version control. Validation programs and procedures are taken in the version as of creation of archive.
- Tags.
  - The content of TAGS and TAGS_AUDIT table as well as the tag definitions.
- Queries.
  - All query and query_item information from Clintrial Resolve.
- Codelists.
  - All referenced codelists with code, value, short and long label.
- Gain additional potential benefit from archiving Oracle dump files besides the converted data (e.g. in XML format).
  - Dump files are easy to create and for a certain period of time are reloadable into Oracle and retrievable.
- Database lock / unlock information including reason.

# Not part of Clintrial study data archive

- Dictionaries.
  - Dictionary content is typically stored together with SAS data for clinical evaluation. Together with study data, the coding result (associated code) is of interest rather than code and decode.
- Classify Omissions.
  - Tracking information about how omissions became Synonyms is of no interest. If queries have been raised which led to modification of .as reported. terms, this information can be tracked in Resolve data and on DCFs.
- Users, Roles, Access.
  - For closed studies only read access exists on data. Access modifications are not audited in the database and cannot be tracked nor listed out of the database.

# From Clintrial to ODM






sanofi aventis
Because health matters

# Accomplishments

- Phase 1: Extract Data Clintrial v3.3
  - Achieved: use of tool TOAD; it is Possible to Read Older Oracle Versions of CT3.
  - Extracted Data saved in a ‘Staging Area’
- Phase 2: Transform to CDISC/ODM v1.2
  - Achieved: XML/ODM files created
  - ODM DTD v1.2 Used as a Data Standard
  - Verified to Conform to ODM v1.2 with ‘ODM Checker’
- Phase 3: Read ODM/XML Files
  - Partially Achieved: Small Files can be Read, but not Large Files
- Phase 4: Archive to ADELE
  - Partially Achieved: Archive not Completed but Verified by ADELE Team

# Problems Encountered (1/3)

- It takes too Long to Archive one Study
  - Approach: Archive one Study at a time
    - Extract protocol, data, etc.
    - Map data, field by field, transform each panel, etc.
  - Resolution: Extract/Map/Transform Many Like-Studies Together
- Use of XML Viewers is not Effective
  - Approach: Use ODM Viewer to Read ODM Files
    - No XML Viewer was effective reading all XML files created
    - ODM viewer is very good for small files
    - Unusable for normal large XML files, e.g., from one panel
  - Resolution: XML files are huge so they must be divided

# Problems Encountered (2/3)

- No Solution to Divide & Form 'Ideal' XML Files
  - Approach: Divide files by patient, submission structure, etc.
    - Too much Effort to Analyze each Study Separately to Determine Best Structure
    - Too many Small Files to Manage
  - Resolution: Do Not Use ODM Viewer to Read ODM Files
    - Purpose of Reading Archived ODM Files is to resubmit, re-analyze, etc. SAS will be Used in Almost All Cases; Verified w/BioStats

# Problems Encountered (3/3)

- Develop a User – Oriented Tool
  - Approach: User Managed Archiving
    - Too much effort on User Interface, etc. with no Archiving Results
  - Resolution: Extract/Map/Transform Periodically Many Studies at the Same Time
- Effort to learn each XML+ODM+Clintrial is huge
  - Approach: Use Internal Resources; too much discovery time
  - Resolution: Use External Experts and Learn from Them

# Lessons Learned

- Global View Needed
  - Do not Archive Study-by-Study
  - Revise Archiving Methodology
- CDISC/ODM: Continue to Use ODM; it Works
  - Phase 1: Data Extraction: Many Studies at Once
  - Phase 2: Transform like Studies Together into CDISC/ODM
  - Phase 3: Extract/Query XML files with SAS
    - Make the Effort to View/Query an Archived file with a Tool Suited for this purpose
  - Phase 4: Archive with ADELE: Continue to Use ADELE
- A User Based Tool is not Efficient
  - Abandon
- Use Industry Experts

# Recommandations

- Use Expert and Dedicated Resources
  - Involve Outside Archiving / ODM Experts
  - Train Internal Resources with them
- Use CDISC/ODM as standard for all Clinical Data Interfaces
  - Archiving is often considered as an ‘After-Thought’
  - Plan at the beginning for Study Migration and Archiving
  - One Standard for All Interfaces: import and export