Clinical Data Acquisition Standards Harmonization (CDASH)

Setting the Global Standard for Clinical Data

CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM

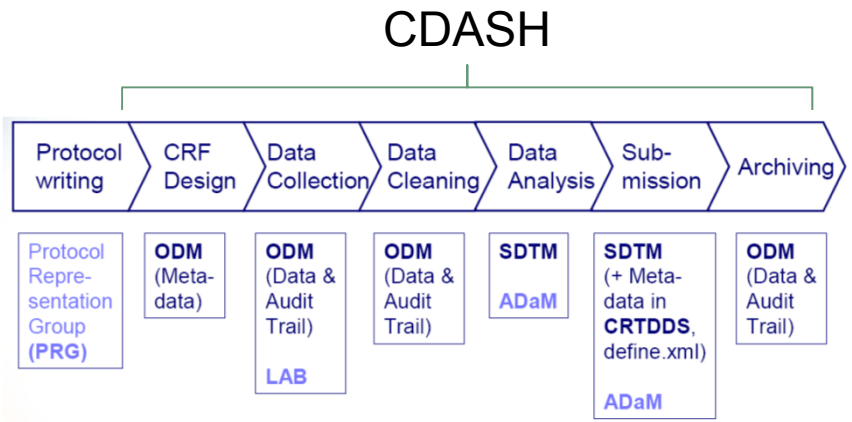
Sanofi Aventis

2008-02-05

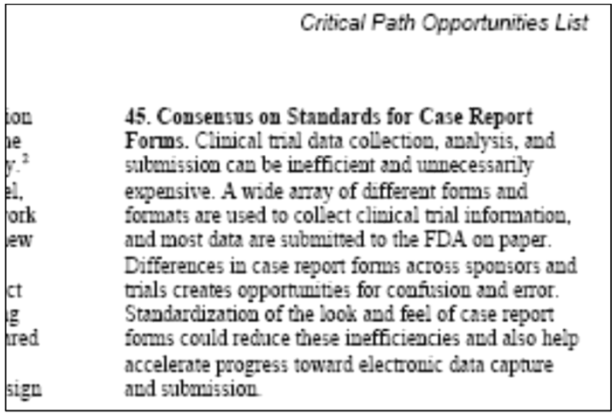
CDASH??

Clinical **D**ata **A**quisition
Standard **H**armonization

- Les 4 objectifs de CDISC / développement de standard pour:
  - L'acquisition de données
  - L'échange de données
  - La soumission de données
  - L'archivage de données



Pourquoi CDASH?



Critical Path Opportunities List

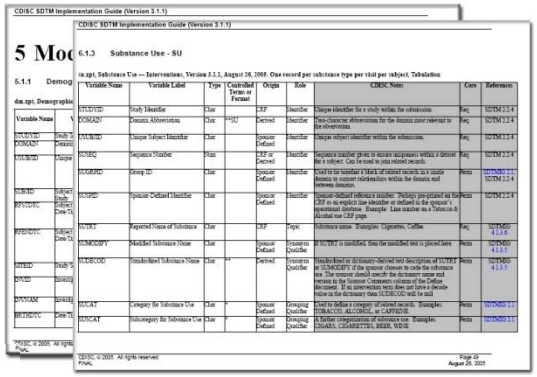
45. Consensus on Standards for Case Report Forms. Clinical trial data collection, analysis, and submission can be inefficient and unnecessarily expensive. A wide array of different forms and formats are used to collect clinical trial information, and most data are submitted to the FDA on paper. Differences in case report forms across sponsors and trials creates opportunities for confusion and error. Standardization of the look and feel of case report forms could reduce these inefficiencies and also help accelerate progress toward electronic data capture and submission.

"From the FDA's perspective, the quality and integrity of the data is paramount. Common standards for case report forms can improve both, and are also a crucial enabler for the biomedical research of the future, for example, to support genotypic and phenotypic evaluation of each subject. We appreciate CDISC taking the leadership role to start now to create the data collections tools for the future," states Dr. Janet Woodcock, Deputy Commissioner and Chief Medial Officer, FDA

- The CDASH project focuses on the FDA Critical Path Opportunity #45, and is supported by
  - a collaborative group of organizations comprised of the Association of Clinical Research Organizations (ACRO),
  - the Association of Clinical Research Professionals (ACRP),
  - the American Medical Informatics Association (AMIA),
  - Baylor College of Medicine,
  - the Clinical Data Interchange Standards Consortium (CDISC),
  - the Clinical Research Forum,
  - the Food and Drug Administration (FDA),
  - the National Institutes of Health (NIH),
  - (the Clinical Research Policy Analysis and Coordination Program, the National Cancer Institute (NCI-caBIG and NCI-EVS), the National Clinical Research Resources (NCRR), the National Library of Medicine (NLM), the National Institute of Child Health & Human Development (NICHD)),
  - the Critical Path Institute,
  - the Pharmaceutical Research and Manufacturers of America (PhRMA),
  - the Biotech Industry Organization (BIO),
  - the Society for Clinical Data Management (SCDM) and Duke Clinical Research Institute.
- CDASH was first announced by Dr. Janet Woodcock, (FDA), at the 2006 DIA Annual meeting.

Travail de CDASH

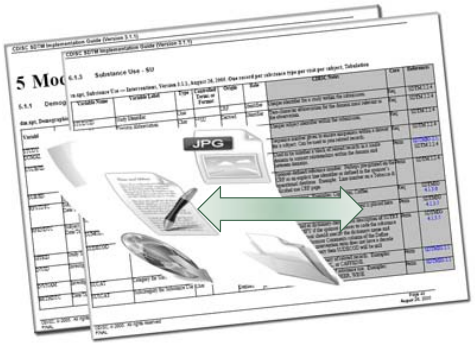
⑤ Mise au point des Packages CDASH



① SDTM

② CRF samples

③ Analyse du contenu ≠ présentation



④ Evaluation des similitudes / différences

pe.xpt, Physical Examination — Findings, Version 3.1.1, August 26, 2005. One record per body system or abnormality per visit per subject, Tabulation

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
STUDYID	Study Identifier	Char						
DOMAIN	Domain Abbreviation	Char	**PE					
USUBJID	Unique Subject Identifier	Char						
PESEQ	Sequence Number	Num						
PEGRPID	Group ID	Char						
PESPID								
PETESTCD								
PETEST								
PEMODIF								
PECAT								
PESCAT								
PELOC								
PEBODSYS	Body System or Organ Class	Char	**	CRF or Derived	Result Qualifier	1. Body system or organ class (Primary SOC) that is involved in an event or measurement from the standard hierarchy (e.g., MedDRA). 2. May be predefined from CRF page or obtained from coding.	Perm	SDTMIG 4.1.3.5
PEORRES	Verbatim Examination Finding	Char		CRF or Derived	Result Qualifier	Text description of any abnormal findings. If the examination was completed and there were no abnormal findings, the value should be NORMAL. If the examination was not performed on a particular body system, or at the subject level, then the value should be null, and NOT DONE should appear in PESTAT.	Exp	SDTMIG 4.1.5.1
PEORRESU	Original Units	Char		CRF or Derived	Variable Qualifier	Original units in which the data were collected. The unit for PEORRES.	Perm	
PESTRESC	Character Result/Finding in Std. Format	Char		Derived	Result Qualifier	If there are findings for a body system, then either the dictionary preferred term (if findings are coded using a dictionary) or PEORRES (if findings are not coded) should appear here. If PEORRES is null, PESTRESC should be null	Exp	SDTMIG 4.1.5.1
PESTRESN	Numeric Result/Finding in Standard Units	Num		Derived	Result Qualifier	Used for continuous or numeric results or findings in standard format; copied in numeric format from PESTRESC. PESTRESN should store all numeric test results or findings.	Exp	SDTMIG 4.1.5.1
PESTRESU	Standard Units	Char	*	Derived	Variable Qualifier	Standardized unit used for PESTRESC or PESTRESN.	Exp	
PESEV	Severity/Intensity	Char	*	CRF	Record Qualifier	The severity or intensity of the exam finding. PESEV is only used to qualify the results of an observation -- it should not be used to record the answer to the question posed in PETEST.. Example: "MODERATE" could be used to describe a PEORRES of "ACNE".	Perm	
PESTAT	Examination Status	Char	**NOT DONE	CRF or Derived	Record Qualifier	Used to indicate exam not done. Should be null if a result exists in ORRES.	Perm	SDTMIG 4.1.5.1
PEREASND	Reason Not Examined	Char		CRF or Derived	Record Qualifier	Describes why an examination was not performed or why a body system was not examined. Example: SUBJECT REFUSED. Used in conjunction with STAT when value is NOT DONE.	Perm	

4. Table 1: PE: Highly Reco

	CRF Data Collection Fields	SDTM Variable Name (*SDTM Core*)					
1	Body System Examined	PETEST *(required)*					
2	Examination Result	PEORRES *(expected)* PESTAT *(permissible)*				Abnormal). If the examination is not performed or not required select Not Done.	Done. Sites should be directed to complete overall assessment for each exam category/body system listed. If examination result is Normal then the value in PEORRES should be NORMAL. If the examination is not done, then the value in PEORRES should be Null and the value in PESTAT should be Not Done. If the exam result is abnormal, then the value of PEORRES should contain the text of the abnormal findings.
3	Abnormal Findings	PEORRES *(expected)*	PEDESC	Text description of any abnormal findings		Record all abnormal findings for the given body system in the space provided.	Text entered under abnormal findings (PEDESC) should be mapped to PEORRES.

5. Table 2: PE: Recommended/Conditional/Optional Data Collection Fields

	CRF Data Collection Fields	SDTM Variable Name (*SDTM Core*)	CDASH Data Collection Field Name	Definition	Applicable Regulations	Instruction to Clinical Site	Implementation/Rationale
1	Sponsor ID (O)	PESPID *(permissible)*		Sponsor defined reference number		N/A	May be pre-printed on the CRF as an explicit line identifier or defined in the sponsor's operational database.
2	Date of Examination (C)	PEDTC *(expected)*	PEDT	Date of examination		Record complete date of examination, day, month and year	The date of examination may be derived from the date of visit and therefore there will not be a separate field. Date and Time should be converted to ISO 8601 format and mapped to PEDTC
3	Time of Examination (O)	PEDTC *(expected)*	PETM	Time of examination		Record military time of examination	Date and Time should be converted to ISO 8601 format and mapped to PEDTC
4	Overall Examination Status (O)	PESTAT *(permissible)*	PEDONE	Used to indicate if exam was		Record whether or not the physical examination was	A subject/page level question can be used asking the site if the physical

6. Table 3: PE: Variable Definition Tables (Alternate Proposal)

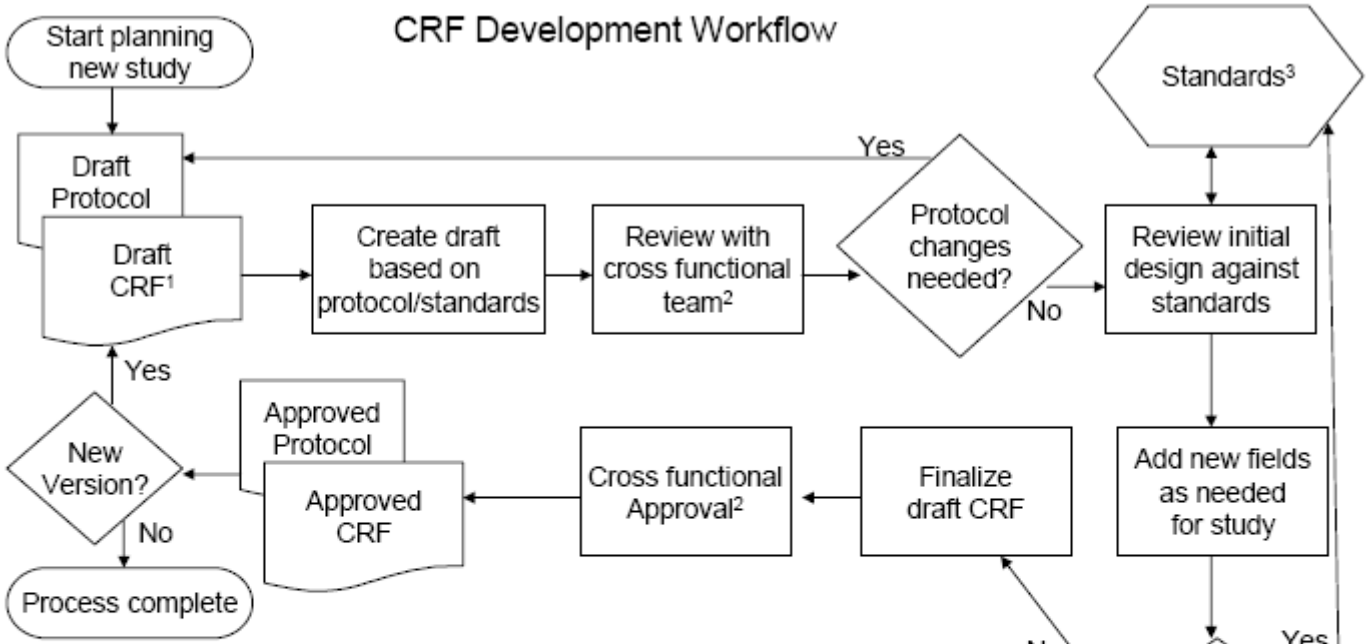
	CRF Data Collection Fields	SDTM Variable Name (*SDTM Core*)	CDASH Data Collection Field Name	Definition	Applicable Regulations	Instruction to Clinical Site	Implementation/Rationale
1	Overall Examination Status (O)	PESTAT *(permissible)*	PEDONE	Used to indicate if exam was performed at this visit.		Record whether or not the physical examination was performed.	BASELINE: If Yes, CRF and CRF Instructions will direct to report all abnormal findings/conditions on appropriate CRF (e.g. Medical History, Baseline Findings, and Adverse Events). POST-BASELINE: If Yes *and* abnormality new or worsened, CRF and CRF Instructions will direct to capture all changes on appropriate CRF (e.g. Post-baseline Assessment, Adverse Events).
2	Date of Examination (C)	PEDTC *(expected)*	PEDT	Date of examination		Record complete date of examination, day, month and year	The date of examination may be derived from the date of visit and therefore there will not be a separate field. Date and Time should be converted to ISO 8601 format and mapped to PEDTC
3	Time of Examination (O)	PEDTC *(expected)*	PETM	Time of examination		Record military time of examination	Date and Time should be converted to ISO 8601 format and mapped to PEDTC

3.5 Common Identifier Variables

The following variables apply across all of the data collection domains.

	CDASH CRF Label/Question	Clinical Database Variable Name *CDASH variables shaded*	Definition	Instruction to Clinical Site	Implementation / Rationale to Sponsors	CDASH Core
1	Protocol/Study Identifier	STUDYID (*required*)	Unique Identifier for a study within a submission.		This is typically pre-printed in the header of each CRF page. In an EDC study, this would be hard-coded into the study design. For data received from electronic data providers (i.e., central lab) this information should be provided to the e-data provider and verified during testing of the e-data receipts.	Highly Recommended
2	Site Identifier	SITEID (*required*)	Unique identifier for the site.	Record your clinical site's identifier as defined by the sponsor.	This is typically pre-printed in the header of each CRF page. In an EDC study, this should be pre-populated in the screens provided to the site. For data received from electronic data providers (i.e., central lab) this information should be provided to the e-data provider and verified during testing of the e-data receipts.	Highly Recommended
3	Subject	SUBJID (*required*)	Subject identifier	Record the identifier for the subject.	This is typically recorded in the header of each CRF page. In an EDC study the subject identifiers may be provided to the site using a pre-populated list in the system. For data received from electronic data providers (i.e., central lab) this information should be provided to the e-data provider and verified during testing of the e-data receipts. The subject identifier recorded in the CRF may be combined with other identifiers to produce the SUBJID, or may map directly to the SUBJID.	Highly Recommended
4	Investigator	INVID (*permissible*)	Investigator identifier	Record the sponsor defined identifier for your site investigator.	Study level – Not needed if SITEID is equivalent to INVID.	Optional

References: SCDM's GCDMP (v.4 2005) and GlaxoSmithKline CRF Principles



High Level Overview of CRF Development Best Practices:

- [1]Develop along with the Protocol (and SAP if available)
- [2]Develop as a cross-functional team
  - All data for analysis collected?
  - Possible to collect this way at the site?
  - Sufficient data to address safety?
- [3]Develop and/or enforce standards

3.4 FAQs on Best Practices for Creating CRF Content and Structure

Ref	Question	CRF Type	Best Practice Recommendation	Rationale
1	Should "Yes/No" questions be preferred over "Tick all that apply" questions?	Paper and electronic	• If an assessment can have composite responses (e.g. presence or absence of two or more symptoms), 'Yes/No' questions for each component response (e.g. symptom) are preferred to 'Tick all that apply' questions. • Exceptions to this recommendation might include assessments where the majority of options would be answered 'No'. An example would be the collection of ECG abnormality data where approximately 45 abnormalities may be listed but only a few will apply.	• Yes/No questions provide a definite answer. The absence of a response is ambiguous as it can mean "no" or that the response is missing. • 'Tick all that apply' questions are occasionally needed where the number of options is high.
2	Should there be a standard order for YES/NO response boxes and other standardized lists?	Paper and electronic	• It is recommended that a consistent order of Yes/No responses be used.	• A standard order of Yes/No response boxes facilitates the use of the CRF • Presenting Yes/No responses in a standard order could reduce bias. Add some wording to say it is "one tool" that can be used to reduce bias, but questions should also be carefully worded so they don't introduce bias or lead the investigator to a desired response.
3	What date format should be used for subject and site completed CRF data?	Paper and electronic	• CDASH is recommending an unambiguous date format. • For paper CRFs, or electronic studies in which the date is manually entered, CDASH recommends the format of DD-MON-YYYY for all date collection fields (whether the components are collected as a group or as separate components of day, month and year). • For non-English study data, use a character-based month abbreviation that is recognized in that language. • For electronic data capture, the user may be able to select a date from a calendar, and this would also meet the requirement for an unambiguous date.	• Using the international date format (DD-MON-YYYY) will provide unambiguous dates that will be as the same date by anyone who reads them. For example, the date 06/08/02, can be interpreted as June 8, 2002 or August 6, 2002. • Note: If subject-completed CRF pages are translated into a local language, the international date may make it easier to translate the documents. • Dates are collected in a user-friendly format and then converted to the ISO 8601 format for submission.

Ref	Question	CRF Type	Best Practice Recommendation	Rationale
4	What time format should be used for subject and site completed CRF data?	Paper and electronic	• CDASH recommends the use of a 24 hour clock using the HH:MM:SS format for recording times. 00:00:00 would indicate midnight and start the new date.	• As many of the HH:MM:SS elements should be used as are needed for a particular field. • Subject completed times may be recorded using a 12 hour clock and an A.M. or P.M. designation. The time would then be transformed to a 24 hour clock in the system. • Times are collected in a user-friendly format and then converted to the ISO 8601 format for submission.
5	Should calculated data items be recorded on the CRF?	Paper and electronic	• Calculated fields should not typically be recorded within the CRF when the raw data on which the calculation is based are recorded in the CRF. • An exception is when a treatment and/or study conduct decision needs to be made on those calculations. In those cases it may be useful for the calculated field to be recorded within the CRF. • It may also be useful to provide the site a step-by-step worksheet to record this data.	• Data items which can be calculated from other data captured within the CRF are more accurately reported if they are calculated programmatically in-house using validated algorithms. • Capturing both the source data items and the calculated field would be a duplication of data. • If the calculated field is used to make a treatment and/or study conduct decision, the results of the calculation would be required on the CRF to explain the decision made.
6	Should all data collected on CRFs be databased?	Paper	• Data that are collected on CRFs should usually be databased. • If data are not required for reporting or analysis, but that collecting the data aids the investigator or monitor, it is recommended that data be collected on a worksheet. Worksheets used at the investigator's site are not typically brought in-house and will not subsequently be databased. (examples would be an entry criteria worksheet, or a dose titration worksheet.) • Some data points are collected to facilitate data cleaning, and are not used for reporting or analysis. • Some fields, such as Investigator's Signature, can be verified by the data entry staff, but cannot actually be databased.	• Although the data recorded on worksheets are supporting documentation for key information collected elsewhere in the CRF, these data do not add value to the key information collected and are deemed redundant. • All such worksheets should be considered source documents or monitoring tools, and should be maintained at the site with the study files.

Ref	Question	CRF Type	Best Practice Recommendation	Rationale
7	Should "Was assessment x performed?" questions be collected and/or databased? And Should "Yes/No" exam completed be preferred over "Check if not done" questions?	Paper and electronic	• The database should contain an indication that an assessment was not performed. The mechanism for this may be different from system to system, or from paper to EDC. • In some cases this might be a "Yes/No – assessment completed" question, or "check if not done" box; in others it might be a blank flag or list of values to indicate why data are missing.	• This will provide a definitive indicator to both clinical and statistical programmers of why a data field has missing data. • This will prevent unnecessary data queries to clarify whether an assessment has been performed.
8	Should free text be an option for a response to a specific question? *(Also refer to the Comments Domain for additional information.)*	Paper and electronic	• The general recommendation from CDASH is that the collection of free text comments and general comments pages should be discouraged. Collection of free text should be limited to cases of specific safety or therapeutic need in reporting or analysis, such as Adverse Events, Concomitant Medications or Medical History. • CDASH recommends that questions be specific and clear, rather than open-ended. Instead of free text, or solicited comments fields, CDASH recommends a thorough review of the CRF by the protocol development team to maximize the use of pre-defined lists of responses.	• The collection and processing of free text requires significant resources, and is of limited use when analyzing and reporting clinical data. • Sites may enter data into free text fields that should be recorded elsewhere. • Entering text from these fields into the database is time consuming for data entry and requires Data Management resources to review the text for safety information and inconsistencies with other recorded data.
9	Should data be pre-populated in the CRF?	Paper or electronic	• Pre-printing or pre-populating any data in the CRF is discouraged.	• The CRF should be used as a tool to collect unknown study data. • In general, data should be collected and recorded by the site, not pre-populated.
10	Should location of measurement (e.g., oral temperature, blood pressure from right arm, etc.) be collected for each assessment?	Paper and electronic	• Location data should be collected only when multiple possibilities are present, and the location is required to make a meaningful analysis of the data (e.g. a comparison of blood pressures collected supine, sitting or standing)..	• Location options are only used when the protocol specifies.
11	Should sites be given guidance on how to record verbatim terms for adverse events, concomitant medications or medical history in the CRF?	Paper and electronic	• CDASH recommends that training be provided to the sites so they provide the required information in a reported term to enable meaningful coding. • CDASH recommends not providing actual coding dictionaries to the site for adverse events, concomitant medications or medical history reported terms, as this may bias responses and/or result in inconsistent coding.	• Providing guidance to the site on how the coding dictionaries will be used, and on the importance of clearly associating related terms (e.g., concomitant medications that are given for an adverse event) will facilitate the data verification process, reduce bias and facilitate coding.

- CDASH is focused on the development of consensus-based ‘content standards’--specifically: element name, definition, and metadata -- for a basic set of global data collection fields based on the CDISC SDTM model
- The initial scope of the project is focused on the ‘safety data domains’ to support clinical and medical research and safety reporting
- CDASH est différent de SDTM (notamment pour les variables dérivées) – mais conservation du niveau de typage des informations (required, expected, permissible…)

- CDISC held the CDASH project kick-off meeting in October 2006 to initiate the first three project streams (sub-groups),
- and to organize interested volunteers into working groups aligned with these project streams.
  - A total of 9 streams have been initiated and working since then.
  - They have developed initial consensus versions for the following safety data domains:

- adverse events (AE),
- concomitant medications (CM),
- Demographics (DM),
- subject characteristics (SC),
- inclusion/exclusion criteria (IE),
- medical history (MH),
- substance use (SU),
- physical exam (PE),
- vital signs (VS),
- drug accountability (DA),
- Exposure (EX),
- Comments (CO),
- protocol deviation (DV),
- disposition/end of study (DS),
- Lab (LB),
- and ECG (EG).

Conclusion

- One key goal of this initiative is to facilitate the participation of investigators and investigative site personnel in clinical trials by allowing them to enter data in a common format across trials.
- This harmonization will ensure that there is an integrated flow of data from site through submission and warehousing/archive

=> Illustration of CDASH standardization into ODM