SDTM aCRF Guideline

-

Guideline for SDTM annotations in Case Report Forms

Summary and Recommendations for Best Practice

© 2020 by Authors, Published under Creative Commons License (CC BY-NC-ND) 4.0

Version	V1.0	
Date	2020-11-20	
Authors	Task Lead:	Markus Stoll
	Co-Lead	Stefanie Sturm (mainanalytics GmbH)
	Co-Authors:	Christina Paul
		Eva Bervar (Bayer AG)
		Michael Schmitz (Bayer AG)
		Petra Rein
		Stefan Bordasch (AbbVie Deutschland GmbH & Co. KG)
		Torsten Petsching (Boehringer Ingelheim Pharma GmbH & Co. KG)

Table of Content

1. Introduction

Background

The annotated Case Report Form (aCRF, acrf) links the data collection fields used to capture study data to the corresponding variables in the study database. For every user of the study data the aCRF is a critical part of the study documentation. It enables the user to understand how the study data were collected and to trace back from study analysis results to the origin where it was collected.

As such it is also an integral part of study data submissions to authorities and needs to follow associated guidelines and standards. It is obvious that making the aCRF submission ready right from the beginning will facilitate submission preparation and will save time and costs.

By searching the internet there are only a few guidelines available by authorities but a broad number of recommendations available from different organizations that provide varying levels of details on how an aCRF should look like, on the format of the document and any naming conventions.

Discussions in the German Speaking CDISC User Network (CDISC DACH/DEUN UN) revealed that a consolidated and comprehensive guidance would be beneficial.

In July 2017 a view volunteers of the CDISC DACH UG built a sub team to develop a supporting Guidance for creating submission ready aCRFs, based on best practice and public available regulatory documentation in the web. A poster and a paper about the preliminary results were presented at the PhUSE EU Connect 2018 Conference in Frankfurt, Germany.

Purpose

The authors of this guideline try to collect and consolidate the available technical requirements, recommendations, formats and technical prerequisites. In addition, best practices and examples are included on how to populate annotations on a CRF (paper or electronic) to make it submission ready.

Disclaimer

The authors of this guideline do not provide instructions on how SDTM should be interpreted in general.

The content in this document is result of the authors understanding or interpretation of the available information listed in this document and the authors long term experience in the pharmaceutical industry for the last 20 years.

The statements, examples, recommendations and references contained in this document have not been reviewed, authorized or approved by any official body or authority and should therefore always be considered as the authors' non-binding and private opinions/recommendations. The authors make no warranties that the information or references contained in this document are accurate or complete, and therefore disclaim any liability for any damages or disadvantages that may arise in connection with the use or distribution of this information.

At least some of the sources referenced in this document do change on a regular base (e.g. FDA - Technical Conformance Guide). The recommendations based in this document are based on the document versions specified. It is explicitly recommended to check for updates on a regular base.

Identified Main Sources

The following sources have been identified and reviewed by the authors of this guideline for information or submission relevant annotation rules. In all identified official sources only a few general recommendations on SDTM annotations were available. There are almost no details or clear specifications provided. Submission relevant prerequisites or requirements were only identified in the underlined sources below. These can be found / looked up in chapter 2 (Prerequisites / Requirements).

- CDISC Web Page
  - QRS Rules
  - SDTM-MSG (Metadata Submission Guideline for SDTM)
- FDA
  - Study Data Specifications
  - Portable Document Format Specification
  - Study Data Technical Conformance Guide
- PMDA
- PhUSE
- XML ODM Team

The following sources have been checked, but there was no information found regarding SDTM annotated CRFs:

- TMF Uni Münster
- EMEA

2. Prerequisites / Requirements

FDA eCTD Portable Document Format (PDF) Specifications (v4.1, 21 Sep. 2016)

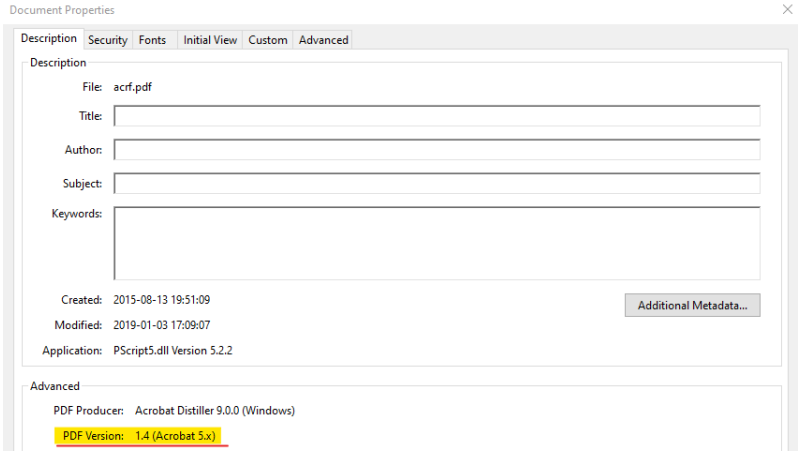
PURPOSE

These specifications are for the creation of documents in Portable Document Format (PDF) for submission to CDER or CBER, that align with ICH M2 recommendations[1] and that are in a format that the receiving Center currently supports. For purposes of this document, “supports” means the receiving Center has established processes and technological infrastructure to support the receipt, processing, review and archiving of files in the specified standard format. PDF is an open, published format created by Adobe Systems Incorporated (http://www.adobe.com). Software from a variety of sources can be used to create files in the PDF format.

VERSION

PDF versions 1.4 through 1.7, PDF/A-1 and PDF/A-2 are acceptable for documents. Submitted PDF files should be readable by Adobe Acrobat X, should not require additional software or plugins to be read and navigated, and should be text searchable. If plug-ins are used during the creation of a PDF document, prior to submitting the document, ensure that a plug-in is not needed for review or archive.

Check the version of the PDF document in Adobe Acrobat under:
File -> Properties -> Description Tab -> PDF Version



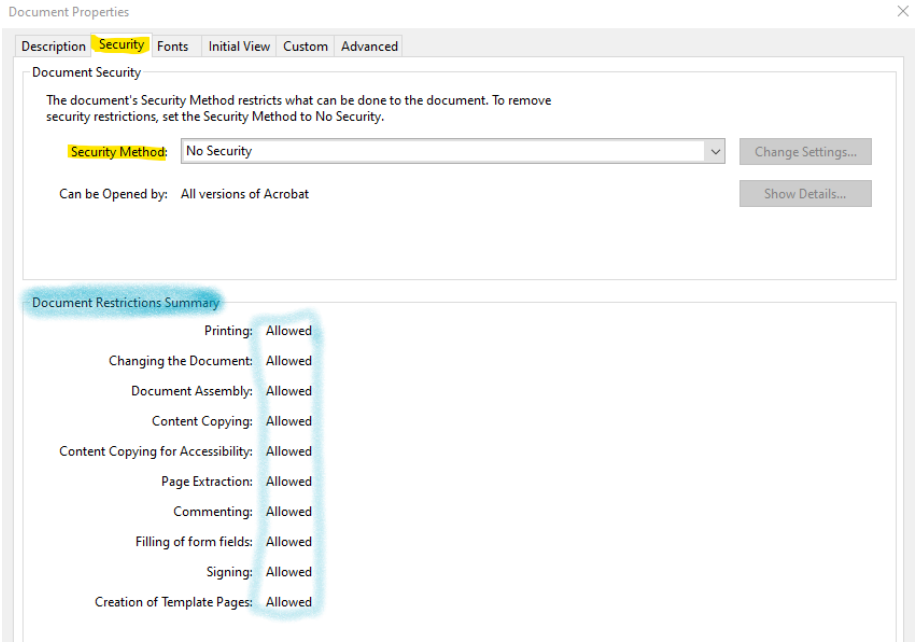
[1] ICH, "M2 Recommendations & Technical References): ICH has recommended several file formats for the exchange of regulatory documents

PDF files must not contain JavaScript; dynamic content which can include audio, video or special effects and animations; attachments or 3D content[2]. Do not include PDF annotations in documents[2]. Ensure that all hypertext links in documents remain active after conversion to PDF/A.

SECURITY

Do not activate security settings or password protection. The integrity of the submitted files is maintained through Agency security and archival processes. A copy of the files, generated from the submitted files, will be provided to the reviewer. The reviewer should be able to print, select text and graphics, and make changes to text, notes and form fields using the provided copy. FDA Forms downloaded from the FDA Forms website contain security settings that prevent changing the documents. These forms should be submitted as provided, with no additional security added and without removing the provided security settings.

On the *Document Properties > Security* tab check there is no *Security Method* selected and the *Document Restrictions Summary* view shows no restrictions are applied.



[2] Exception noted in: FDA, "Portable Document Format (PDF) Specifications" (v4.1, 21 Sep. 2016): “Special Considerations for Promotional Labeling and Advertising Material”, p.10 or see here in this document

FONTS

Fully embed all non-standard fonts. PDF viewing software automatically substitutes a font to display text if the font used to create the text is unavailable on the reviewer's computer. In some cases, font substitution can occur even when the fonts are available. For example, Helvetica or Times are substituted even if available on the reviewer's computer. Font substitution can affect a document's appearance and structure, and in some cases, it can affect the information conveyed by a document.

Font availability to the reviewer is ensured if all non-standard fonts are fully embedded. When fonts are embedded, all characters for the font should be included, not just a subset of the fonts being used in the document. Inspect documents to make sure all non-standard fonts are fully embedded prior to submission.

Font embedding does not always solve the problems that occur when a reviewer tries to copy and paste text from a PDF document into another software format. If the font is not available on the reviewer's computer, font substitution results, even if the fonts are embedded. This problem is avoided if the fonts are restricted to the standard fonts listed in Table 1.

Table 1 - List of Standard Fonts

Font type	Font name		Font type	Font name
Serif	Times New Roman		Non Proportional	Courier New
	Times New Roman Italic			Courier New Italic
	Times New Roman Bold			Courier New Bold
	Times New Roman Bold Italic			Courier New Bold Italic
Sans Serif	Arial		Other	Symbol
	Arial Italic			Zapf Dingbats
	Arial Bold			
	Arial Bold Italic			

Use font sizes ranging from 9 to 12 point[3]. Times New Roman 12-point font is recommended for narrative text. When choosing a point size for tables, a balance should be made between providing sufficient information on a single page that may facilitate data comparisons while still achieving a point size that remains legible. Generally, point sizes 9-10 are recommended for tables; smaller point sizes should be avoided. Ten-point fonts are recommended for footnotes.

[3] Exception noted in: FDA, "Portable Document Format (PDF) Specifications" (v4.1, 21 Sep. 2016): "Special Considerations for Promotional Labeling and Advertising Material", p.10 or see here in this document

When creating documents which include scanned images, ensure that any resizing of the image does not reduce the effective font size below the recommended size.

Black is the recommended font color[3] except that blue can be used for hypertext links. Light colors do not print well on grayscale printers. Any colors used should be tested prior to submission by printing sample pages from the document using a grayscale printer.

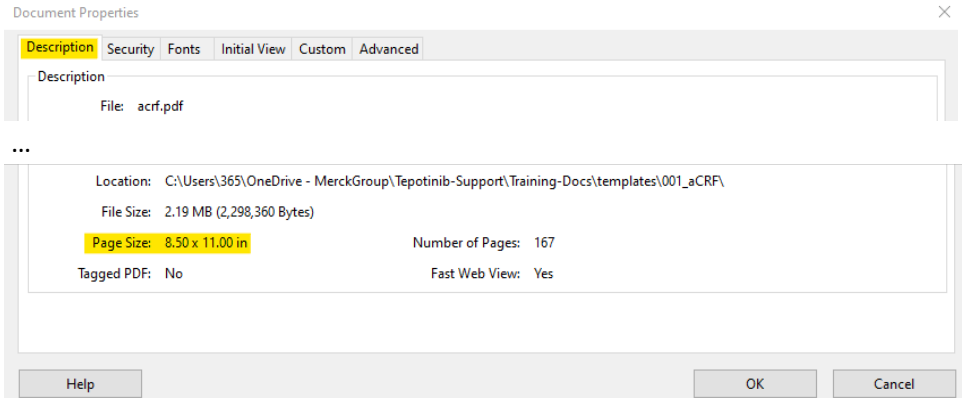
PAGE ORIENTATION

Save the page orientation for proper viewing and printing within the document. Proper page orientation eliminates the need for reviewers to rotate pages. For example, setting page orientation of landscape pages to landscape prior to saving the PDF document in final form ensures a correct page presentation.

PAGE SIZE AND MARGINS

Set up the print area for pages to fit on a sheet of paper that is 8.5 inches by 11 inches. A margin of at least 3/4 of an inch on the left side of page avoids obscuring information when pages are subsequently printed and bound. Setting the margin for at least 3/8 of an inch on the other sides is sufficient. For pages in landscape orientation, a margin of 3/4 of an inch at the top allows more information to be displayed legibly on the page. Header and footer information should not invade the specified margins (i.e., header and footer information should not appear within 3/8 of an inch of the edge of the 8.5 by 11 inch page), so the text will not be lost upon printing or being bound. These margins allow printing on A4 as well. Oversized documents (e.g., CAD drawings or other specialized documents) and promotional materials submitted in PDF format should be created according to their actual page size.

In Adobe Acrobat you can check for the Page Size under
Menu > File > Properties > Description Tab > Page Size



SOURCE OF ELECTRONIC DOCUMENTS

Avoid image-based PDF files whenever possible. PDF documents produced by scanning paper documents usually have poorer image resolution than PDF documents produced from electronic source documents such as word processing files. Scanned documents are generally more difficult to read and do not allow the reviewer to search or copy and paste text for editing in other documents. If scanned files must be submitted, they should be made text searchable where possible. If optical character recognition software is used, verify that imaged text is converted completely and accurately.

METHODS FOR CREATING PDF DOCUMENTS AND IMAGES

Use the dpi settings in Table 2 for scanning documents. Scanned documents scanned at a resolution of 300 dots per inch (dpi) ensure that the pages of the document are legible both on the computer screen and when printed and, at the same time, minimizes the file size. The use of grayscale and color significantly increases the file size and should be used only when these features improve the reviewability of the material. After scanning, avoid resampling to a lower resolution. A captured image should not be subjected to non-uniform scaling (i.e., sizing). See the following table for resolutions for various images.

Table 2 - Scanning Resolution

Document type	Resolution
Handwritten notes	300 dpi (black ink)
Plotter output graphics	300 dpi
Photographs – black and white	600 dpi (8 bit gray scale)
Photographs – color	600 dpi (24 bit RGB)
Gels and karyotypes	600 dpi (8 bit grayscale depth)
High pressure liquid chromatography	300 dpi

IMAGE COMPRESSION TO REDUCE FILE SIZE

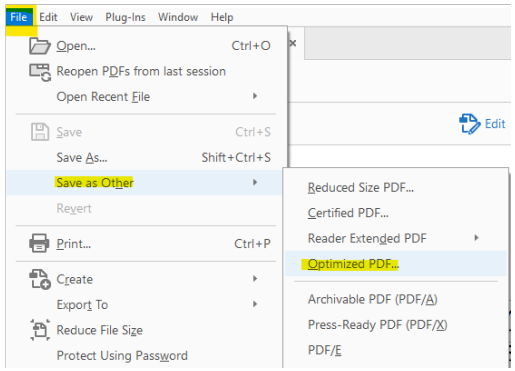
Compress files using either Zip/Flate or CCITT Group 4. File compression is a method for reducing file size. Some methods of compression can result in loss of data and can introduce compression artifacts that affect the reviewability of the information. The following two methods provide lossless compression.

- Zip/Flate (one technique with two names) for lossless compression of color and grayscale images is specified in Internet RFC 1950 and RFC 1951.
- CCITT Group 4 Fax compression technique recommendations for lossless compression of black and white images is specified in T.6 (1988) - Facsimile coding schemes and coding control functions for Group 4 facsimile apparatus.

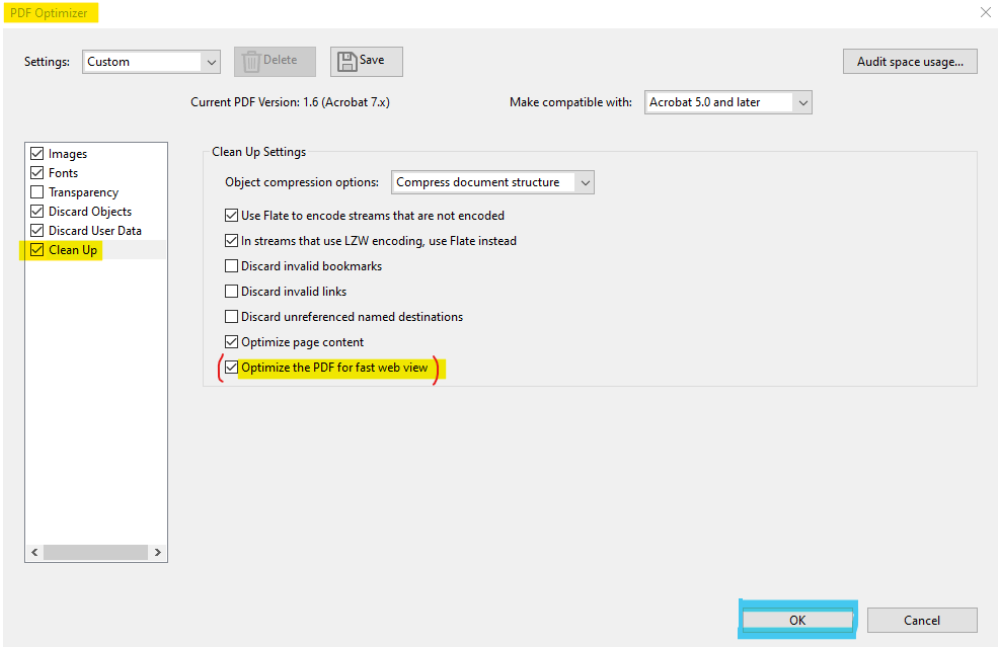
OPTIMIZE FOR FAST WEB VIEW AND OPENING SETTINGS

Create files from source documents using the “Optimize the PDF for fast web view” option to reduce file sizes and file opening times.

Select from the menu bar: File > Save as Other > Optimized PDF

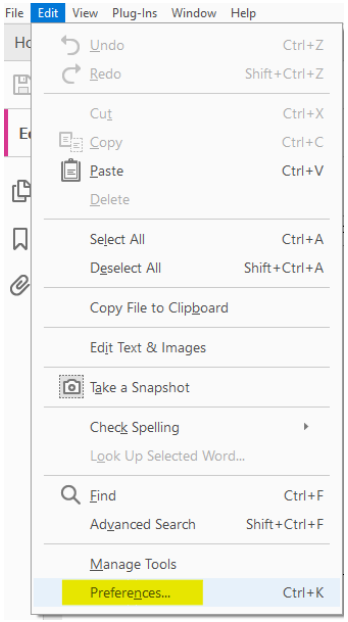


In the “PDF Optimizer” popup menu select the “Clean Up” submenu from the selection list on the left. In this submenu make sure to check the “Optimize the PDF for fast web view” checkbox on the bottom of the property list before saving via the “OK” button.

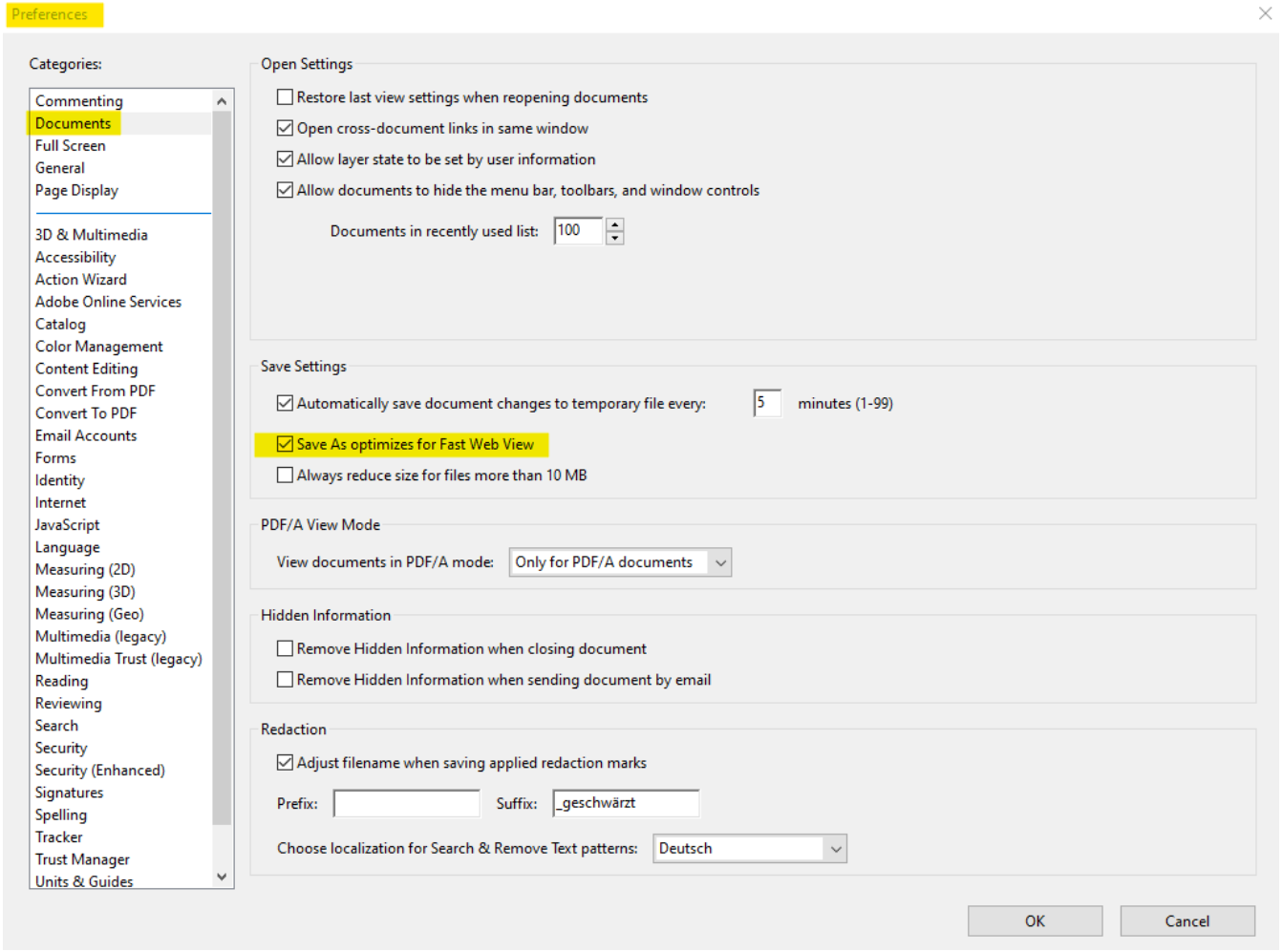


Activate “Fast Web View” compatibility in Adobe Acrobat Preferences:

Menu > Edit > Preferences…



… > Documents: activate tick-box “Save As optimizes for Fast Web View”



Set the correct PDF Opening Settings:

File > Properties > Initial View: Submenu “Layout and Magnification”

- Navigation Tab: Bookmarks Panel and Page
- Page Layout: Default
- Magnification Default
- Open to page: 1

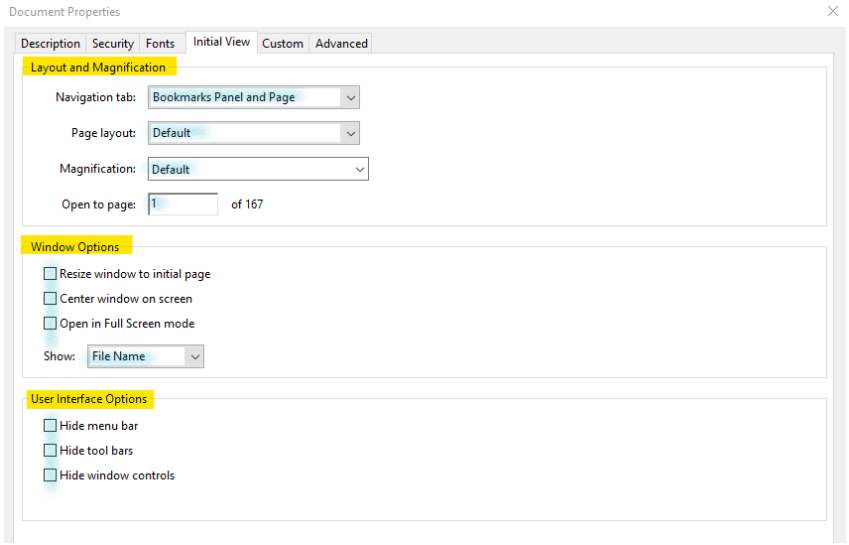
Submenu “Window Options”

- Checkboxes: Not ticked
- Show: File Name

Submenu “User Interface Options”

- Checkboxes: Not ticked

Check the Opening Settings under *Document Properties > Initial View*



Check the Fast Web View Settings under *Document Properties > Description*

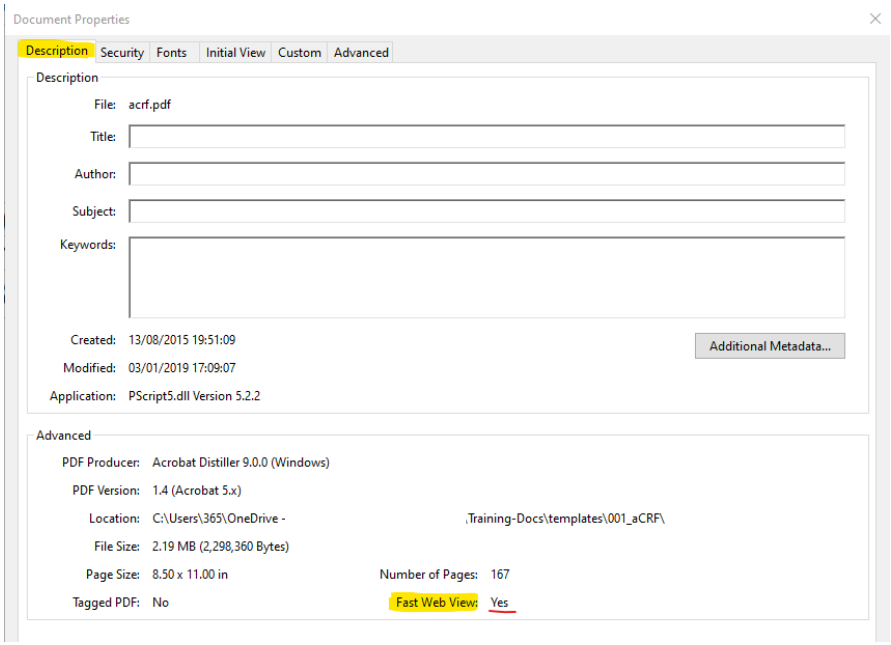


IMAGE COLOR MATCHING

Because color varies from monitor to monitor, it is difficult to ensure that the reviewer will see exactly the same color as in the actual image. However, for printing, there is more control over the color by using CMYK (Cyan, Magenta, Yellow, Black) color model as opposed to the RGB model. Pantone Matching using the color profile provided by CMYK ensure color consistency for printing. The International Color Consortium (ICC)[4] color profile specification is used when PDF documents are printed.

USE OF THUMBNAILS

PDF documents do not need embedded thumbnails.

DOCUMENT NAVIGATION

A table of contents (TOC), hypertext links and bookmarks provide essential navigation through PDF documents. Include a hypertext linked TOC and bookmarks in documents 5 pages or longer. Use hypertext links throughout the body of all documents to link to supporting annotations, related sections, references, appendices, tables or figures that are not located on the same page as the narrative text. Hypertext links in text can be designated by rectangles using thin lines or by blue text. A consistent method of designating links in a document avoids confusion. Hypertext links that open a file or document should be set to open the file or document in a new window. Using relative paths when creating hypertext links minimizes the loss of hyperlink functionality when submissions are loaded onto network servers; both absolute links that reference specific drives and links to root directories do not work once the submission is loaded.

The document TOC helps the reviewer navigate to the information of interest within the document that is not provided in the submission table of contents. For documents with a table of contents, provide bookmarks and hypertext links for each item listed in the table of contents including all tables, figures, publications, other references, and appendices that are essential for navigation through documents. The use of invisible rectangles and blue text in the table of contents for hypertext links avoids obscuring text. Other help for navigation includes a bookmark hierarchy identical to the table of contents; up to four levels deep in the hierarchy.

When creating bookmarks and hyperlinks, set the magnification setting to “Inherit Zoom” so the destination page displays at the same magnification level used in the primary document.

[4] ICC – International Color Consortium, see References

INITIAL VIEW SETTINGS

Set the Navigation Tab to open to “Bookmarks Panel and Page.” This sets the initial document view when the file is opened. If there are no bookmarks, set the Navigation Tab to “Page Only.” Page Layout and Magnification should be set to “Default.”

PAGE NUMBERING

In general, it is easier to navigate through an electronic document if the page numbers for the document and the PDF file are the same, with the initial page of the document numbered as page one. There is an exception when a document is split because of its size and the second or subsequent file is numbered consecutively to that of the first or preceding file.

NAMING PDF FILES

Use lower case characters and avoid using special characters except hyphens and underscores in file names. Special characters to avoid include punctuation, spaces, or other non-alphanumeric symbols. The current FDA validation criteria and the ICH eCTD specification both provide additional guidance on allowable special characters in file names.

SPECIAL CONSIDERATIONS FOR PROMOTIONAL LABELING AND ADVERTISING MATERIAL

Promotional materials submitted in PDF format may need special consideration to ensure accurate representation of the actual image. PDF restrictions for font size, color, and annotations stated in this document are not applicable to these materials. Since color varies from monitor to monitor, it is difficult to ensure that the reviewer will see exactly the same color as in the actual image. Provide images at the highest resolution and depth practical. For photographs, the image should be obtained with a resolution of at least 600 dpi. Documents that are available only in paper should be scanned at resolutions that will ensure the pages are legible both on the computer monitor and when printed; at least 600 dpi is recommended. Promotional material should be submitted according to its actual size when practical. When an image size is altered, the original dimensions must be stated. Images of three-dimensional promotional pieces must show all sides and components.

Promotional materials submitted in PDF format may need special consideration to ensure accurate representation of functionality. For example, screenshots of websites submitted in PDF format should contain links that allow the reviewer to click on them to simulate navigation in the actual website. Dynamic content such as audio, video, special effects, animations, attachments or 3D content are permitted if embedded in the PDF and playable via Adobe Acrobat X without the need for plug-ins or other special software.

FDA, “Study Data Specifications” (V2.0, July 18, 2012)

Source: FDA, Study Data Specifications, Chapter 5.4 „Annotated Case Report Form“

5.4 Annotated case report form

Definition

This is a blank case report form with annotations that document the location of the data with the corresponding names of the datasets and the names of those variables included in the submitted datasets.

Specifications

The annotated CRF is a blank CRF that includes treatment assignment forms and maps each item on the CRF to the corresponding variables in the database. The annotated CRF should provide the variable names and coding for each CRF item included in the data tabulation datasets. All of the pages and each item in the CRF should be included. The sponsor should write *not entered in database* in all items where this applies. The annotated CRF should be provided as a PDF file. Name the file *blankcrf.pdf*[5].

FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019)

4.1.4.6 Annotated Case Report Form (aCRF) for SDTM

An Annotated Case Report Form (aCRF) is a PDF document that maps the clinical data collection fields used to capture subject data (electronic or paper) to the corresponding variables or discrete variable values contained within the SDTM datasets. Regardless of whether the clinical database is legacy or SDTM compliant, an aCRF should be submitted.

The aCRF should be provided as a PDF with the file name “acrf.pdf.” (Previously acrf.pdf was called blankcrf.pdf).

The SDTM Metadata Submission Guidelines should be used for additional information on annotated CRFs (See CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31).

[5] FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019): section 4.1.4.6, p.32 The naming of the file is changed to “acrf.pdf”. As the TCG is updated on a regular base and is more recent than the eCTD PDF Specifications it is best practice to use the file name “acrf.pdf” for the SDTM annotated case report form.

The aCRF should include treatment assignment forms, when applicable, and should map each variable on the CRF to the corresponding variables in the datasets (or database). The aCRF should include the variable names and coding for each CRF item.

When data are recorded on the CRF but are not submitted, the CRF should be annotated with the text "NOT SUBMITTED." There should be an explanation in the SDRG stating why data have not been submitted.

8.3.2.1 Traceability Issues with Legacy Data Conversion (Tech Conformance Guide)

FDA does not recommend a particular approach to legacy study data conversion, but rather explains the issues that should be addressed so that the converted data are traceable and adequate to support review.

Table 3 presents some of the issues that can be observed during a review when legacy study data are converted to SDTM and submitted with legacy analysis datasets.

Table 3 - Traceability Issues: Legacy Data Conversion to SDTM Only

1. Limited ability to determine location of collected CRF variables in the converted SDTM data unless the legacy aCRF is re-annotated.
2. Limited traceable path from SDTM to the legacy analysis data.
3. Limited ability to replicate/confirm legacy analysis datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
4. Limited ability to confirm derivation of intermediate analysis datasets or custom domains.
5. Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

Table 4 presents the issues when legacy study data and legacy analysis data are independently converted to SDTM and ADaM formats, respectively, rather than ADaM datasets being created directly from the SDTM datasets (converted from legacy study data).

Table 4 - Traceability Issues: Independent Legacy Data Conversion to SDTM and ADaM Issues

1. Limited ability to determine location of collected CRF variables in the converted SDTM data unless the legacy aCRF is re-annotated.
2. Limited traceable path from SDTM to the legacy analysis data.
3. Limited ability to replicate/confirm legacy analysis datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
4. Limited ability to confirm derivation of intermediate analysis datasets or custom domains.
5. Limited traceable path from SDTM to the ADaM datasets.
6. Limited ability to replicate ADaM datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
7. Limited traceable path from ADaM to the Tables, Figures and the Clinical Study Report (CSR).
8. Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

Table 5 presents the issues when legacy data are converted to SDTM and ADaM formats in sequence (i.e., converting legacy study data to SDTM and then creating ADaM from the SDTM). The key concern is the traceability from ADaM to the Tables, Figures and CSR.

Table 5 - Traceability Issues: Legacy Data Conversion to SDTM and ADaM in Sequence

1. Limited ability to determine location of collected CRF variables in the converted SDTM data unless the legacy aCRF is re-annotated.
2. Limited traceable path from SDTM to the legacy analysis data.
3. Limited ability to replicate/confirm legacy analysis datasets (i.e., analysis variable imputation or derived variables) using SDTM datasets.
4. Limited ability to confirm derivation of intermediate analysis datasets or custom domains.
5. Limited traceable path from ADaM to the Tables, Figures and the CSR.
6. Difficulty in understanding the source or derivation methods for imputed or derived variables in integrated/pooled data, supplemental qualifiers, and related records.

8.3.2.2 Legacy Data Conversion Plan and Report (Technical Conformance Guide)

Sponsors should evaluate the decision involved in converting previously collected non-standardized data (i.e., legacy study data) to standardized data (i.e., SDTM, and ADaM). Sponsors should provide the explanation and rationale for the study data conversion in the Reviewer's Guide. To mitigate traceability issues when converting legacy data, FDA recommends the following procedures:

1. Prepare and Submit a Legacy Data Conversion Plan and Report.
    a. The plan should describe the legacy data and the process intended for the conversion.
    b. The report should present the results of the conversions, issues encountered and resolved, and outstanding issues.
    c. The plan and report should be provided in the SDRG.
2. Provide an aCRF, for clinical data, that maps the legacy data elements.
    a. Sponsors should provide two separate CRF annotations, one based on the original legacy data, and the other based on the converted data (i.e., SDTM) when legacy datasets are submitted. The legacy CRF tabulation data should include all versions and all forms used in the study.
3. Record significant data issues, clarifications, explanations of traceability, and adjudications in the RG. For example, data were not collected or were collected using different/incompatible terminologies, or were collected but will not fit into, for example, SDTM format.

Legacy data (i.e., legacy aCRF, legacy tabulation data, and legacy analysis data) may be needed in addition to the converted data.

Other Recommendations and Best Practice

Pre-requisites for the underlying document(s) of the annotated CRF

- Final or approved Data Collection Tool (DCT)/CRF should be blank, which means no raw database annotations
- It should not contain any blank pages
- It must be complete with all pages/CRF modules
- It should include additional collection documents (e.g. questionnaires, diaries, lab print outs) if these data are NOT transcribed into the CRF
- It should contain only unique pages
- Amendments which lead to changes in the DCT/CRF should be included

3. Format

File Format and Naming Conventions

As stated in the FDA Study Data Technical Conformance Guide, the aCRF is a PDF document and should be provided with the file name ‘acrf.pdf’.

Page Numbering

Page numbering should be done on the final aCRF that includes the printable Table of Contents (TOC), if needed. For more details refer to section “5. Table of Contents (TOC)”.
When referring to a page number on the aCRF use the page as it is defined within the PDF document and not the original page number printed on a CRF page.

Annotations

According to the FDA Study Data Technical Conformance Guide the aCRF should map each data collection field to the corresponding variable or discrete variable value contained within the SDTM dataset. To meet this obligation annotations for SDTM domains and SDTM variables should be done on each CRF page.
For more details and examples refer to section “6. Annotations”

Annotated Comment/Instruction Boxes

During study conduct additional comment/instruction boxes that explain the SDTM mapping could be helpful for the creation of the SDTM datasets. But they need to be removed prior to the creation of the final aCRF for a submission.

4. Bookmarks

The annotated CRF is required to contain bookmarks.
According to the CDISC Metadata Submission Guideline (see References) bookmarks should be created in two ways (**dual bookmarking**):

- Chronological - by time points according to study event schedule with study-level bookmarks (e.g. AE) available last

- Alphabetical - by CRF topics or DCT (sequence used in EDC tool). Available bookmarks within a topic have to be ordered chronologically according to study event schedule, too.

Bookmarks have to reflect forms as many times as needed to give evidence how the data were collected; e.g. “Subject Status” is bookmarked in visits 1, 3, and 5;. all three bookmarks would link to the unique subject status page.
Magnification has to be set to “Inherit Zoom”.
Bookmarks have to be created at least in chronological order - this has to be confirmed by reviewer.

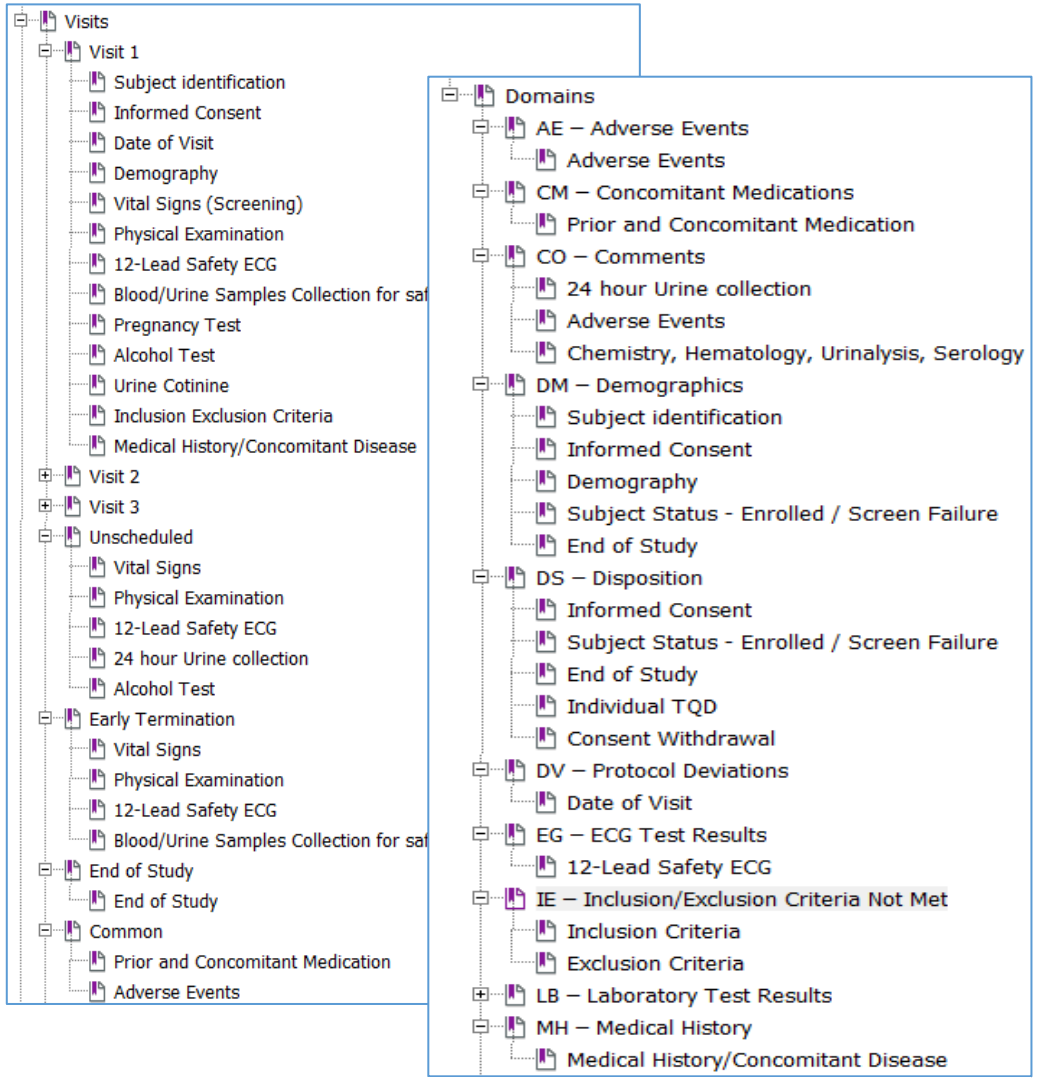


Figure 4-1 chronologically bookmarks by visits and alphabetically bookmarks by domains

5. Table of Content (TOC)

Bookmarks act as Table of Contents (TOC) for the reviewers. The aCRF has to have a separate TOC when more than 5 pages are available. TOC should be printable as well as hyperlinked to contents.
Several commercially available PDF tools create a TOC based on the bookmarks. However, it is also possible to create a TOC using word processing software when using 'SAVE AS' PDF with Options 'Include non-printing information' -> 'Headings' and 'Document structure tags for accessibility'.

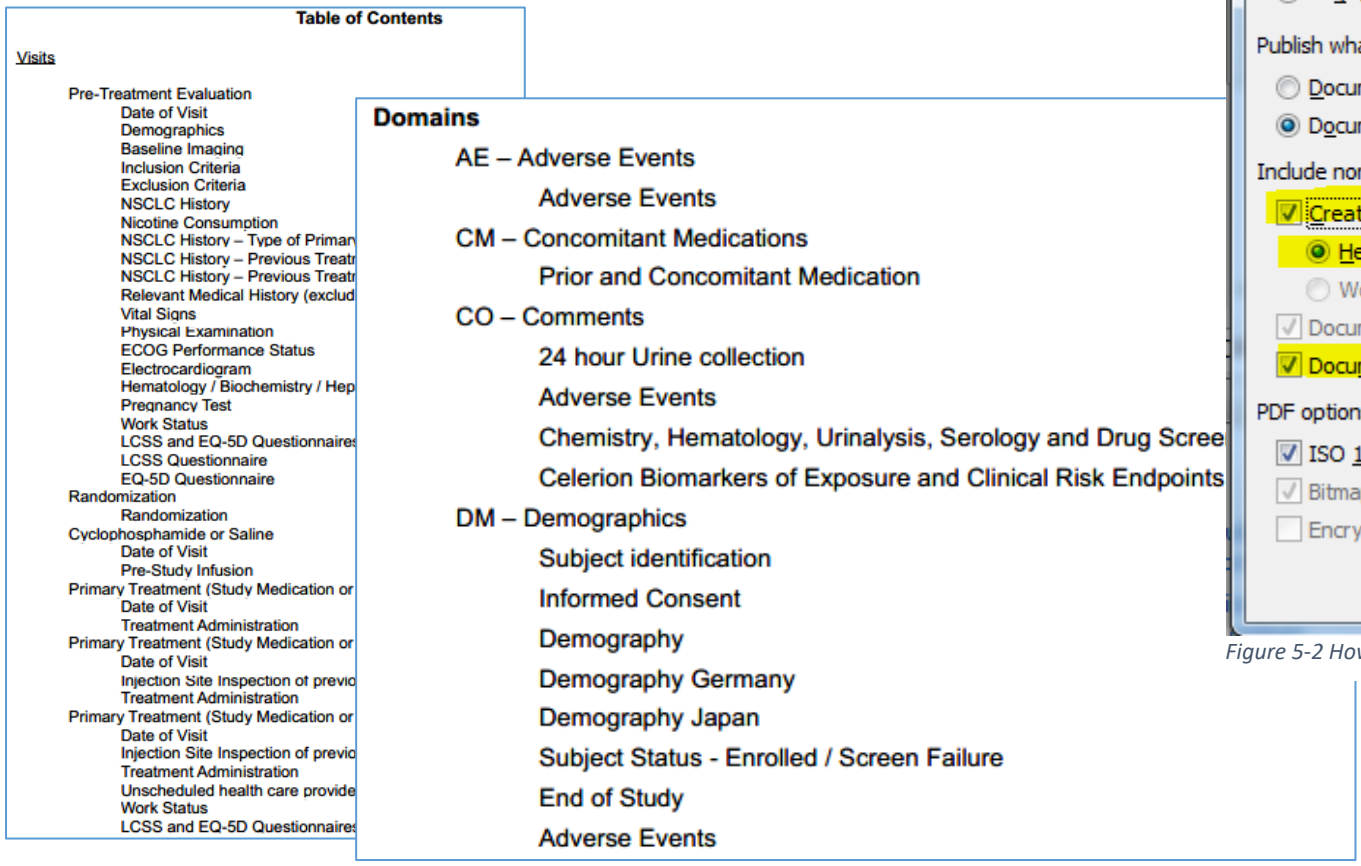


Figure 5-1 Table of Content by Visits and Domains

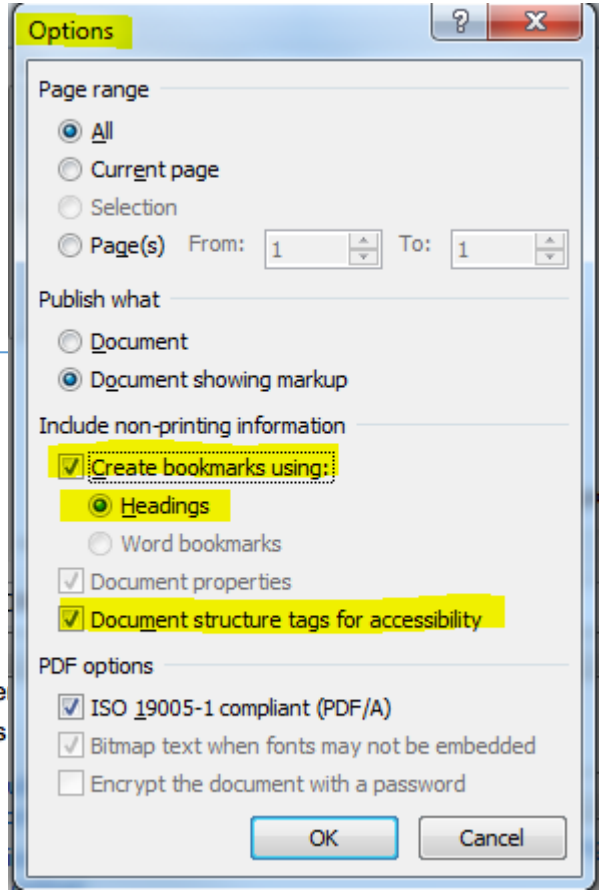


Figure 5-2 How to integrate TOC from word documents

6. Annotations

General Assumptions

An Annotated Case Report Form (aCRF) is a PDF document that maps the clinical data collection fields used to capture subject data (electronic or paper) to the corresponding variables or discrete variable values contained within the SDTM datasets. […] It should include the variable names and coding for each CRF item.[6]

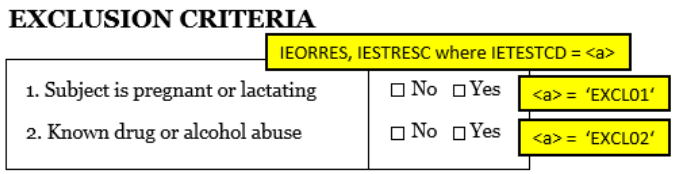
Sponsors may choose any available tool for creating annotations. Irrespective of the tool used, the annotations should be searchable, (i.e. text-based), to enhance the review process. Since the annotated CRF supports the review process, the annotations should reflect the data that are expected to be submitted within the SDTM datasets. In the event that data were intended to be collected for a variable, but none actually was, the annotated CRF will represent the data that would have been submitted had data been received. It is not necessary to re-annotate the acrf.pdf to indicate that data were not collected.[7]

Currently some sponsors include the entire casebook in the acrf.pdf (former blankcrf.pdf) while others include only unique forms. CDISC recommends to include and annotate unique forms only.[8] In any case, all included pages need to be annotated.

On each page of the annotated CRF, the version (number and/or date) of the annotation should be documented on each CRF page (e.g. at the bottom).

All text in the annotations that represent variable and domain names should be capitalized. If possible, the annotations should not obstruct any text on the CRF page. A sponsor may choose to resize the domain annotation based on the CRF layout.[9]

If it is necessary to abbreviate annotations due to insufficient space on the CRF page, this should be done in the following way using pointed brackets:



[6] FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019): section 4.1.4.6, p.32
[7] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1, p.17
[8] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1.1, p.17
[9] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1.2, p.18

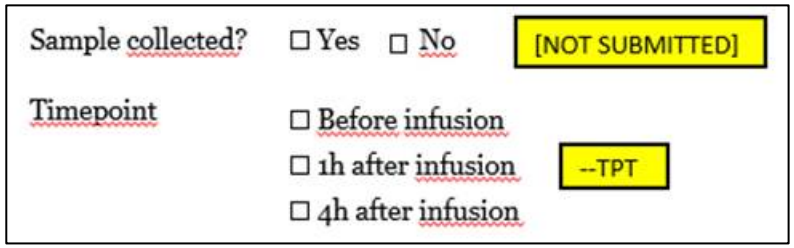
If repetitions of completely identical pages or modules are included in the aCRF, e.g. because of a different visit name or a different scheduled time point label, annotations for these pages/modules should not be repeated. In a minimum aCRF approach only unique pages would be included and if pages are used multiple times this would be reflected by linking them multiple times in the visit bookmark structure. If the sponsor decides to include all the CRF pages, including all repeating pages, a reference "ANNOTATIONS ON PAGE X" might be added to the repetitions, with X being the page number of the referenced page as displayed in the complete PDF file. Additionally, a hyperlink can be included to the respective annotated reference page to facilitate the navigation within the acrf.pdf. In case there are any differences in the annotation for a page or module, as for example different timepoint information or category which would end up in a smallest difference of the annotation, referencing is no longer possible and all annotations for the complete page or module have to be provided.

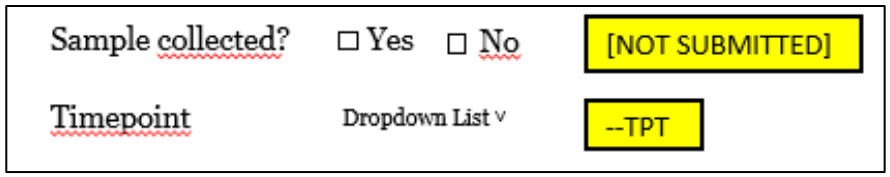
Carriage returns within the annotation field should be avoided. When a page is printed, all annotations must be included and readable.

Additional comments related to annotations within the CRF should be avoided, for example further explanations or reason for changes. Further comments or explanations should always be included in the define.xml and/or the cSDRG.

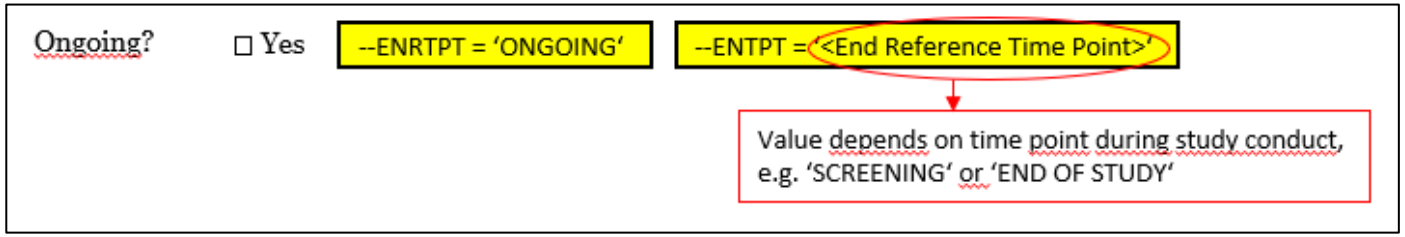
Examples:

A) Two examples where the timepoint annotation do not change if the module is used multiple times in the CRF (If re-used annotations could be referenced):





B) Example where the timepoint annotation will change if CRF module is used at different timepoints. All instances of the module would require a full annotation as the timepoint itself is part of the annotation and will change:



Domain Level

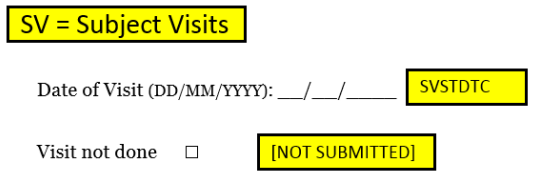
Location of annotations

Each domain that is represented on a CRF page should have its own annotation on the upper left side of the CRF page with the 2-letter domain code and domain name. To support a programmatic QC of the CRF annotations, we recommend to repeat/include Domain annotations on top of all consecutive CRF pages, in case a CRF module spans over multiple pages. Otherwise it can be very complex or even impossible to perform proper automated checks of the annotations. In this version of the guideline we do not focus on machine readability. However, this topic is planned to be included and discussed in a future version of this guideline.

Domain names rather than dataset names should be annotated. To distinguish the domain level annotations from the variable annotations a slightly larger font can be used for the domain annotations. If more than one domain exists on a page, each domain annotation, and all of its variables, should be color-coded. It is not necessary to continue the color scheme for a domain across CRF pages.[10]

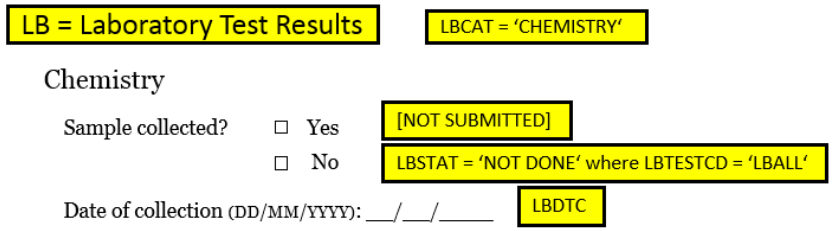
[10] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1.2, p.18

Example:



Whenever a --CAT or –SCAT variable value is applicable to an entire CRF module, the annotation should be done at the top of the page next to the annotation of the domain. Other variables, e.g. --METHOD, can also be added.

Example:



If the annotations on the CRF module are mapped to multiple domains a respective color coding could be used to correlate the –CAT and –SCAT annotation on top to their corresponding domain.

In case multiple –CAT or –SCAT annotations are required for a single domain on a CRF module the annotations should be placed as meaningful as possible on the CRF module to allow an easy correlation to the corresponding TESTCD annotations.

For each annotated variable a single annotation box should be used. There is only one exception: date variables.

Example: see section Date / Time fields.

SUPP--

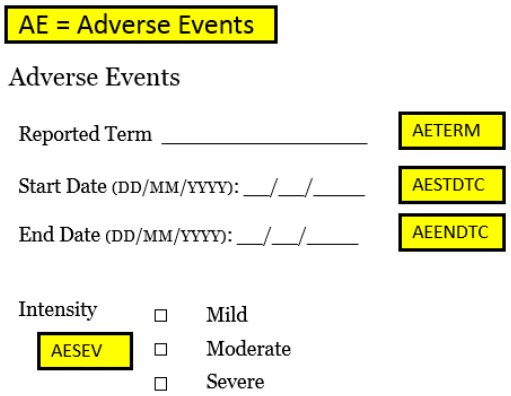
SUPPQUAL dataset names do not need to be annotated since SUPPQUAL variables are annotated as part of the main domain.[11] For examples, see chapter Element Level/SUPP--.

Element Level

General Assumptions

When a CRF item is mapped directly into the SDTM structure, the SDTM variable has to be annotated.

Example:



If the variable to be annotated is used for more than one domain it can be entered using -- as domain placeholder.

Example: see section Date / Time fields.

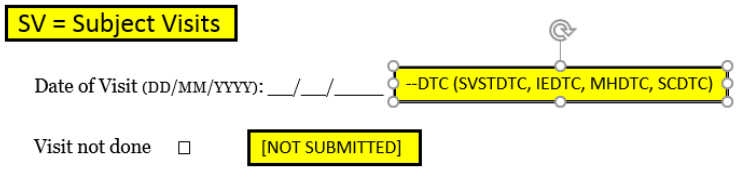
[11] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1.2, p.18

Date / Time fields

If the visit date collected as SVSTDTC is also the date of collection in other domains within this visit, the respective --DTC variables of the other domains belonging to this visit should be annotated as well.

The annotation of the domain code and name (see Domain Level – Location of annotations) should only reflect to SV domain for these multiple --DTC variables. A detailed description for the derivation of the respective --DTC variables can be provided in the define.xml/cSDRG.

Example:



In the define.xml, the origin of these --DTC variables should be listed as “CRF page x”, with “x” being the page of the respective CRF module, [12]even though the individual date variables were derived from the Date of Visit module and were not directly collected on the CRF.

--TEST/--TESTCD

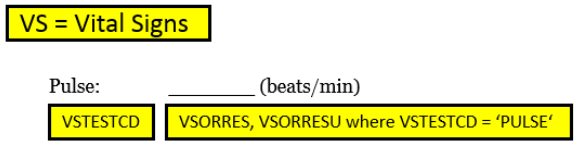
Summary:

Because of the vertical nature of the SDTM findings domains, it may be necessary to include the value of “--TESTCD” in the annotations. For example, annotating a field as simply VSORRES or VSORRESU is not sufficient; it is necessary to indicate the VSTESTCD to which the result or units applies.[13]

[12] CDISC.org, “define.xml v2.0” Specifications (Define-XML-2-0-Specification, 2014-04-24): Section 4.2.2.2, p.22-27 / 4.5, p.42-43 / 5.3.6.1.1, p.59 / 5.3.11.3, p.80
[13] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1.3, p.18

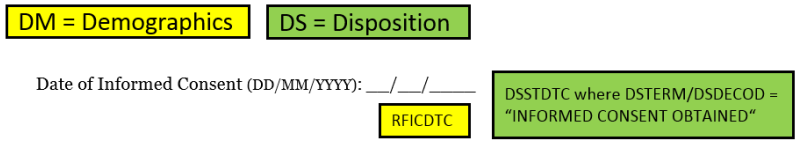
It is sufficient to annotate the respective --TESTCD; the link to the corresponding --TEST value will be described in the define.xml.

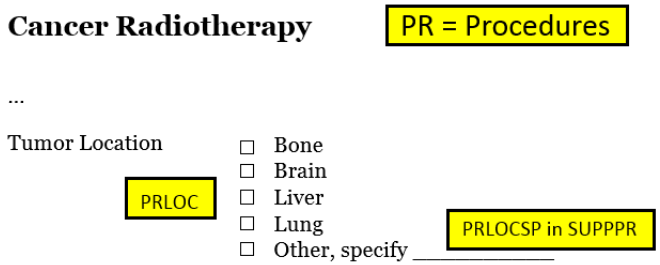
Example:



Results of medical items that are collected in one operational variable but have to be stored in different variables in SDTM have to be annotated appropriately to show the different locations of the content.

Example:





SUPP--

In general, variables used to support key analyses should not be represented in SUPPQUAL.[14]

When annotating SUPPQUAL variables, annotate the QNAM value and the SUPPQUAL dataset. The rationale for this approach is that the review tools join the supplemental values with the correct data row(s) from the parent domain and the reviewer needs to know only the variable name and that it originated in the supplemental dataset.[15]

Example:

RACEOTH in SUPPDM

If further specification of a variable is required, e.g. when an entry in a SUPPQUAL dataset links to a certain entry in the parent domain only, then a further specification in the annotation can be added:

Example:

DOSUOTH in SUPPCM where CMDOSU = "OTHER"

RHYTHOTH in SUPPEG where EGTESTCD = "RHYTHM" and EGORRES = "OTHER"

[14] FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019): section 4.1.1.3, p.22
[15] CDISC.org, “Study Data Tabulation Model Metadata Submission Guidelines” (SDTM-MSG v1.0, 2011-12-31): section 4.1.4, p.18

Another possibility for the annotation can be “SUPPxx.QVAL when QNAM=”abcd”.

Example: SUPPDM.QVAL when QNAM = “RACEOTH"

If multiple selections can be made for one data collection field on the CRF (e.g. due to a “Check all that apply” instruction on the CRF), the value for the respective variable within the parent domain should be set to “MULTIPLE” and SUPP-- should be used to store the individual responses.[16]

Example: ACN1-3 in SUPPAE where AEACN = "MULTIPLE"

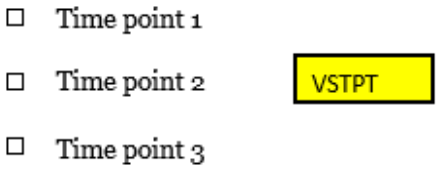
--GRPID

When the --GRPID variable is used within a domain, a short description how the --GRPID is used may be added to the aCRF. If the description is too complex, it should only be included in the define.xml/cSDRG.

--TPT, --SPID

If a single examination is repeated at multiple planned time points within a visit, the variable --TPT (planned time point name) is used to capture the information and to distinguish between the individual time points.

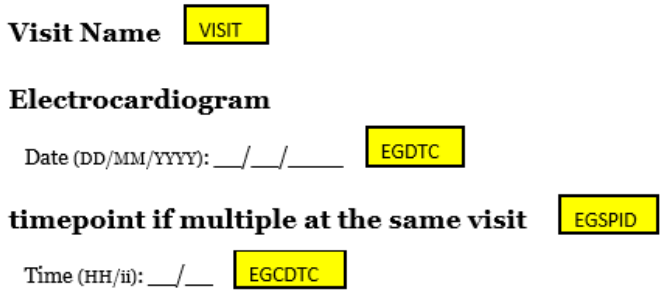
Example:



[16] CDISC.org, “Study Data Tabulation Model Implementation Guide” (SDTM-IG) (v3.2, Portfolio, 2013-11-26): Section 4.1.2.8.3, p.13

If repeating examinations are done for a single time point (e.g. multiple ECG measurements are performed in order to evaluate the mean results for these measurements) and the --TPT variable is not sufficient to capture the time point information, the sponsor-defined identifier --SPID variable may be used to distinguish the individual measurements that are done for a single time point.

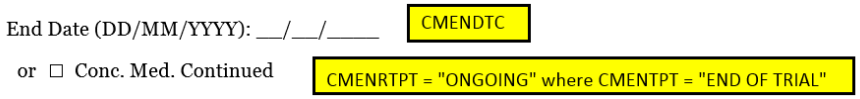
Example:



Ongoing

If “Ongoing” at a study defined reference period can be ticked on the CRF, this information should be stored in the variables --ENTPT and --ENRTPT. The “Ongoing” is annotated to --ENRTPT, while --ENTPT specifies the corresponding reference period.

Example:



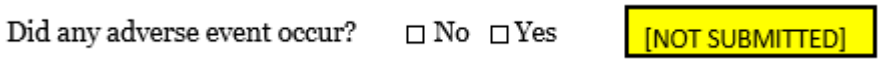
„NOT SUBMITTED" Annotations

All data collection fields on the CRF should be annotated. When data are recorded on the CRF but are not submitted, the CRF should be annotated with the text "NOT SUBMITTED."[17] No further SDTM variable annotation will be allocated to these fields.

For fields set to [NOT SUBMITTED], it should be checked with the responsible department, e.g. Biostatistics, that the respective information is not required for analysis.

[NOT SUBMITTED] can be assigned to a specific question, module or a complete page. If a complete CRF page is not submitted it is best practice to have a single annotation on the very top of the not submitted page without any further annotations. Instead of the general expression [NOT SUBMITTED] used for single or multiple data collection fields you might also use [PAGE NOT SUBMITTED] or [CRF MODULE NOT SUBMITTED].

Example (specific question):



Example (complete page)

OTHER

"Other, specify" variables should be annotated with the suffix "OTH". If the related variable name contains more than 5 characters then it is sufficient to have only "OT" or "O" to keep to the requirement for a maximum of 8 characters for the variable name.

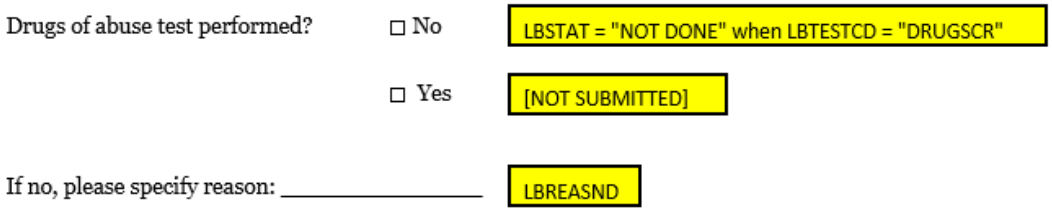
[17] FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019): section 4.1.4.6, p.32

Example:

Not Done / --REASND

Not done information that is prespecified/solicited on the CRF and will be submitted has to be annotated to --STAT (Completion Status). --STAT can be related to a specific question, module or a complete page. In every case it is recommended to have a “Reason not done” (--REASND) variable included if the information is collected on the CRF.

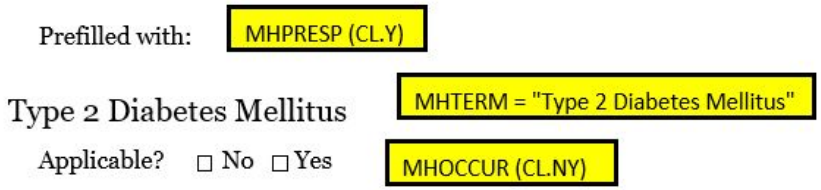
Example:



Pre-specified Events

When using pre-specified terms in the CRF, the variables --PRESP and --OCCUR should be used together with --TERM. The variable --TERM should contain the pre-specified text.

Example:



Controlled Terminology

There are different guidance's on how controlled terminology items should be included in the aCRF.

According to the Metadata Submission Guideline v1.0, the aCRF should not contain additional entries such as codelist codes (especially for drop-down lists in eCRFs).

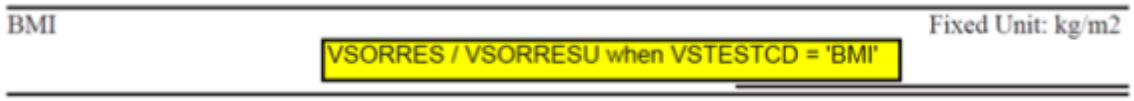
According to the Study Data Technical Conformance Guide v4.3, the aCRF should include the variable names and coding for each CRF item[18].
Note: There is no further guidance provided what "coding" means in this context. Whether this means to include the corresponding NCI codelist name only or each corresponding NCI codelist value(s) or the NCI codelist codes, is not clear. Providing the NCI codelist codes could be a problem if sponsor defined codelist or terms are used in the aCRF which are not available in the NCI CDISC controlled terminology catalogues.

There are different approaches how to handle the respective annotations in the aCRF.

Following the guidance provided by CDISC in the Metadata Submission Guide, only the codelist value for the parameter test code would be included in the annotated CRF.

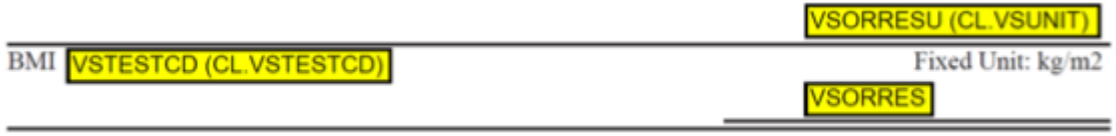
[18] Note: Currently no one of the document authors is reporting that CT annotations are included in their submitted acrf.pdf documents to the authorities. Up to end of 2019 also no one is aware of any FDA request or query in a submission because of missing CT annotations in the SDTM acrf.pdf.

Example:

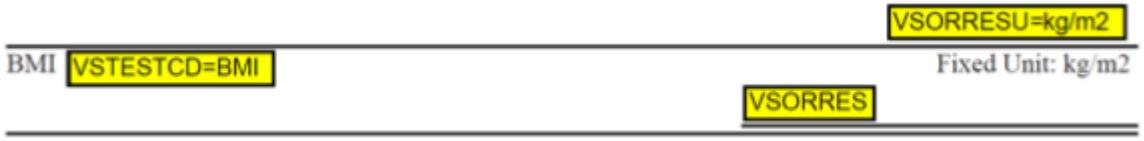


Following the guidance provided by the FDA in the Study Data Technical Conformance Guide, this might not be sufficient as it does not include the "coding" for each CRF item. But as it is not clear which kind of / level of coding is referred to in this context, even trying to follow the Study Data Technical Conformance Guide might lead to different results.

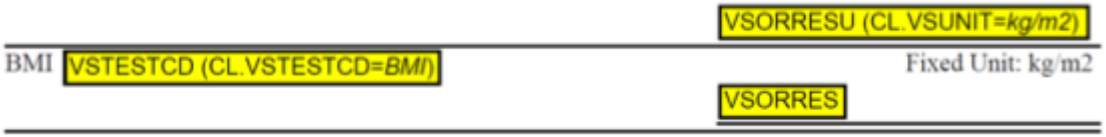
Annotation of codelist name only:



Annotation of codelist value only:



Annotation of codelist name and codelist value:



FINDINGS ABOUT

For any CRF data mapped to FA, the annotation of FAORRES should also indicate the corresponding FATESTCD and FAOBJ.[19]

Example:

RELREC

If relationships between data collected on the CRF pages are described using RELREC, this should be mentioned in the annotated CRF.[20]

Example from SDTM MSG, chapter 4.1.5:

Adverse events are collected on the Adverse Events CRF as numbered running records and the AE number is mapped to AESPID. If a subject discontinues due to an AE, the relationship is established by entering the associated AE number (AESPID) on the Termination CRF. A RELREC record is created to link the discontinuation due to an adverse event record in DS to the related adverse event record in AE via AESPID. This relationship is documented on the Termination CRF in the annotation:

Linked to related AE record via RELREC

[19] CDISC.org, "Clinical Data Acquisition Standards Harmonization" (CDASH IG v2.0, 2017-09-20): chapter 8.3.15
[20] CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): section 4.1.5, p.18

Alternative example from CDASHIG v2.0, chapter 8.1.4:

Questionnaires, Ratings, and Scales (QRS) Domains

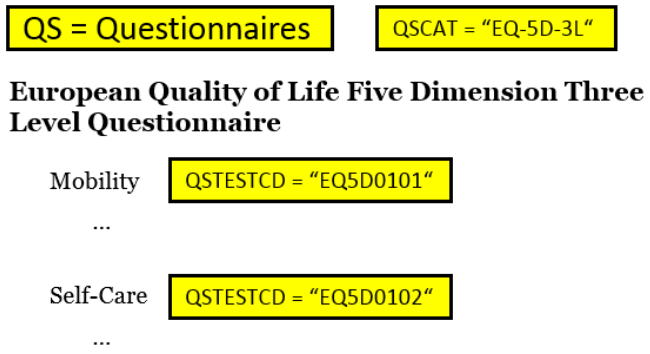
SDTM IG version 3.2 only describes the Questionnaires (QS) domain, while in SDTM IG version 3.3 additional QRS domains Function Tests (FT) and Disease Response and Clin Classification (RS) have been included.

If any questionnaire, rating or scale (QRS) is part of the aCRF, the US English version should be used for the annotation whenever possible. If no US English version is available, a local non-English version may be used instead. In this case any special local language characters have to be avoided because these might lead to SAS issues.

If a link regarding the used labels between the US English version and any local language version is needed, this can be handled in the cSDRG, but should not be included in the aCRF itself.

If more than one version of QRS exist, e.g. European Quality of Life Five Dimension Three Level Scale / European Quality of Life Five Dimension Five Level Scale, Brief Pain Inventory / Brief Pain Inventory Short Form, the information which version has been used will be reflected in QSCAT and QSTESTCD.

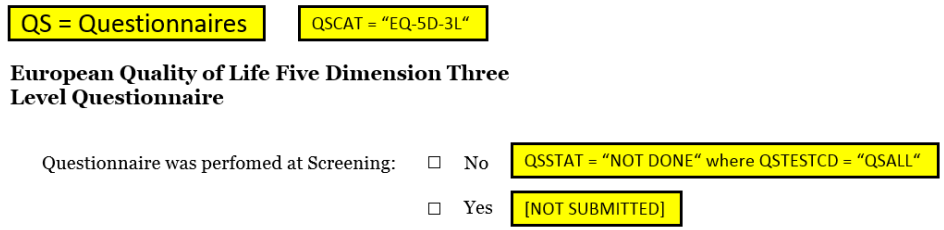
Example:



Codes like “UNKNOWN”, “Not Available” should not be added (afterwards) to a defined code list to collect any ‘missing information’ at the end of a trial. If such a code is an ‘official’ answer option in a published QRS item, then a respective term will be part of the approved code list.

If a complete QRS was not performed at the expected time point / visit etc., --TESTCD should be set to “QSALL”, “FTALL”, “RSALL” respectively.

Example:



Questionnaires (QS)

Depending on the setup of a questionnaire, some items may be logically skipped. The responses for these logically skipped items should be recorded and included in the submission dataset. See FDA STUDY DATA TECHNICAL CONFORMANCE GUIDE[21] for further information on how these data should be reflected in the QS domain.

External Data

Including the annotation of external data, i.e. data not directly collected and documented in the CRF, such as e.g. laboratory data or ECG data, into the annotated CRF is a helpful tool to facilitate full documentation and traceability to all source data.

One option would be to add an extract of the respective raw data specifications (variable names and labels) to the aCRF and include the SDTM mapping for the different items.

Example:

Data Transfer Specifications

Dataset PK PC = Pharmacokinetics Concentrations

Variable Name	Format	Comment
PATNUM USUBJID	$10	i.e. 010-012345
VISIT VISIT	$20	i.e. C1D1
TIMEPOINT PCTPT	$9	PRE-DOSE or POST-DOSE
ANALYTE PCTESTCD	$40	ANALYTE1 or ANALYTE2
CONCENTRATION PCORRES	$200	
…		

[21] FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019): section 4.1.1.3, p.22

If the annotation for external data is included in the aCRF, this might have an impact on any automated processes for define.xml creation and it should be carefully checked that the origin of the respective data is correctly assigned (e.g. eDT rather than CRF page x).

Another possibility would be to have all required information for the proper mapping of the external data to the respective data in the SDTM database already included in the data transfer specifications/agreement (DTS/DTA). This would allow the highest flexibility as a respective template including all required information for the sponsor specific SDTM implementation could be provided upfront to the data vendors or included as attachment into the DTS/DTA early in the setup process. This process helps to avoid surprises or gaps in the late project phase when timelines are tight and resources are a precious good if external data transfers as Pharmacodynamic, Immunogenicity and Biomarker data might be delivered after data base lock. In this case any missings or additions to the expected data files might then have a huge impact and can set timelines to risk. Especially when dealing with local labs a lot of data vendors still have only small up to no CDISC experience and knowledge. Therefore, a proper and gapless mapping or at least manually added annotations for all external data in the respective DTS/DTA or similar documentation should be considered as best practice.

7. Ongoing Trials (Drafted, Not included in Guideline Version 1.0)

8. Special Cases (Drafted, Not included in Guideline Version 1.0)

Update Process

Versioning

Legacy Data

9. Tumor Domains (Drafted, Not included in Guideline Version 1.0)

10. Automation of Annotations (Drafted, Not included in Version 1.0)

11. Glossary (Drafted, Not included in Version 1.0)

12. Version and Change History

Version	Date Published	Updated Chapter/Section	Details/Notes
V1.0d Draft for Review	2019-09-25	First published draft of Version 1.0 for Public Review in German CDISC User Network Community	
V1.0	2020-11-20	Initial Version 1.0	Added details in chapter 2, several text revisions, update of all footnotes and document references Added chapter 13) Links & References

13. Links & References

References:	ICH, "M2 Recommendations & Technical References": http://estri.ich.org/recommendations/index.htm ICC – International Color Consortium: http://www.color.org/index.xalter FDA, Study Data Specifications (V2.0, July 18, 2012): https://www.fda.gov/media/83880/download FDA, Study Data Standards Resources Page: https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources FDA, "Portable Document Format (PDF) Specifications" (v4.1, 21 Sep. 2016): https://www.fda.gov/media/76797/download FDA, "Study Data Technical Conformance Guide" (v4.4, Oct 2019): https://www.fda.gov/media/131872/download CDISC.org, “define.xml v2.0” Specifications (Define-XML-2-0-Specification, 2014-04-24): https://www.cdisc.org/standards/data-exchange/define-xml CDISC.org, "Study Data Tabulation Model Metadata Submission Guidelines" (SDTM-MSG v1.0, 2011-12-31): https://www.cdisc.org/standards/foundational/sdtm CDISC.org, “Study Data Tabulation Model Implementation Guide" (SDTM-IG) (v3.2, Portfolio, 2013-11-26): https://www.cdisc.org/standards/foundational/sdtm CDISC.org, “Clinical Data Acquisition Standards Harmonization” (CDASH IG v2.0, 2017-09-20): https://www.cdisc.org/standards/foundational/cdash