PhUSE US Connect 2018

Poster PP11

Common Define.xml File Issues Seen During FDA's JumpStart Service

Peter Glass, IBM, Washington, DC, USA
Caitlin Anderson, IBM, Washington, DC, USA
Margo Cohen, IBM, Washington, DC, USA
Kathryn Matto, IBM, Washington, DC, USA
Austin Taylor, IBM, Washington, DC, USA
Sergiy Sirichenko, Pinnacle 21, Plymouth Meeting, PA, USA
Max Kanevsky, Pinnacle 21, Plymouth Meeting, PA, USA
Michael Beers, Pinnacle 21, Plymouth Meeting, PA, USA
Mary Doi, FDA, White Oak, MD, USA
Alan Shapiro, FDA, White Oak, MD, USA
Barbara Witczak, FDA, White Oak, MD, USA
Jesse Anderson, FDA, White Oak, MD, USA
Lilliam Rosario, FDA, White Oak, MD, USA

ABSTRACT

The data definitions file, or the define.xml file, is a critical part of the electronic dataset submission for regulatory review because it defines the metadata structures that should be used to describe the datasets, variables, controlled terminology, and codes. A well-developed and complete data definition file allows the FDA review team to quickly understand the structure of the datasets within the submission and provides a reference that can be used throughout the review cycle. Through the course of providing JumpStart Services on 130 applications since 2013, the FDA's JumpStart team has had the opportunity to work with many define.xml files from a variety of applicants. This poster will highlight how define.xml files can be used during application review and the common issues with them.

INTRODUCTION

The JumpStart service provides CDER review teams with an assessment of the composition and quality of the clinical trial data submitted within an application. JumpStart runs a series of drug clinical trial data analyses early in the review process to assess data composition, quality, analyses options, and tools for the analyses, so reviewers better understand the data and have the information to conduct an effective evaluation of the drug submission. JumpStart provides these data findings to the reviewers within two weeks of the receipt of a new submission, giving reviewers time to clarify issues or make requests of the organizations developing the drug before proceeding in the review process. JumpStart also provides data analyses that highlight areas of concern and may warrant further attention. From inception, the JumpStart service has provided this assessment to over 130 applications and in July of 2014, JumpStart was recognized as a "Secretary's Pick" by Secretary Sylvia Matthews Burwell and Deputy Secretary Bill Corr at the HHS Innovates Ceremony.

JUMPSTART BENEFITS AND IMPACT

JumpStart is transforming the scientific regulatory review process in several ways:

- The JumpStart data fitness assessment and exploratory safety analyses are improving the review process for getting safe and effective drug products to the market.
- The initial assessment on data fitness gives reviewers and sponsors the opportunity to address potential issues early in the review cycle and avoid delays.
- By providing targeted exploratory safety analysis, JumpStart enhances the efficiency of safety signal identification and risk analysis. This has a direct impact on patients who expect the approval of safe and effective products that are appropriately labeled for marketing.
- JumpStart work session discussions encourage collaboration and communication among different disciplines and informatics experts. By building and enhancing these ties with one another, reviewers and informatics experts can improve and speed communication with sponsors to resolve data issues or questions.

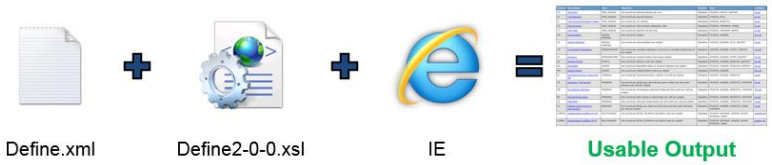
- JumpStart provides an opportunity for clinical reviewers to take advantage of modern tools and technologies, and expose more people to technical innovation.

COMMON DEFINE.XML FILE ISSUES SEEN DURING FDA'S JUMPSTART SERVICE

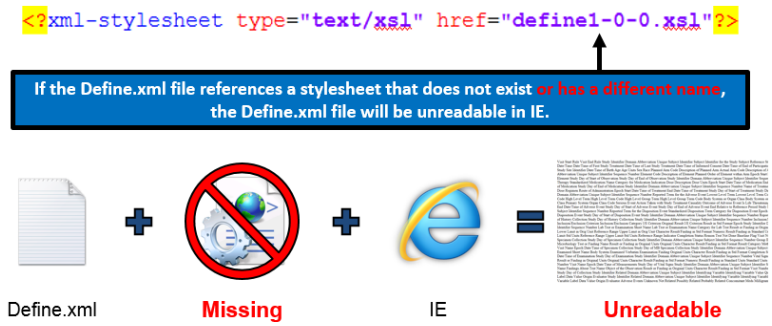
Part of the JumpStart service is an introduction to and a review of the Define.xml file submitted with study data. The Define.xml file is a critical part of the electronic submission for regulatory review because it defines the metadata structures that should be used to describe the datasets, variables, controlled terminology, and codes. A well-developed and complete data definition file allows the FDA review team to quickly understand the structure of the datasets within the submission and provides a reference that can be used throughout the review cycle. Define.xml file is also an important source of machine-readable metadata, which can be utilized for automated processes. If a Define.xml file has an issue that causes ambiguousness of data or affects the file's usability, it can cause delays to a review team's analysis. The JumpStart team has had the opportunity to work with many Define.xml files from a variety of applicants and has tracked common issues that can occur in the Define.xml file. This poster will highlight some issues that the JumpStart team has encountered during its analysis.

XSL STYLE SHEET REFERENCE ERRORS

During the review process, XML files are viewed using different web browsers including Microsoft Internet Explorer. However, with Internet Explorer, in addition to the XML file, XSL files are needed which describe how data in an XML file is to be presented to the user.



A common issue which results in an unusable Define.xml file is using an incorrect stylesheet reference.



LIMITED DETAIL IN COMPUTATIONAL ALGORITHMS

Computational algorithms in the Define.xml file allow review teams assess applicant methodology and allow for the verification of derived variables. Algorithms that provide comprehensive detail and are presented in an easy-to-understand format allow for quicker verification and a more thorough understanding of submitted data.

Example of an Algorithm with Limited Detail:

Variable	Label	Derivation/Comment
RFPENDTC	Date/Time of End of Participation	Last date of contact

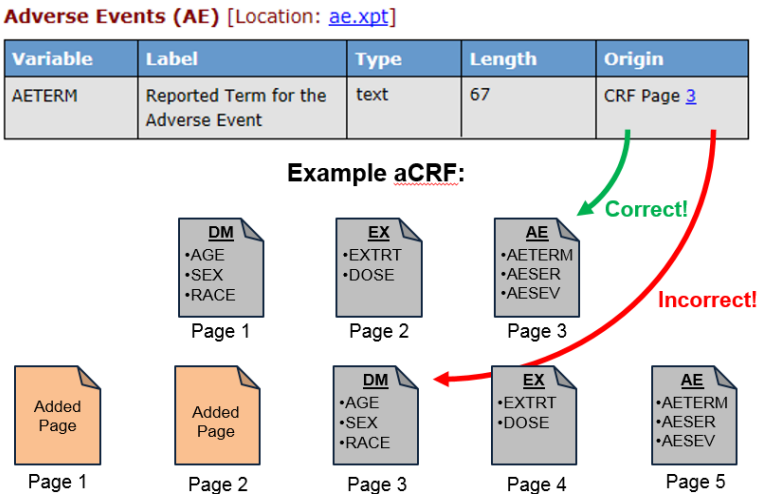
Example of an Algorithm with Added Detail:

Variable	Label	Derivation/Comment
RFPENDTC	Date/Time of End of Participation	Last known day of contact with subject; If DM.DTHDTC is not NULL then RFPENDTC = DM.DTHDTC. Else RFPENDTC is equal to MAX of --DTC, --STDTC, and --ENDTC variables from all domains.

INCORRECT ACRF LINKS

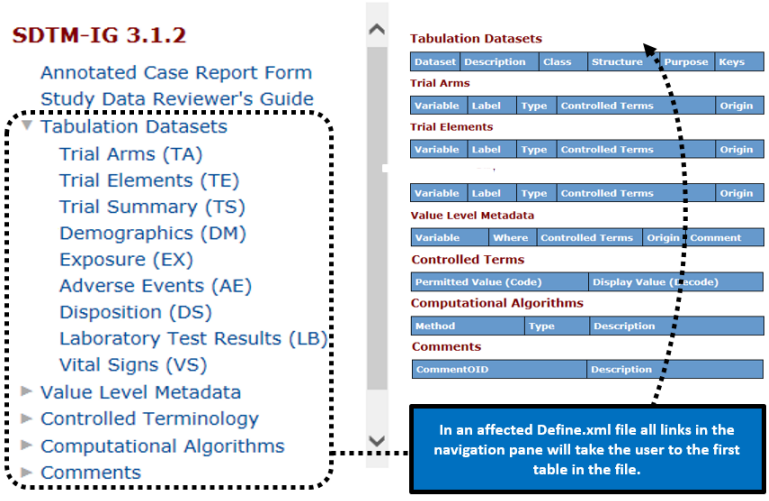
One of the most useful functions of the Define.xml file is its ability to link a variable's record in the file to the page where the information was collected in the annotated Case Report Form (aCRF).

A common issue is an incorrect aCRF link. This may be caused when a file name for an aCRF is changed without updating a reference in Define.xml file. For example, original "blankcrf.pdf" or "acrf-study-1.pdf" renamed to new "acrf.pdf". Another cause can be adding pages to the aCRF without updating links in the Define.xml file.



BROKEN LINKS IN NAVIGATION PANE

As the Define.xml file can be a very long document with many segments, easy navigation is important. The navigation pane running along the left-hand side of the Define.xml file is critical to a reviewer to help manage the size of the document. However, a common issue is a navigation pane that takes the user to the top of the document when any link in the pane is selected. This issue can be due to an outdated or incorrectly modified stylesheet.



MISSING MEDDRA VERSION OR SDRG MEDDRA VERSION MISMATCH

The MedDRA version used in the coding of the adverse events needs to be specified in every application. Various analytical tools can then be used to verify and analyze the coding of the adverse events. Other tools can read this important metadata when the MedDRA version is provided as an 'External Code List' attribute and allow for quick identification of the MedDRA version.

External Dictionaries

Reference Name	External Dictionary	Dictionary Version
Adverse Event Dictionary	MedDRA	20.0

To avoid confusion, the version stated in the Define.XML file should match the SDRG.

1.2 Study Data Standards and Dictionary Inventory

Standard or Dictionary	Versions Used
Medical Events Dictionary	MedDRA v18.1

COMPUTATIONAL ALGORITHMS THAT REFERENCE INTERNAL DATASETS

As noted in the issue to the left, CDER review teams depend on clear, reproducible computational algorithms in the Define.xml file to verify the methodology of derived variables. This can be difficult if algorithms contain references to internal applicant datasets or variables which are not submitted in the data package.

Example of an Algorithm with References to Internal Datasets and Variables:

Variable	Label	Derivation/Comment
BRTHDTC	Date/Time of Birth	cat(put(demo.brth_year,4.),'-',put(demo.brth_month,z2.),'-',put (demo.brth_day,z2.))

This algorithm references a dataset (demo) and variables (brth_year, brth_month, brth_day) that cannot be found in any submitted data.

A PhUSE working group is developing a "Define.xml Completion Guide". http://www.phusewiki.org/wiki/index.php?title=Define-XML_V2.0_Completion_Guidelines_%26_Style_Sheet_Recommendations

MISSING OR MERGED CODELISTS

Controlled Terminology (CT) tables in the Define.xml file, also known as codelists, help reviewers quickly examine terminology used by applicants and provide a reference for any codes reviewers might encounter during their investigation of study data. Missing codelists can cause errors during validation due to the lack of information to help reviewers understand how data was collected.

Unit

Permitted Value (Code)
beats/min
g
g/L
kg
kg/m2
mg
mL
mmhg
mmol/L
U/L
ug
umol/L

The merged Unit codelist on the left contains units from three domains.

LB Unit

Permitted Value (Code)
g/L
mmol/L
U/L
umol/L

VS Unit

Permitted Value (Code)
beats/min
kg
kg/m2
mmhg

CM Unit

Permitted Value (Code)
g
mg
mL
ug

CONTACT INFORMATION

Learn more about JumpStart at: https://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm397921.htm