

Metadata, Study data standards, Master data, terminology, interoperability: key concepts underlying compliance to FDA guidance on electronic submission

Isabelle de Zegher, PAREXEL Informatics, Belgium

ABSTRACT

Management of metadata and master data is key for system and data integration. However these are ill defined concepts and there is a lot of confusion in our industry on what they mean, how to manage and implement them. As part of the Emerging Technologies working group with FDA/PhUSE, we started a project around “metadata definition”. Its goal is to develop common industry definitions, supported by operational examples, around metadata, master data, controlled terminology and interoperability....

This talk will present the results of the project by focusing on questions such as: what is a study specific data standard versus enterprise metadata, what a metadata repository (MDR) is, how does metadata related with master data, what is controlled terminology and how to manage it more efficiently. The second part of the talk will focus the importance of these concepts while ensuring compliance with the FDA February 2014 guidance on electronic submission.

Introduction

Metadata management, value level metadata, master data management, pooling, integration, aggregation, .. are all terms that we use and hear increasingly. Yet experience shows that different people in our industry do not always mean the same thing when using these terms. This is the same case for the word “football” between American and European people: same word, different games. And in an area which is already complex, this increases confusion unnecessarily.

As part of the PhUSE CSS initiatives, we decided to put a working group together to define all terms that we believe are not well understood today. This paper present the results of this working group as well as some of the lessons learned from PAREXEL informatics, in our approach around metadata management, to support more efficient compliance to data standards – which are themselves metadata.

PhUSE CSS Emerging Technology Working group

FDA/PhUSE Computational Science Symposium (CSS) is a collaborative effort between industry and the FDA to work on implementation of data standard. In 2013 we launched the Emerging Technologies working group (http://www.PhUSEwiki.org/wiki/index.php?title=Emerging_Technologies), a working group related to specific computational science topics, tools, technologies, and approaches. At the start, the following topics were mentioned as being the one with the most interested: semantic web technology, analysis metadata, cloud computing and big data.

Big data being considered as a second priority, during 2013 and 2014, only 3 working groups were set up. This paper describes the result of the metadata working group.

Metadata data definition

The objective of this working group was to develop common definitions around metadata and related aspects across the industry.

We first identified a list of terms considered as being used differently across different organizations, and we grouped these terms into the following categories:

- metadata management,
- controlled terminology and value sets
- master data management,
- interoperability,
- data aggregation, pooling and integration.

As the purpose of the document is to be of use for operational people, it contains not only definitions but also a short description and example of usage. Whenever possible, the definitions are built from those existing definitions from FDA guidance's, CDISC glossary, check cross industry definition (e.g. Gartner). Reference to the source definition is provided either directly with the definition or in the reference section.

When we started we believe, this could be done in 3 months with a meeting every other week. We ended up working for 12 months with a meeting every week.

The document with all the definitions resulting from the working group can be found in

http://www.PhUSEwiki.org/wiki/index.php?title=Metadata_Management

During the talk we will focus on the terms that we considered as being the more confusing i.e. metadata, master data, controlled terminology, pooling.

Metadata and compliance on electronic submission guideline

1.1 Lessons learned:

While working with the PhUSE CSS working group, within PAREXEL Informatics we worked as well on a new approach to metadata management with customers. Putting together the more academic work within PhUSE CSS and the very operational work in customer project we came to the following conclusions

- Data Standards – or structural metadata - are required for electronic submission, and needed for Data Integration.
- Efficient Data Integration and compliance to regulatory standards does not start after pooling (retroactive approach) ; it starts with the protocol (proactive approach)
- A proactive approach is based on 2 components
 - Agreement on study “Master” Data (study ID, visit ID, ..) and related descriptive metadata
 - Definition of study structural metadata – aka study specific data standards – as a subset of the enterprise wide variables and value sets contained in a Metadata repository (MDR)
- To be manageable, data standards variables defined in an MDR need to be grouped in semantically meaningful “clinical research concepts” (CRC)
- Classical code lists should be replaced by value sets, linked to a code system in a controlled terminology server

1.2 Practically: what is pro-active data integration and electronic submission?

With these in mind the study set-up will fundamentally change.

The protocol analyst will first identify and agree the key master data for the study (such as product id, site id, visit id, trial id) and make sure all the stakeholders have a clear description of the master data when setting up their system for data collection. For site and investigator this ideally should be supported through master data management system linked to the CTMS.

The second step the protocol analyst needs to ensure is to define the study data standards i.e. select clinical research concepts relevant for the study and then generate the list of variables (CDASH for data collection – linked to SDTM). Ideally this should be automated through a concept based MDR who includes definition of concepts, composed by semantic meaningful group of CDASH variables, which are in turn linked to SDTM variables.

PhUSE 2014 – DH08

In parallel with process change we need new tools to manage metadata; we need to move away from variable based metadata repository to concept based metadata repository. This concept based MDR should have the following key features

- Grouping variables into semantically meaningful clinical research concepts (following industry wide patterns as defined in BRIDG)
- Linking data collection (CDASH) to data submission (SDTM) variables
- Linked with controlled terminology managed in a separate Controlled Terminology server

CONCLUSION

To make progress as an industry around data standards and electronic submission, as we all as data integration we first need to make sure we speak the same language and stop any confusion around meaning of words.

If we want to solve the challenge of data integration – while ensuring regulatory compliance - we need to change the way we consider data standards from a retro-active way (building define.xml at submission time) to a pro-active approach (study data standards - underlying defined.xml - defined at study set-up). This requires process changes as well as availability of new tools to manage metadata and controlled terminology.

References

The document referred into this paper can be found in.

http://www.PhUSEwiki.org/wiki/images/c/cc/ET-project- Metadata definitions-V1-2014_07_15.pdf

ACKNOWLEDGMENTS

The following people actively participated to the authoring and review of the Metadata definition document

Mitra Rocca (co-chair)	Mitra.rocca@fda.hhs.gov
Marcelina Hungria (co-chair)	mhungria@dicoregroup.com
Yun Oldshue	yun.oldshue@takeda.com
Kenneth Stoltzfus	kenneth.m.stoltzfus@accenture.com
Julie James	julie_james@bluewaveinformatics.co.uk
Tim Church	tim.church@torch.uk.net
Gregory Steffens	Gregory.steffens@novartis.com
Praveen Garg	Praveen.Garg@iconplc.com
John Leveille	jleveille@d-wise.com
Aimee Basile	abasile@celgene.com
Sam Hume	shume@cdisc.org

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dr. Isabelle de Zegher

Senior Director, Clinical Information Management

PAREXEL Informatics

Parc des Collines

T +32 2 767 16 48

M +32 478 48 28 54

Brand and product names are trademarks of their respective companies.