Paper CD09

Managing ADaM Metadata

Ann Fitzherbert, GlaxoSmithKline, London, UK
Khilit Shah, GlaxoSmithKline, London, UK

ABSTRACT
This presentation will focus on GSK's implementation of a simple system (The ADaM Metadata Toolkit) to provide tools for the centralised management and consumption of GSK standard ADaM metadata.

Within GSK the toolkit consists of four major components, with varying levels of requirements and user access. A key feature of the toolkit is to store and maintain relationships, allowing versioning of the ADaM metadata by date and / or version of the global ADaM IG.

ADAM METADATA

Within CDISC, the ADaM metadata is seen as a central component to the whole process of a CDISC compliant submission. It's the framework (or backbone) for your data (and in this case, simply does not only refer to your datasets, but also to the results that are produced), that helps to convey to the reviewer what is in the data.

The metadata helps the reviewer by allowing analytical result to be traced back to the analysis dataset, collected data, and ultimately source (e.g. CRF, lab data). As well as structural information, such as length and format, the metadata in a CDISC provides fields to indicate the source (e.g. either another variable, or a derivation etc), at dataset, variable, value and result levels

ADaM metadata focuses on four levels; Dataset, Variable, Value (parameter level) and optionally the results. We needed a system to manage the first three kinds of metadata in a standard and structure manner. Within GSK, we have done this using the ADaM Metadata Toolkit.

INTRODUCTION TO THE TOOLKIT
In designing a system to manage our ADaM metadata the following requirements were listed which were then used to design each of the component parts of the toolkit. Some of the basic components such as a Database, Web interface and Excel tools were defined upfront as these were central to the toolkit's design.
- Provide a central repository ("**ADaM Metadata Database**") to store key ADaM metadata in a single relational database which maintains the appropriate relationships between different item levels
- Provide versioning to allow definition of the ADaM metadata by date and/or version of the global ADaM IG, as appropriate for the type
- Provide a web interface ("**ADaM Metadata Viewer**") for users to view the latest version of the metadata
- Provide an Excel tool ("**ADaM Metadata Spec Builder**") to allow users to selectively extract any version of the metadata into a single Excel workbook
- Provide an Excel tool ("**ADaM Metadata Manager**") to allow domain owners to define updates to the metadata in the repository
- Provide a process for loading updates to the repository
- Provide a process to generate a Define.xml file for the latest ADaM metadata
- Provide a process to generate a Define.xml file from the ADaM Metadata Spec Builder

The ADaM Metadata Toolkit will comprise four main components;
- ADaM Metadata Database (AMD), a relational (Oracle) database which stores the metadata
- ADaM Metadata Viewer (AMV), a web tool to allow viewing of the latest metadata
- ADaM Metadata Spec Builder (AMSB), an Excel tool to allow extraction of any version of the metadata into Excel
- ADaM Metadata Manager (AMM), an Excel tool to allow content owners to define updates to be loaded into database

ADAM METADATA SYSTEM

The ADaM Toolkit will work together as shown by the diagram below:
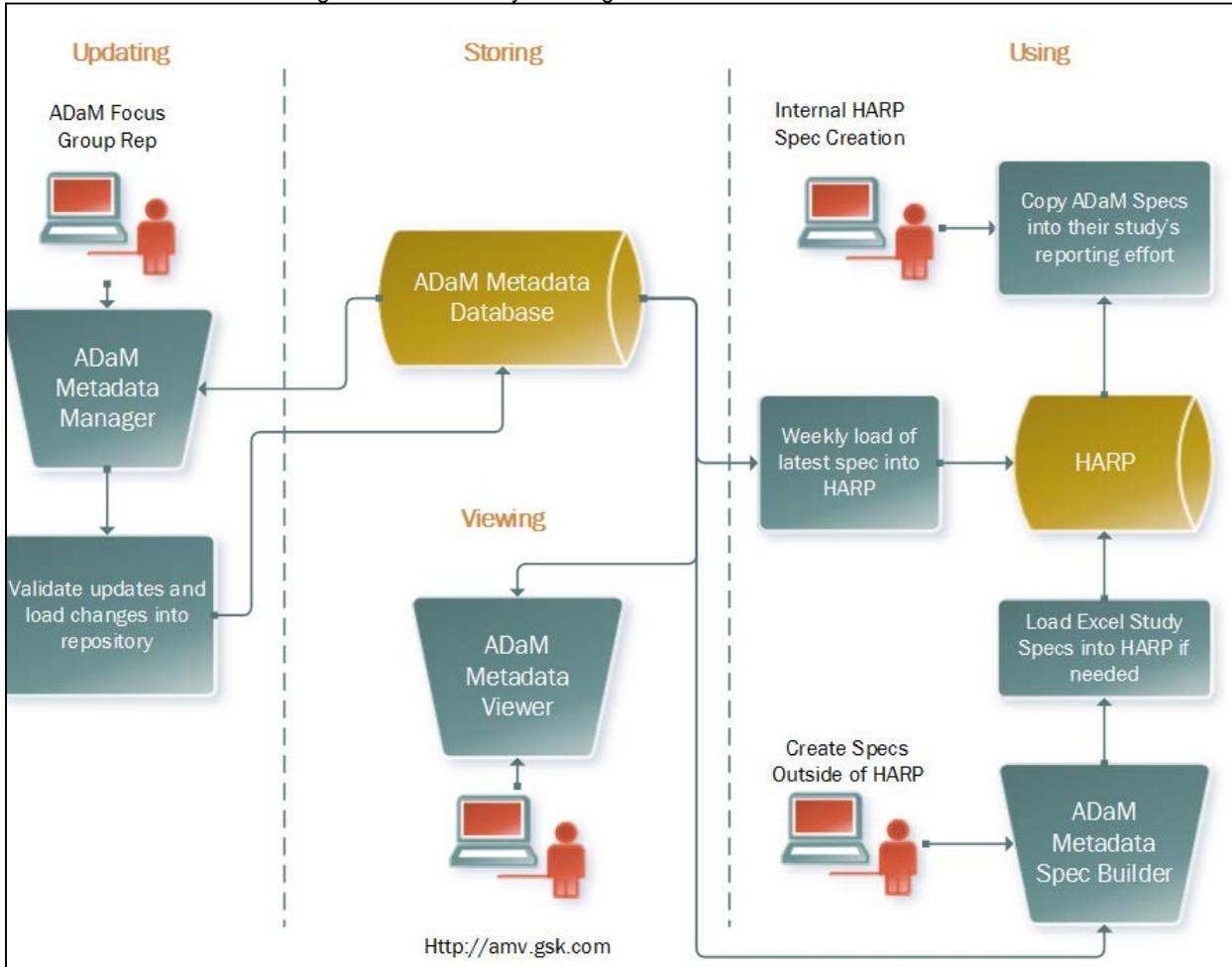
FIGURE 1 - ADAM METADATA TOOLKIT

ADAM METADATA DATABASE (AMD)

The ADaM Metadata Database is a relational (oracle) database that stores the metadata. It has a series of tables containing all of the information required for each domain specification. Each table has its own level of version control ranging from the controlled terms list that has no version control to the specs table which has IG version control. Key tables within the AMD include:
- The Domains table : contains the metadata for each domain and its attributes;
- The specifications table: links together the information from the domain table along with Variables, Value Level Metadata, controlled terms and computational methods, each of which is stored in a separate table of the database.

The specification versions are managed using the IG version control. There are two columns associated with each specification that are "From" and "To" IG Guide.
- "From IG Guide" is populated when a spec is entered into the database
- "To IG Guide" (initially missing). When a specification changes as per the IG, the old spec has the "To IG Guide" populated with the old version and a new spec is created for the new IG.

The variables within a spec are also IG version controlled. If a new IG calls for a change to a variable, a new variable will be created with the new IG in the "From IG Guide" column and the old version of the variable will be closed out using the "To IG Guide" column. If the new IG version doesn't require a change to a variable that variable will be left as it is in the database with the "To IG Guide" left blank.

2

Along with the IG versioning each piece of metadata in the database also has "From Date" and "To Date" attributes associated with it. This allows updates and new versions to be stored in the same way as the IG version.

The database has been set up so that the specifications are additive. We hold Core, Therapeutic Area (TA), and Indication level specifications.
- Core is the lowest level and all the core variables will be available in the TA and Indication level specifications.
- The TA and Indication specific variables and metadata are then added to the core specification.

This inter-dependence means that any changes made to the core specifications get cascaded down to the TA and Indication levels. If this is not appropriate for a certain variable, a TA or Indication version of that variable can be added to the database to replace the core one.

The positive aspect of this design means that less space is required in the database to store the metadata. This makes the AMD unique from the many currently available off the shelf solutions that are designed using an object model. In those solutions, the model is often driven top to bottom, so if you were to make a change to the Core domains/variables you can choose whether to approve these changes to follow through to the TA or Indication level domains/specifications.

Controlled terms within the database are not version controlled as there is currently no requirement for this. All that is stored is the domain and type that are associated with the controlled term. Computational Methods are handled in the same manner.

Value Level Metadata and the Controlled Term table that stores the code/decode pairs have internal control only. This means they don't have any date or IG version attributes but they do have an internal version number. If a VLM or Controlled Term changes then the new version is added to the table and the internal version number is incremented. When the standards are pulled using the ADaM Metadata Viewer or Spec Builder, the latest version is always picked up from these tables.

ADAM METADATA VIEWER (AMV)

The ADaM Metadata Viewer is a web based tool to allow users to easily view the latest standards in the database. The AMV gives read only access to the ADaM Metadata Database.

Upon accessing the AMV the page that you are taken to is shown below and has the following actions:

FIGURE 2 - AMV

The AMV has a series of tabs, each providing specific functionality to view the ADaM metadata information held in the database.

- Within the ADaM Domains tab, all domains currently held in the database are displayed. Additional information such as class, source, data structure, key variables and a description of the domain is provided alongside.
- The domain names are clickable and clicking on any single domain will load the full specification for that domain on the ADaM Specs tab; alternatively, users can use the "drop down" filters on the top of the page to jump to the specific domains.
- Filters are also enabled to navigate to a set of "class" such as 'BDS' 'OTHER' 'ODS' etc and that will reduce the table to show only domains within that class.
- Below the Filter buttons there are two tools buttons, one which "resets all filters", to bring you back to this main view. The other button "Export to Excel" will produce a worksheet called 'Domains.xls' which shows the same information as this page on the website.

Once you have selected the domain you wish to view, you are taken to the specs tab of the website.
On the ADaM Specs Page you can view one spec or all specs in a table with each spec following on after the pervious.

- There are filters to select the Therapeutic area and the Indication you wish to see.
- The user also has the option to select to view a variable; for example a user could choose to view the variable AVAL and this will bring back a row for each domain that contains AVAL.
- The variable name is clickable and automatically links to the Master variable list tab. The Master Variable List will show only the latest version of a variable.
- If a spec has a core and therapeutic version then if the TA filter is applied the page will show all variables for the selected TA, if TA filter is not applied then you will see both the CORE and TA variables shown on the page.
- Value Level Metadata (VLM) is also a clickable column that links to the VLM tab and shows only the Values for the one you have clicked into from the specs page. The regular VLM Name, TA, Indication and reset all filters are available on this tab so you can view all VLM that is available, or you can use the PARAMCD filter to filter down the table this way too.
- The computational method and controlled term tabs act in the same manner as the VLM tab, enabling users to view the latest version of these items.

ADAM METADATA SPEC BUILDER

The Spec Builder toolkit facilitates the creation of ADaM datasets specifications, Value Level Metadata (VLM), Controlled Terminology (CT) and Computational Method (CM) in an Excel format. This can then be sent onto the study teams / CROs reporting the study.

Invoking the spec builder, the tool makes a connection to the database with further options to retrieve a "default" spec based on ADaM IG version, Therapy Area and Study Indication. Optionally, you can filter by date to retrieve previous versions of the specs.
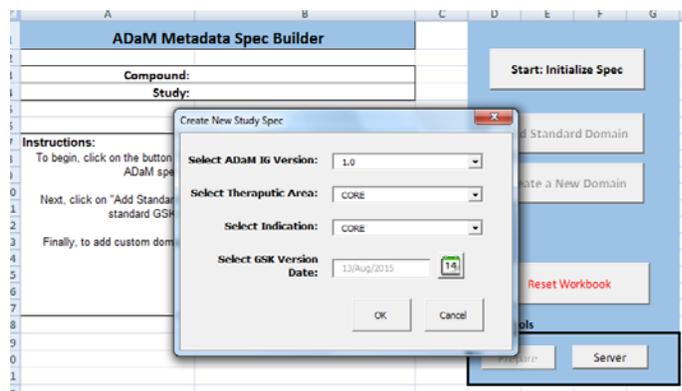
Additional DOMAINs can be added onto the newly created SPEC file.
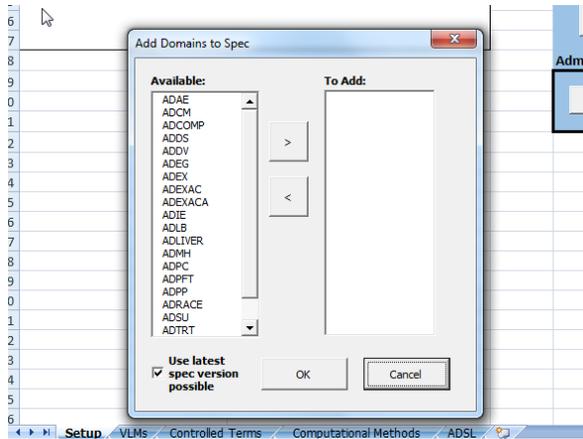
If study specific CTs, VLMs or CMs needed to be defined, separate worksheets exist within the SPEC file that can be updated. Following this, the information can be "added" onto existing variables within each of the domain

The user also has the ability to add new variables within existing domain specs as per study requirements. If its deemed that additional variables may be suitable for more studies, then this information can be added to the ADaM Metadata Database via the ADaM Metadata Manager.

ADAM METEDATA MANAGER
The ADaM Metadata Manager is an Excel tool that allows content owners to define updates to the specifications held in the database. Upon opening the tool the user is presented with an option to load a domain, they are able to select an existing domain or create a new domain if appropriate. They may also specify the Therapeutic Area and Indication for the domain they wish to load.
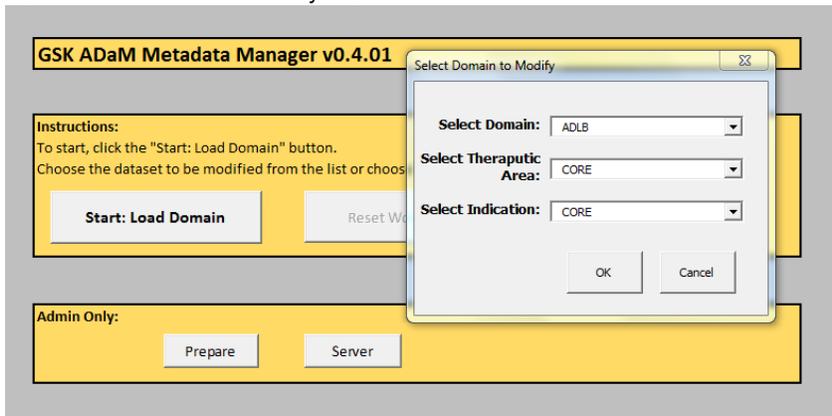
Upon clicking the OK button the domain is loaded into the Excel workbook on a new tab. Once the domain is loaded the content owner can make any changes they need to any number of variables or other metadata.

If the content owner is editing a TA or Indication level specification the only variables they will be able to alter are the TA or Indication variables. If a change is needed to a core variable then the core version of the spec must be loaded to make the update.

Once the update has been made in the tool the Excel workbook is saved and sent to a toolkit administrator to be uploaded into the database.

GLOSSARY

ADaM	Analysis Data Model
AMD	ADaM Metadata Database
AMV	ADaM Metadata Viewer
AMSB	ADaM Metadata Spec Builder
AMM	ADaM Metadata Manager
BDS	Basic Data Structure
CDISC	Clinical Data Interchange Standards Consortium
CM	Computational Method
CRO	Contract Research Organization
CT	Controlled Term
HARP	Application tool used by statisticians and programmers at GSK to facilitate the analysis and reporting (A&R) process
IG	Implementation Guide
ODS	Occurrence Data Structure
TA	Therapeutic Area
VLM	Value Level Metadata

CONCLUSION

The ADaM Metadata Toolkit is the current method that enables the management, storage and viewing of the metadata that has thus far been developed. The toolkit is user friendly enabling the metadata to be viewed by version and/or date and allows updates to the metadata by administrators.

Uploading of updated metadata is loaded into the database via a SAS® script. The creation of a define.xml from AMSB is also via a SAS script. These SAS scripts complete the requirements of the toolkit.

The toolkit provides a complete system to manage GSK's ADaM Metadata and the initial user feedback has been very positive on every aspect.

CONTACT INFORMATION

Author Name	Ann Fitzherbert & Khilit Shah
Company	GlaxoSmithKline
Address	Stockley Park West, 1-3 Ironbridge Road,
City / Postcode	Uxbridge, Middlesex, UB11 1BT, UK
Email:	ann.h.fitzherbert@gsk.com khilit.m.shah@gsk.com
Web:	https://www.gsk.com/

Brand and product names are trademarks of their respective companies.