**Meeting Minutes for the ETL Demo:**

**Presentation: Ordinal Data - Ptolemy.V**

**Demo: 3:00-4:00 pm, May 25, 2017**

**NCI Webex:** https://cbiit.webex.com/cbiit/j.php?MTID=m69858d7154a2dbde79a883708aa0d2f9

|  |
| --- |
|  |
| **TCON:** 1-650-479-3207 Call-in toll number (US/Canada) |

**Minutes:** Alec Petkoff, Mitra Rocca

|  |  |  |  |
| --- | --- | --- | --- |
|  | **CDM Harmonization Team Members** | **Organization/Office** | **Role** |
| X | Mitra Rocca | FDA CDER OTS | Lead |
| X | Ken Gersing | NIH/NCATS | Co-lead |
|  | Lisa Schick | NIH/NCATS | Team Member, NCATS |
| X | Robinette Renner | NIH/NCI | Team Member, NLM |
| X | Aras Efterkhari | NIH/NCI | NCI Project Manager |
| X | Yelena Kagan | NIH/NCI | Team Member, NCI |
| X | Susan Lumsden | HHS/OS | ASPE Project Officer |
| X | Liz Amos | NIH/NLM | Team Member, NLM |
|  | Vojtech Huser | NIH/NLM | Team Member, NLM |
| X | Rashedul Hassan | FDA CDER OTS | CDER OTS technical member |
| X | Denise Warzel | NIH/NCI | Team Member and NCI technical lead |
| X | Lisa Lang | NIH/NLM | Team Member, NLM technical lead |
| X | Rachael Roan | NIH/NLM | Team Member, NLM |
| X | Albert Taylor | HHS/ONC | Team Member and ONC lead |
| X | Scott Gordon | FDA CDER OSP | Project Manager, responsible for coordination of tasks across agencies |
| X | Alec Petkoff | FDA CDER | Engility Contractor – Project Manager Support |
| X | Yaffa Rubinstein | NIH/NLM | Team Member, NLM |
| X | Elad Sharon | NIH/NCI | Team Member, Oncology Researcher |
|  | Sean Khozin | FDA/OHOP | Team Member, Oncology Medical Reviewer |
| X | Duc Nguyen | NIH/NLM | Team Member, Computer Scientist |
| X | Fatima Frye | FDA/CDER | Team Member, FDA COR |
|  | Christophe Ludet | NIH/NLM | Team Member, NLM |
| X | Dheeraj Gobburu | NIH/NCATS | Technical Lead (NCATS) |

**Introduction of all participants:**

* Introduction of the HHS PCORTF CDMH team.
* Introduction of Ordinal Data (formerly Life Data Systems) guest: Bill Tulskie

**Overview**

* Developed to allow the combining of information and resources.
* Designed to translate data in to a standardized data model. Makes it easy to convert clinical data into a model they are more familiar with. Also makes it easier to publish subsets of data. And easy to regenerate data with new information.
* The target audience is clinical researchers.
* Part of the architecture uses an open source MonetDB.

**Ptolemy.V Demo**

* Opening page is a guide, lets you know what you can do.
* Uses Source Data Sets. It has a browsing function. You can drill down in the data elements. (E.g. sex). Then you can get into related data elements.
* Dates are available, but did not have them in the demo for simplicity.
* Many things are customizable.
* Easy way to find things out is to use the Source Data Search. You can browse the data sets the same way.
  + Do you have the ability to do more nuanced mapping, or complex transformation? Yes, that can be done as a customization. It is not pre-built, but the architecture allows us to do that. We will be building up that library over time, especially for standard conversions.
* Registering a new data set: Page is just fill-in (uses sample information). Then Browse the file share to see what data you may want to take in. List of sources could also include cloud sources.
* Has a data dictionary.
* Search is built on keywords in the indices.
  + Any method to the ordering? No, because the list is all inferred. IF the list had a set of mappings you would be able to see those.
* Operations Buttons - puts elements into the translation builder. In the “Cart” you can see everything for test 2, but it also included related data sets (automatically put in the cart). Also gives standard data elements.
  + What is the rational in bringing in the other files because they have some attributes? Because we are at the stage where we would be doing some analysis. So Ptolemy.V is telling me that these other files have the elements. You could include or not include.
* Generates standard data. Puts onto Excel.
  + Issue - Files we are using will be fairly large. - The code that does the translation is written in Java so that it moves faster.

**Discussion:**

* What’s ahead: Ptolemy.V 2.4 comes out in mid-June.
* Adding a source file data dictionary generator; Additional search filtering,(browse and select); Improved translation builder Usability Aids; User workflow aids and improvements; Translation search and select.
* Release in July/August will include: Expanded relationship types and ontological search model; Automated refresh or translations; Admin console for translation job tracking; Enhanced output file and excel file browser; New mapping functions
  1. Will a data set automatically map into a destination model? You could build a translation ahead of time with dummy data, and then when you get real data, use that. You can also expand the data dictionary file to include redefined mappings. There is no roadblock to expanding the data dictionary to remapping to published standards.
  2. Coding question, mapping relationships: Can you make it so that anytime you see female they become a 2? Answer, yes. The way that would be implemented, is that you can do a search for preexisting mappings. Search, find, then apply function.
  3. Issue concerning scale of data: Any concern over really big data? Scale does not give me concern. It depends on how much hardware you are willing to use to do the job. More of a concern would be how many columns, but MonetDB solved that.
* All tables can be queried.
* You can also have a direct SQL interface
* Would like to follow up on use of standards.
* When you generate an output table, we keep the file as well as all the tables used to build it.
* All the ETL tools need customization.
* Our team needs what tools we need.

**Presentation of HHS/ASPE PCORTF project focusing on development of a Cross Network Directory Service (CNDS) and Discussion Meeting with the HHS PCORTF Common Data Model Harmonization Team around Ordinal Data - Ptolemy.V Demo; CNDS Presentation and other items: 4:00-6:00 pm**

**Ptolemy.V Demo**

**Advantages:**

* Can handle unlimited columns, leveraging MonetDB.
* Underlying capabilities can be customized for each customer.
* Simple to use by researchers vs. data managers only.

**Disadvantages:**

* Ontology is not fully developed. (will be developed next month).
* Creating data dictionaries in next release of their tool.
* Looking at improvements and usability of their GUI.
* Basis merge and transform data, but not an off the shelf tool.
* May not be scalable for large volume of data.