

Workshop:**Semantics to Support the NCI Cancer Research Data Commons (CRDC)**

Tuesday May 15 and Wednesday May 16, 2018

Workshop Goals

The expected outcomes of the workshop and follow-up documentation will be:

- 1) A requirements and evaluation document, which identifies and summarizes a set of high-level requirements, to address semantic needs and potential approaches for evaluation testing of the Cancer Data Aggregator (CDA);
- (2) A set of options with pros & cons for using and/or extending current methods and resources (e.g. NLP) to: (i) support semantic query capabilities; (ii) facilitate metadata annotation; and (iii) minimize efforts for data validation and submission;
- (3) A set of recommendations to support ongoing engagement with the community to ensure the semantics underlying the CDA improve and evolve as people contribute to and use the CDA.

Agenda**Day 1: May 15**

- | | |
|------------------|-----------------------------------------------------------------------------------------------------------------------|
| 9:00 - 9:15 am | Introductions and Agenda Overview (Bron Kisler and Melissa Haendel) |
| 9:15 - 9:40 am | <u>Pediatric Cancer Data Commons</u> , Dr. Sam Volchenboum |
| 9:40 - 10:05 am | <u>APOLLO: A High Profile Use Case with Unique Challenges for the CRDC</u> , Kathleen Darcy |
| 10:05 - 10:30 am | <u>Project Genomics, Evidence, Neoplasia, Information, Exchange (GENIE)</u> , James Eddy/Sara Gosline |
| 10:30-10:45 am | <u>NCIt OBO Edition and its Use in Phenotype/Genotype Efforts across Non-Cancer and Cancer Diseases</u> , Jim Balhoff |
| 10:45 -11:00 am | <u>Modeling Landscape Analysis</u> , Bron Kisler |

- 11:00 -11:10 am Explanation of breakout sessions and assignments (lead facilitators, note taking, reporting)
- 11:10-11:30 am Break (One breakout group to move to NCI conference room)
- 11:30-12:30 pm **Breakout Session 1: Key Questions related to Semantics for Query, Data Aggregation, and Analysis** (3 groups w/ Bron, Melissa and Betsy to facilitate)
Objective: Identify the key capabilities and approaches needed to support semantics across CRDC nodes
- 12:30 - 1:30 pm Lunch break
- 1:30 - 2:00 pm **Breakout Session 1 (Continued)**
- 2:00 - 3:45 pm **Breakout Session 2: Key Questions related to Semantics for Data Submission** (3 groups w/ Bron, Melissa and Betsy to facilitate)
Objective: Identify the key processes and approaches to provide semantics support for data submission
- 3:45 - 4:00 pm Break (Everyone return to the main meeting room)
- 4:00 - 5:00 pm **Session 3 (All Together): Landscape Summary and Feedback from Participants**
Objectives:
1. Report and feedback from breakout groups
2. Agree on focus for Day 2
Compile requirements, including information about what people are currently using and why; discuss pros/cons of approaches to prepare for developing recommendations on Day 2.
- 5:00 - 5:15 pm Group Photo
- 6:00 – 7:00 pm Meet for Happy Hour Social at Homewood Suites

Day 2: May 16

- 9:00 - 9:30 am Q&A and Discussion from Day 1
- 9:30 - 9:45 am Explanation of Breakout Sessions (4 groups w/ 10-12 people per group; people self-select roles (Note taker, reporter)
- 9:45 - 10:30 am **Breakout Session 3: Begin to formulate recommendations for semantics to support Query, Data Aggregation, Analytics, and Data Submission**
- Objective: Discuss the pros/cons of various approaches to support cross-node query, data aggregation, analytics and data submission from Day 1. Review existing semantic standards and approaches and discuss pros/cons. Where do they fall short? Where are there mismatches? What metadata is needed? Identify pros/cons/gaps
- 10:30-10:45 am Break
- 10:45 am - Noon **Breakout Session 3 (Continued): Finalize recommendation ideas in bullet form (including rationale) for supporting Query, Data Aggregation, Analytics, and Data Submission**
- 12:00 -1:00 pm Lunch
- 1:00 – 2:00 pm Final Session (All Together) Report Back from Breakouts
- 2:00 – 2:15 pm Wrap up/ Next steps