#### **PhUSE /FDA Project Request**

1. Project Title:

|  |
| --- |
| Going Translational with Linked Data *[Updated Scope June 2019]* |

1. Workgroup Category:

**Emerging Trends & Technologies**

**Optimizing the Use of Data Standards**

**Standard Analyses and Code Sharing**

**Nonclinical Topics**

**Educating for the Future**

**Data Transparency**

Please State

1. Affected Stakeholders(s):

|  |
| --- |
| Industry and regulatory agency stakeholders responsible for creating, coding, validating, reviewing, and/or consuming both clincial and nonclinical data (SDTM and SEND datasets). |

1. Project Scope:

|  |
| --- |
| This project builds on previous successful modeling of the SDMT domains DM, VS, EX, TS, and AE and conversion instance data to Linked Data as Resource Description Framework (RDF). The original follow-on project was approved to extend the concept to additional domains. At the CSS 2019 conference, the team recognized that the project goals should change to focus on smaller, modular projects that can be released over time to demonstrate incremental value in a shorter time frame (instead of merely extending to additional SDTM domains). The proposed new focus increases cooperation with the Food and Drug Administration and the Maintenance and Support Services Organization (MSSO, for MedDRA).  This new proposal includes (but is not limited to) the following sub-projects:  **a) MedDRA as RDF**  Development and release of R and SAS scripts that convert MedDRA ASCII files to RDF. Due to licensing limtiations, only the MedDRA RDF instance data that supports a subset of the test project (CDISCPILOT01) will be released online. The conversion scripts will be available publicly from a dedicated Github repository that will include documentation and usage instructions. The project team has the interest of and permission from the MSSO for this subject.  **b) Automating Conformance Checks for DM and TS domains in SEND and SDTM**  The team will upate existing SDTM ontology for the TS and DM domains to accomodate SEND. SPIN, SHEX, and SHACL will be evaluated for their potential to model conformance checks for both SEND and SDTM, working in cooperation with the Center for Drug Evaluation and Research, with Dr. Lilliam Rosario as FDA sponsor. The project will evaluate the efficiency of replacing manual conformance checks using an automated Linked Data approach, with potential efficiencies for both FDA and industry.  **c) Unique Identifiers for Pharma**  A sub-group was formed after the PhUSE EUConnect18 to futher develop the concept of "Study URI" presented by Kersin Forsberg and Daniel Goude from AstraZeneca. The group will develop methods and guidance for unique, universal, persistant identifiers for the pharmaceutical industry based on the RDF URI concept. Work is already underway (<https://github.com/phuse-org/LinkedDataEducation/blob/master/doc/URIsForPharma.md>) and a poster is planned for EUConnect19.    **d) Study Ontology & Data Conversion**  Work on Study ontology development may continue as time allows, with primary focus on deliverables for sub-projects a) , b) and c).  Approval of the steering committee will be requested if other sub-projects are identified during the project execution. |

1. Project Dependencies:

|  |
| --- |
| * Participants with expertise in RDF creation and querying, including ontologies and related tools. * Approval from MSSO to obtain MedDRA data for develop. *Obtained Q1 2019.* * Data sets for evaluating SEND conformance criteria |

1. Project Objectives and Timeline:

|  |  |
| --- | --- |
| Project Initiation | Precurrsor: February 2019  ***Restart: July 2019*** |
| MedDRA as RDF | PhUSE CSS 2020 |
| Automating Conformance Checks | PhUSE CSS 2020 |
| Unique IDs for Pharma | PhUSE CSS 2020 |
| Conclusion | PhUSE CSS 2020 |

1. Project Lead(s):

|  |
| --- |
| * Co-lead: Tim Williams, UCB Biosciences * Co-lead: Armando Oliva, Semantica LLC |

1. Project Requestor(s):

|  |
| --- |
| Tim Williams, Armando Oliva |