Statement of Work: RPI Semantics for Imaging Data Curation
RPI PI: Professor Deborah McGuinness

Imaging big data is unique in its inherent high dimensionality, containing a variety of modalities and a deep hierarchy of information embedded in the images. These unique features of imaging data lead to significant logistical and methodological challenges in standardizing and warehousing imaging data. RPI will collaborate on a semantic imaging data resource and provide the semantics and ontology expertise for the proposal.

SA 1. Document, analyze, and benchmark the current state of imaging data curation workflow
The RPI team will review the imaging data curation documentation and collaborate with domain experts to produce a semantic workflow model.

SA2 – Semantify the Imaging Research Warehouse to improve exposure of Radiologic imaging Data Elements (RI-DE) and integrate with other healthcare related data.
RPI will partner with domain experts to build an ontology to cover radiologic imaging data and related data elements. This includes reviewing relevant existing terminologies, identifying mappings and gaps in existing best practice terminologies, and designing an ontology that covers the project needs and is compatible with best practices. RPI will also worth with domain experts to provide semantic extract, transform, and load services. RPI will also work with machine learning partners to represent and ingest machine learning results, maintaining provenance about how the content was determined. RPI will also work to provide tools that suggest appropriate mappings aimed at reducing manual effort requirements for harmonization.

SA3 – Model and Test a new data curation paradigm. The new paradigm provides a prospectively built database as a knowledge resource fully prepared for query with knowledge inherently exposed. It is designed for easy collaboration and pooling of data, with audit tools to document subsequent annotation, versioning and updates regardless of the data source. RPI will build knowledge graphs and provide APIs to query this data model thereby providing increased access to the local and properly credentialed extended research communities**.** Our partners will provide metrics for use in evaluation and RPI will model the metrics and provide algorithms to evaluate based on these metrics.
.
All ontologies, data products, and algorithms will be open source and available for reuse.