Title: *CIF21 DIBBs: PD: Ontology-enabled Polymer Nanocomposite Open Community Data Resource*
Principal Investigator: Linda Schadler, Deborah L. McGuinness, Rensselaer Polytechnic Institute
Award ID: 1640840 Cate Brinson, Wei Chen, Northwestern University

Polymer nanocomposites exhibit unique combinations of properties, which provide exciting opportunities for advanced applications in energy, medicine, optics, and more. A significant limitation to meeting this promise is poor quantitative understanding of the **p**rocessing-**s**tructure-**p**roperty (p-s-p) relationships. A robust cyberinfrastructure is needed to support a data centric approach to fundamental understanding and materials design. The goal of this project is to build on a prototype data resource for polymer nanocomposites (NanoMine), and create an open access, user friendly, living, growing, knowledge graph for the polymer nanocomposites community to facilitate the development of p-s-p relationships and faster nanocomposite design and insertion into advanced applications.

Significant Challenges and Solutions:
A significant challenge in the field of nanocomposites is the high dimensionality and limited knowledge of causal relationships among diverse p-s-p parameters. Because the data is scattered throughout public literature and private files, the data is incomplete, inconsistently recorded and tagged, and not directly accessible. To support the development of both fundamental understanding and design, a robust ontology is required to facilitate improved access, interoperability, and reuse of polymer nanocomposite knowledge. Our starting point is an XML schema tailored for experimental and computational nanocomposites with elements for metadata, material constituents, processing conditions, microstructure, and property response. We are expanding to an RDF/XML schema that includes extensions for custom data and support for versioning so that existing documents can be automatically updated to the current schema without data loss or manual intervention. We are developing an ontology based on this schema that uses existing standards to encode uploaded data and results as manageable units of experimentally or computationally supported knowledge, called Nanopublications. We use Dublin Core Terms and the BIBO ontology for publication metadata, the SemanticScience Integrated Ontology to represent scientific knowledge, and existing chemical databases like PubChem and PolyInfo to identify chemical entities. The ontology will enable greater interoperability with existing content and it is being used to provide the terminology definitions for a knowledge graph.

A second challenge is recruiting a community of users to NanoMine in order to create a larger data set. We are using four approaches: First, we developed a prototype tool that makes it easier to upload data to NanoMine. Second, we will leverage open source visualization tools and provide semantically enhanced browsing and visualization tools to aid users in discovery of new relationships. Third, we continue to develop tools to allow the user to, for example: characterize their nanofiller dispersion, predict dispersion, and model properties such as glass transition temperature or dielectric breakdown strength. Fourth, advanced features will be implemented to improve user-friendliness, such as: the ability to handle complex conditional data queries, interactive visualization of query results, and data analysis of correlations among parameters and descriptors for concept development, design recommendation and knowledge discovery. A final challenge is developing quality and uncertainty metrics for the data integrated with the ontology along with an automated/semi-automated curation and rigorous evaluation process to validate crowdsource input before insertion into NanoMine.