Fraunhofer Institute for Algorithms and Scientific Computing SCAI

- Expertise: Materials Modelling
  - Department: Virtual Materials Design)
    (headed by Jan Hamekers)
    - Multiscale modeling, numerical simulation and high performance computing in materials science, nanotechnology and chemistry
    - High-dimensional problems, optimization, data analysis, machine learning, Uncertainty Quantification (UQ)
    - Computer-aided materials and molecular design, Integrated Computational Materials Engineering
- Expertise: Interoperability (ITEA Project VMAP)
    - VMAP - A new Interface Standard for Integrated Virtual Material Modelling in Manufacturing Industry (Coordinated by Klaus Wolf)
- Expertise: Ontology Generation and Semantic Services
  - Department Bioinformatics: see following slides
    (headed by Martin Hoffmann-Apitius)

Competency in Ontology Generation and Semantic Services at Fraunofer SCAI

Martin Hofmann-Apitius

Fraunhofer SCAI:
a track record in applied semantics

- Fraunhofer SCAI has a long-standing track record in applied semantics. This becomes evident, when we analyse the publications of the institute in this context (the following is a selection, not the full publication record in this area!):
  - Ivchenko, O., Younesi, E., Shahid, M., Wolf, A., Müller, B., & Hofmann-Apitius, M. (2011). PLIO: an ontology for formal description of protein–ligand interactions. *Bioinformatics*, *27*(12), 1684-1690.
  - Gündel, M., Younesi, E., Malhotra, A., Wang, J., Li, H., Zhang, B., ... & Hofmann-Apitius, M. (2013). HuPSON: the human physiology simulation ontology. Journal of biomedical semantics, 4(1), 35.
  - Malhotra, A., Younesi, E., Gündel, M., Müller, B., Heneka, M. T., & Hofmann-Apitius, M. (2014). ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. Alzheimer's & dementia: the journal of the Alzheimer's Association, 10(2), 238-246.
  - Malhotra, A., Gündel, M., Rajput, A. M., Mevissen, H. T., Saiz, A., Pastor, X., ... & Kotelnikova, E. (2015). Knowledge retrieval from PubMed abstracts and electronic medical records with the Multiple Sclerosis Ontology. PloS one, 10(2), e0116718.

Fraunhofer SCAI: tools for the generation of ontologies and terminologies

- Fraunhofer SCAI has substantial expertise in applied text mining. Terminologies and ontologies play a central role for text mining workflows:
  - Ontologies are being created after simple analyses of the way, how we communicate knowledge about a topic. Amongst these simple analyses are n-gram analyses and other methods of term-frequency analysis
  - Ontologies are created using protégé (the Stanford University knowledge lab editor); we usually map our domain-specific ontologies to widely adopted upper level ontologies
  - Ontologies and terminologies generated are “validated” using application scenarios: we test them for their usability in document retrieval, in text mining and information extraction scenarios. A domain-specific “ground truth” provides the gold standard for the assessment
  - We organise “jamborees” with domain experts to get feedback on the initial draft of a new ontology. We do not favour automated methods for ontology construction (so-called “ontology learning”), as these fail usually to be accepted by domain experts and users
  - Fraunhofer SCAI uses OLS (ontology lookup service, an open source - service developed and maintained at the European Bioinformatics Institute (EBI)) as the central administration and provenance tool for ontologies and terminologies

Fraunhofer SCAI: applied semantics

- Fraunhofer SCAI is using ontologies and terminologies to generate computable, knowledge-based models of a given scientific context:
  - Ontologies provide the “semantic invariables” to refer to in modelling approaches based on RDF (resource description framework) and specialised syntax for knowledge-based modelling such as OpenBEL (www.openbel.org)
  - In brief: one of the “side effects” of ontologies is the fact, that the terms in an ontology become part of a “controlled vocabulary”. We can disambiguate concepts and terms; we can define objects that carry a unique name and a unique identifier. That makes them ideal subjects to refer to, if we want to clarify, what we talk about
  - RDF and OpenBEL are “triple-based languages”, they allow to represent knowledge in the form of “subject-predicate-object” (or “object – relationship – object”) triplets. We can call these triples “knowledge atoms”. Smallest knowledge unit we can think of. Other people call them “nano-publications”.
  - Nano-Publications can be used to represent “all we know” about a domain; that domain can be homogeneous or heterogeneous, simple or complex. Algorithms that run on triple-encoded large knowledge graphs (such as “reverse causal reasoning”) allow to use the knowledge graph for the computing of the concordance between patterns in measured data and the computable representation of what we already know in a given domain