



NUI Galway
OÉ Gaillimh



Digital Enterprise Research Institute



A Roadmap for Navigating the Life Sciences Linked Open Data Cloud

Ali Hasnain, Ronan Fox, and Helena Deus
{ali.hasnain,ronan.fox,helena.deus}@deri.org

Introduction

Navigating through Open Life Sciences Datasets in the Linked Open Data (LOD) Cloud is difficult as most datasets are available only as isolated SPARQL endpoints and they do not a schema or a formalism to aid in the assembly of queried.

We have devised an active roadmap for navigating the linked life sciences cloud that illustrates all the possible "roads" or "links" between concepts in the LOD cloud.

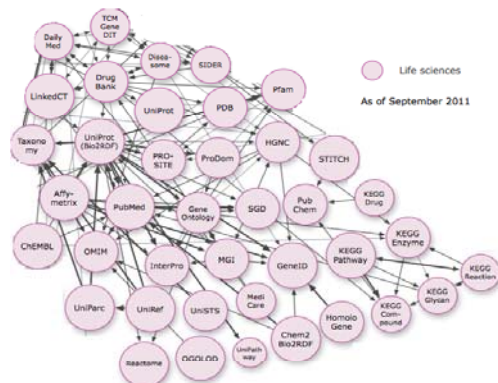


Fig 1: The Life Sciences Linked Open Data Cloud consists of a collection of biological datasets where entities are interlinked

Motivation

- Meaningful biological correlations for personalized medicine.
 - Improvement in drug discovery and development process.
 - Entity identification
 - Protein/Drug, Drug/Drug or Protein/Protein interactions.
- "Drug" data from different overlapping sources can be seen in fig 2.

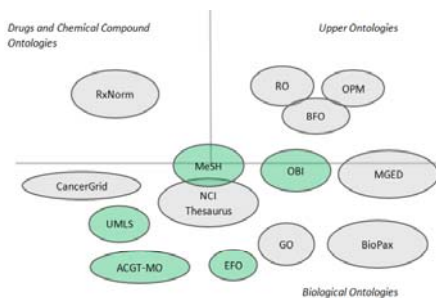


Fig 2: The number of "ontologies" available for representing biological knowledge is abundant. However, these remain disconnected from the data and knowledge they were built to represent

Method

Our method is divided in two main phases:

1. **Catalogue Development:** Data collection, analysis and arrangement.
2. **Roadmap Development:** Link creation based on Syntactic, Semantic and Domain matching of concepts in the catalogue as can be seen in fig 3.

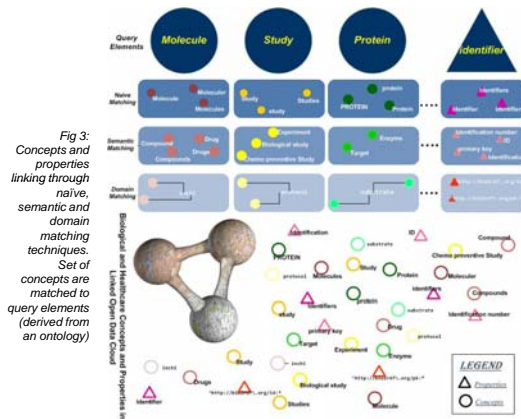


Fig 3: Concepts and properties linking through naïve, semantic and domain matching techniques. Set of concepts are matched to query elements (derived from an ontology)

Results

Technique	Results
Syntactic Matching	<http://bio2rdf.org/ns/biopax#Protein> <http://www.w3.org/2000/01/rdf-schema#subClassOf> <http://bio2rdf.org/biopax_resource:Protein>
Semantic Matching	<http://www4.wiwiw.fu-berlin.de/drugbank/resource/drugbank/targets> <http://www.w3.org/2000/01/rdf-schema#subClassOf> <http://bio2rdf.org/biopax_resource:Protein>
Domain Matching	<http://bio2rdf.org/ns/kegg#Enzyme> <http://www.w3.org/2000/01/rdf-schema#subClassOf> <http://bio2rdf.org/biopax_resource:Protein>

Table 1: Results- Example links based on syntactic, semantic and domain linking.

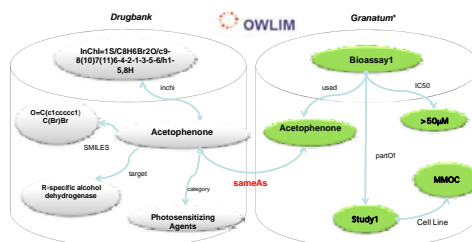


Fig 4: Our methodology enables the immediate integration of biological datasets from different sources, even if their identifiers are not reused

Acknowledgement

* This work has been funded by an EU- project "Granatum" which is partially funded by the European Commission under the Seventh Framework Programme in the area of Virtual Physiological Human (ICT-2009.5.3).



ENTERPRISE IRELAND



Enabling Networked Knowledge