Erik B. Myklebust[1,2], Ernesto Jimenez-Ruiz[2,3], Jiaoyan Chen[4], Raoul Wolf[1] and Knut Erik Tollefsen[1,5]

[1] Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, 0349 Oslo, Norway

[2] Department of Informatics, University of Oslo, Gaustadalléen 23B, 0373 Oslo, Norway

[3] City, University of London, Northampton Square, Clerkenwell, London EC1V 0HB, United Kingdom

[4] Department of Computer Science, University of Oxford, 15 Parks Rd, Oxford OX1 3QD, United Kingdom

[5] Norwegian University of Life Sciences (NMBU), Universitetstunet 3, 1433 Ås, Norway

Semantic Data Access for Toxicological Research

Background

Introduction. Environmental risk assessment uses large amounts of disparate datasets. Therefore, integrating source datasets into a unified data structure is desirable. In this study, we leverage the suite of Semantic Web tools to achieve such a coherent integration.

Methods. We integrate chemical effect concentration data from ECOTOX with external sources, which include more extensive information about chemicals and species used in ECOTOX. These sources are in different formats, and we transforms them into the Resource Description Framework (RDF), a standard of the Semantic Web, called the Toxicological Effects and Risk Assessment Knowledge Graph (TERA). TERA enables the interoperability between sources and the use of the suite of Semantic Web tools, such as query and reasoning engines.

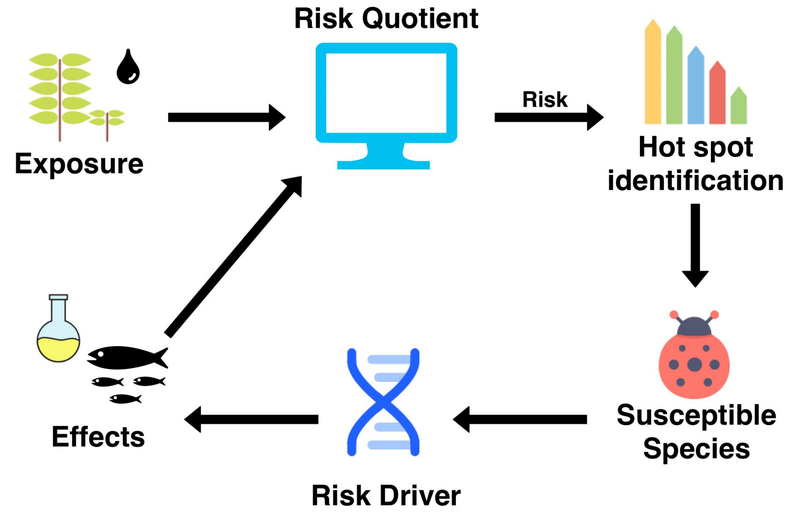

Figure 1. An ecological risk assessment pipeline.

Data Integration

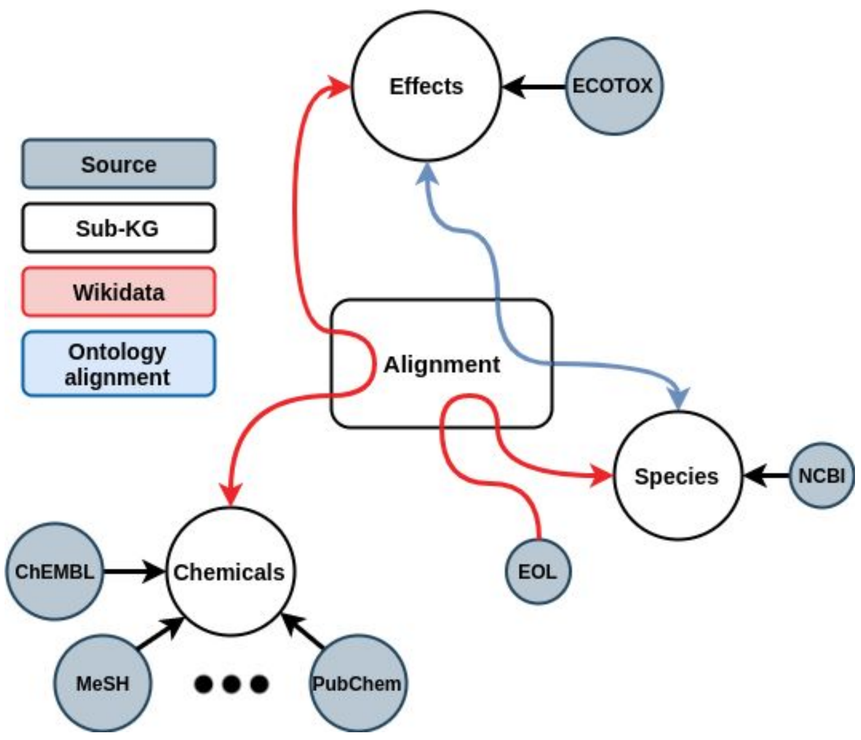

Figure 2. The interconnected parts of TERA.

Figure 2 shows how the data sources are integrated into the TERA knowledge graph.

Effects data is transformed from the tabular U.S. EPAs ECOTOXicological knowledge base into RDF. The chemical identifiers in ECOTOX are aligned to ChEMBL using Wikidata while the taxon identifiers are aligned to the NCBI Taxonomy using the alignment tool LogMap. Encyclopedia of Life traits data is imported and aligned with NCBI using Wikidata. MeSH and PubChem are imported and aligned to ChEMBL using Wikidata.

The integrated dataset and tools for data access are available at https://github.com/NIVA-Knowledge-Graph/TERA .

Example

Query 1 extracts all tests where the chemical used is a neonicotinoid insecticide and the test organism weighs less than 5 gram (a proxy for maximum arthropod size) and is a pollinator. A visualization of Query 1 is seen in Figure 3. The colours indicate the paths of the query in TERA.

```
SELECT ?test
WHERE {
   ?test chemical [ type  Neonicotinoid ] ;
            organism [ pollinates [ type Plant ]  ] ;
            organismWeight [ value ?ow ;
                             unit Gram ] .
   FILTER (?ow < 5)
}
```

Query 1.

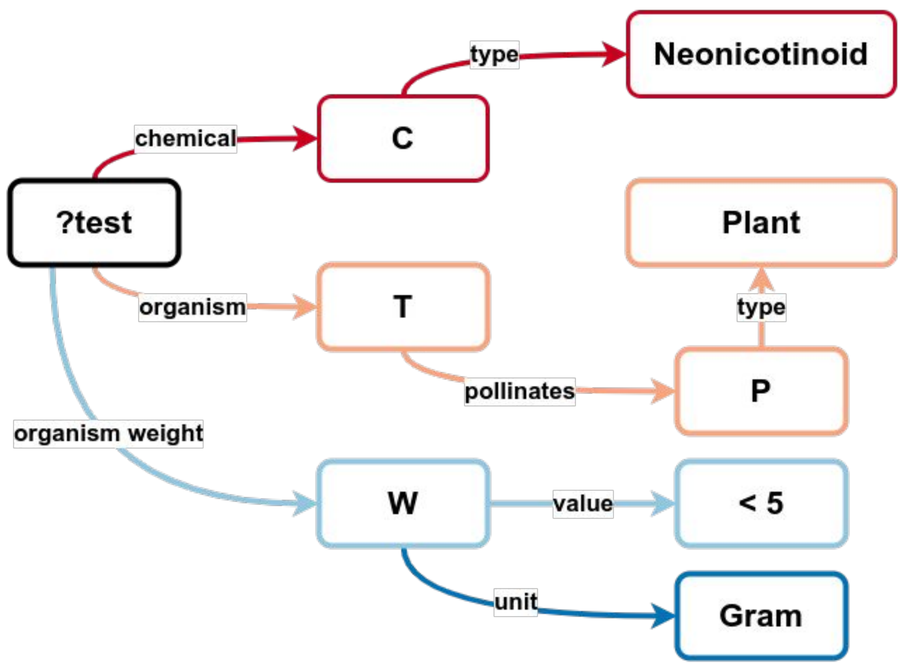

Figure 3. Flow chart corresponding to Query 1.

Future perspectives

Semantic Web tools are currently used by only a limited number of researchers. Therefore, testing and validation from outside this core group is needed. The new user groups can provide valuable insight into new features and use cases for the tools presented here.

Contact info:
Norwegian Institute for Water Research (NIVA), Gaustadalléen 21, N-0349 OSLO, Norway
Email: ebm@niva.no ; ket@niva.no

This work is supported by grant 272414 from the Research Council of Norway (RCN), the MixRisk project (RCN 268294), and the SIRIUS Centre for Scalable Data Access (RCN 237889). The work is organised under NIVA's Computational Toxicology Program (NCTP).

With funding from The Research Council of Norway