A Knowledge Graph for Ecotoxicological Risk Assessment and Effect Prediction

Erik B. Myklebust

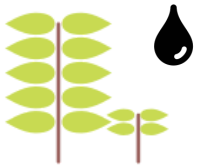
Norwegian Institute for Water Research, Oslo, Norway
Department of Informatics, University of Oslo, Oslo, Norway

Funding by Research Council of Norway (272414).
Supervised by Ernesto Jimenez-Ruiz (U. of Oslo; City, U. of London), Jiaoyan Chen (U. of Oxford), Raoul Wolf (NIVA), Knut Erik Tollefsen (NIVA; NBMU) & Martin Giese (U. of Oslo).

ErikBMyklebust
ebm@niva.no

Norwegian Institute for Water Research

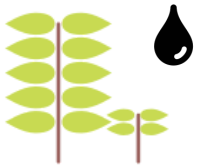
Ecological Risk Assessment



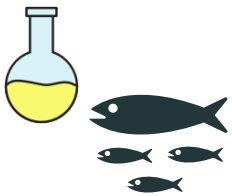
Exposure

Risk assessment is an estimation of cumulative risk on individuals, populations, communities, and ecosystems from chemical pollutants.

Ecological Risk Assessment



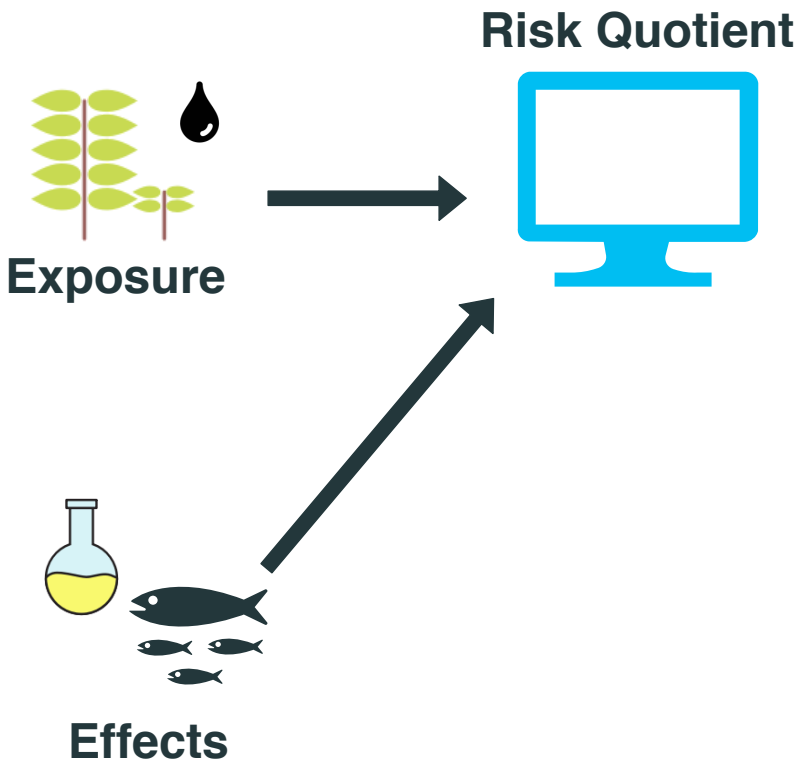
Exposure



Effects

Effect concentrations are found using organism experiments.

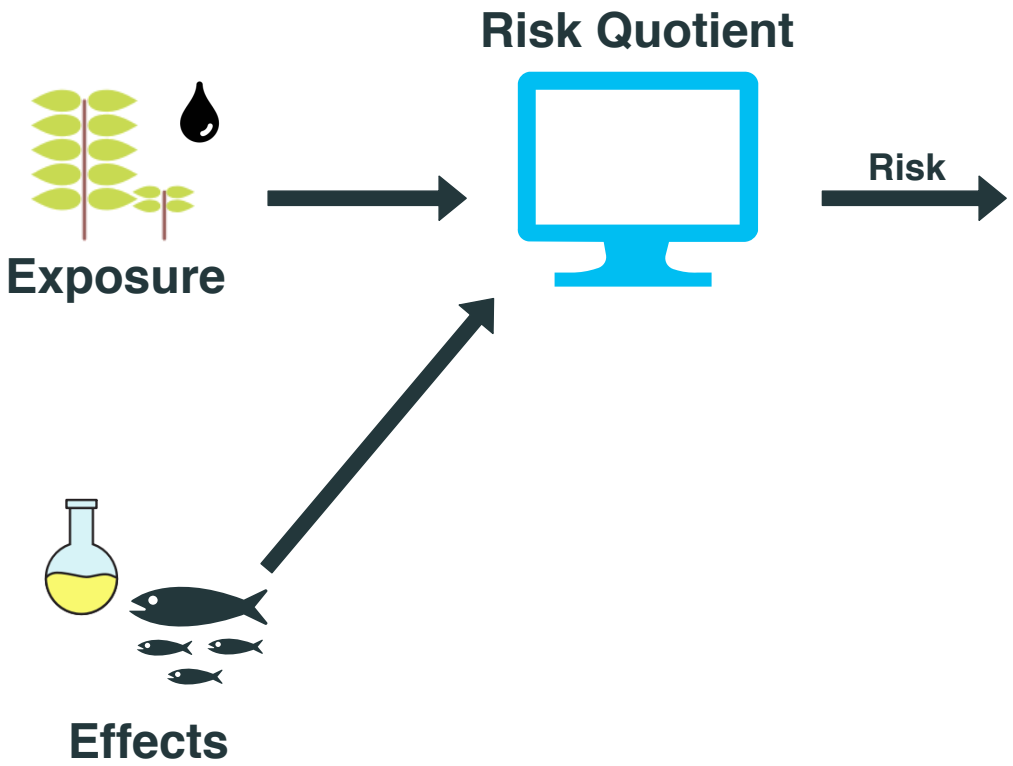
Ecological Risk Assessment



$$RQ = \frac{\text{environmental concentration}}{\text{effect concentration}}$$

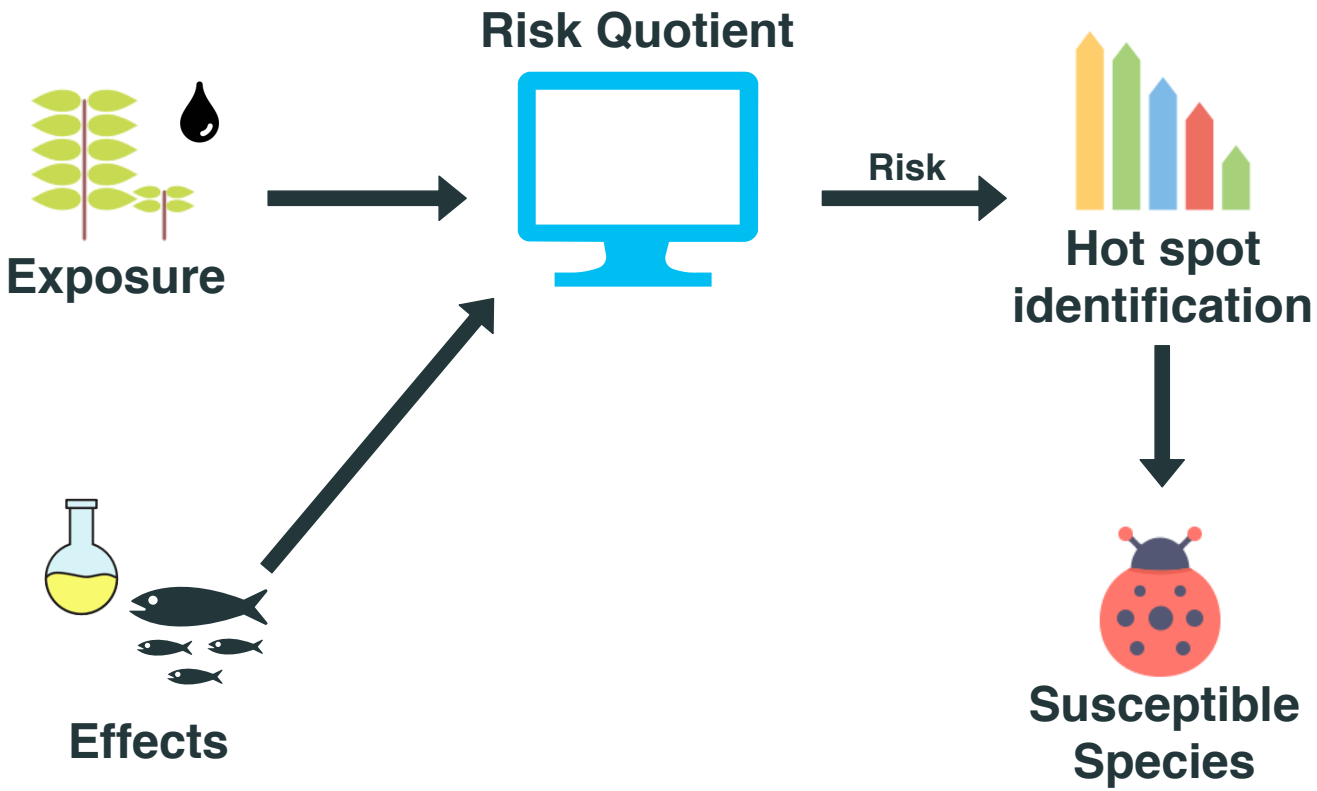
RQs coverage is limited by effect concentration experiments.

Ecological Risk Assessment



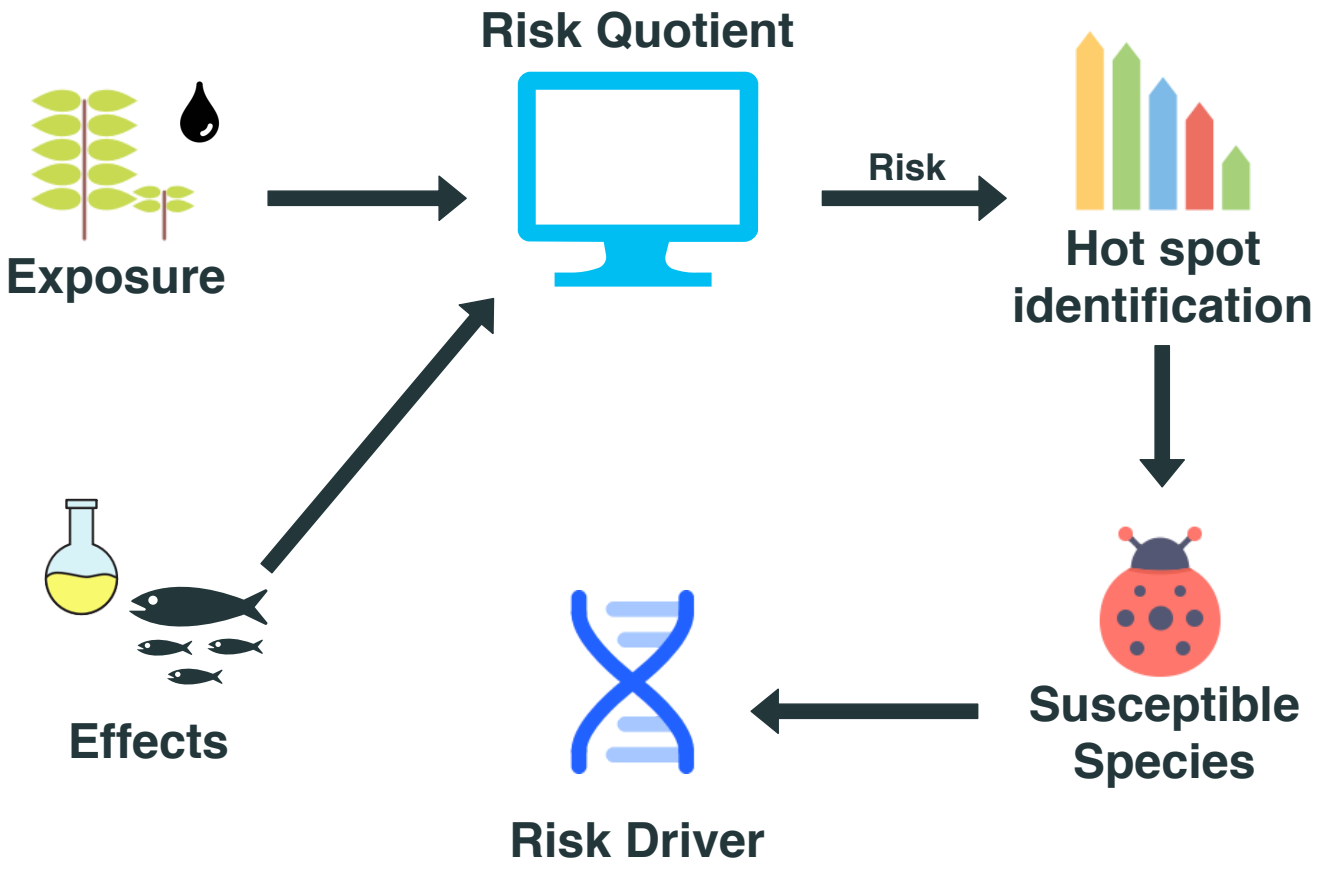
$$\text{risk}_{\text{group}} \approx \sum^{\textit{chemicals}} RQ$$

Ecological Risk Assessment



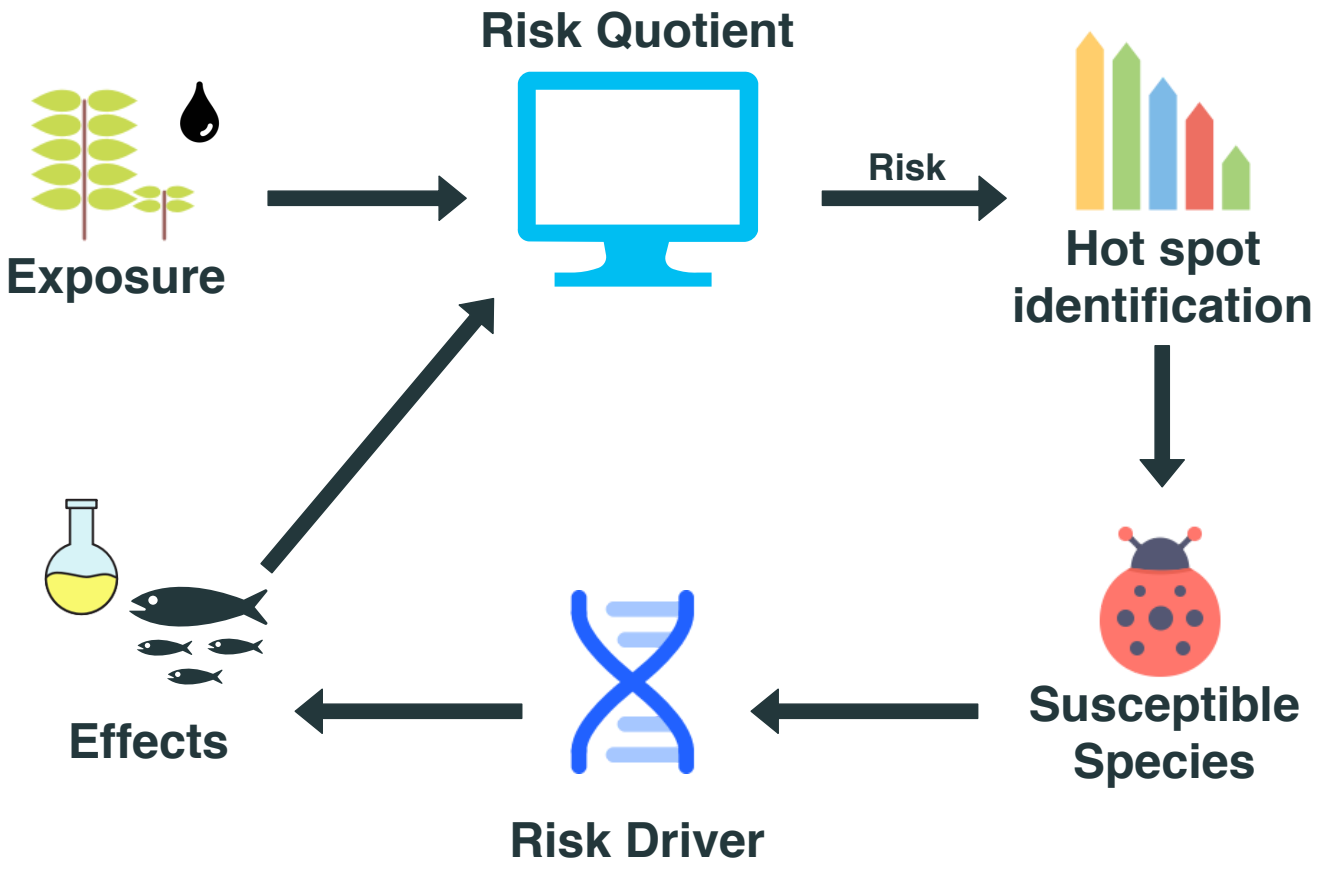
The risk is used to find further susceptible species.

Ecological Risk Assessment



Risk driver describes *how* the chemical affects an organism.

Ecological Risk Assessment



New effect hypothesis are then tested in the laboratory.

Project Goals

1. Construct a knowledge graph that integrates the relevant knowledge for ecological risk assessment.

Project Goals

1. Construct a knowledge graph that integrates the relevant knowledge for ecological risk assessment.
2. Using the knowledge graph to predict ecotoxicological effects.

Project Goals

1. Construct a knowledge graph that integrates the relevant knowledge for ecological risk assessment.
2. Using the knowledge graph to predict ecotoxicological effects. This task is two fold:

Project Goals

1. Construct a knowledge graph that integrates the relevant knowledge for ecological risk assessment.
2. Using the knowledge graph to predict ecotoxicological effects. This task is two fold:
    2.1. Limit the search space for laboratory experiments (binary prediction).

Project Goals

1. Construct a knowledge graph that integrates the relevant knowledge for ecological risk assessment.
2. Using the knowledge graph to predict ecotoxicological effects. This task is two fold:
    2.1. Limit the search space for laboratory experiments (binary prediction).
    2.2. Predict effect concentrations (regression).

Research Questions

Research Questions

RQ1. Can the disparate data sources used in ecotoxicological risk assessment be integrated into a knowledge graph to improve accessibility?

Research Questions

RQ1. Can the disparate data sources used in ecotoxicological risk assessment be integrated into a knowledge graph to improve accessibility?

RQ2. Can the knowledge graph be used to improve (or diversify) ecotoxicological effect prediction over current state-of-the-art models?

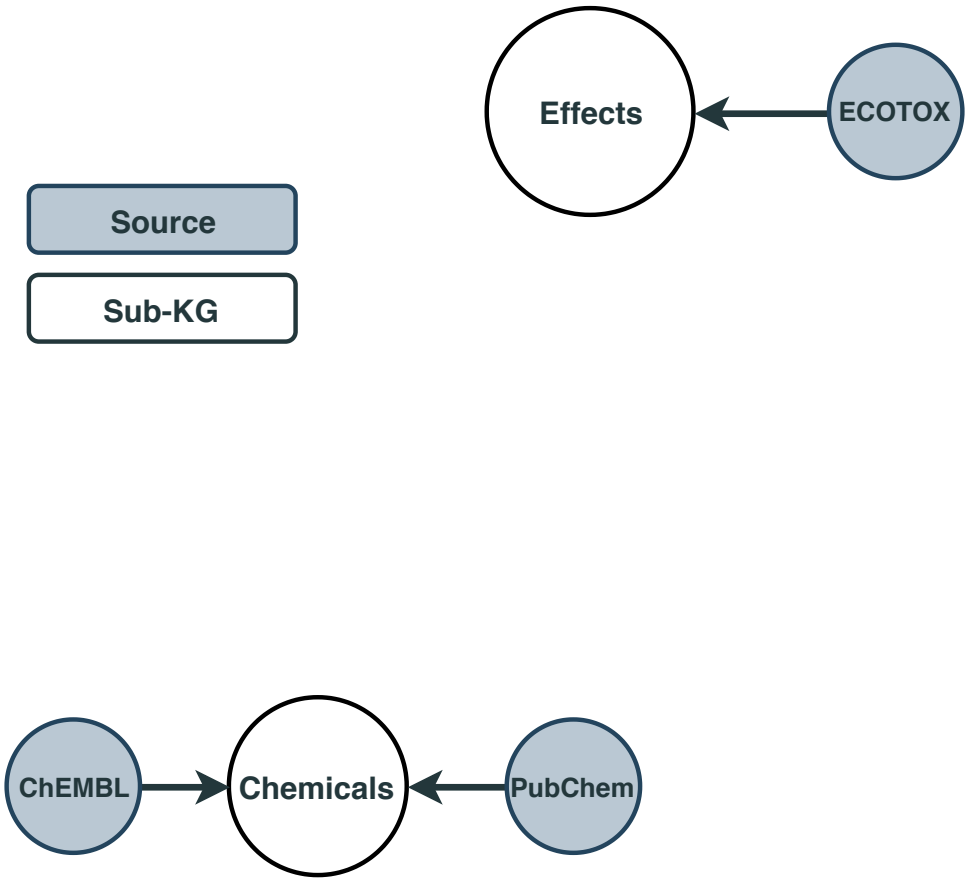
The TERA Knowledge Graph

The Toxicological and Risk Assessment (TERA) knowledge graph integrates data sources varying in format.

The TERA Knowledge Graph

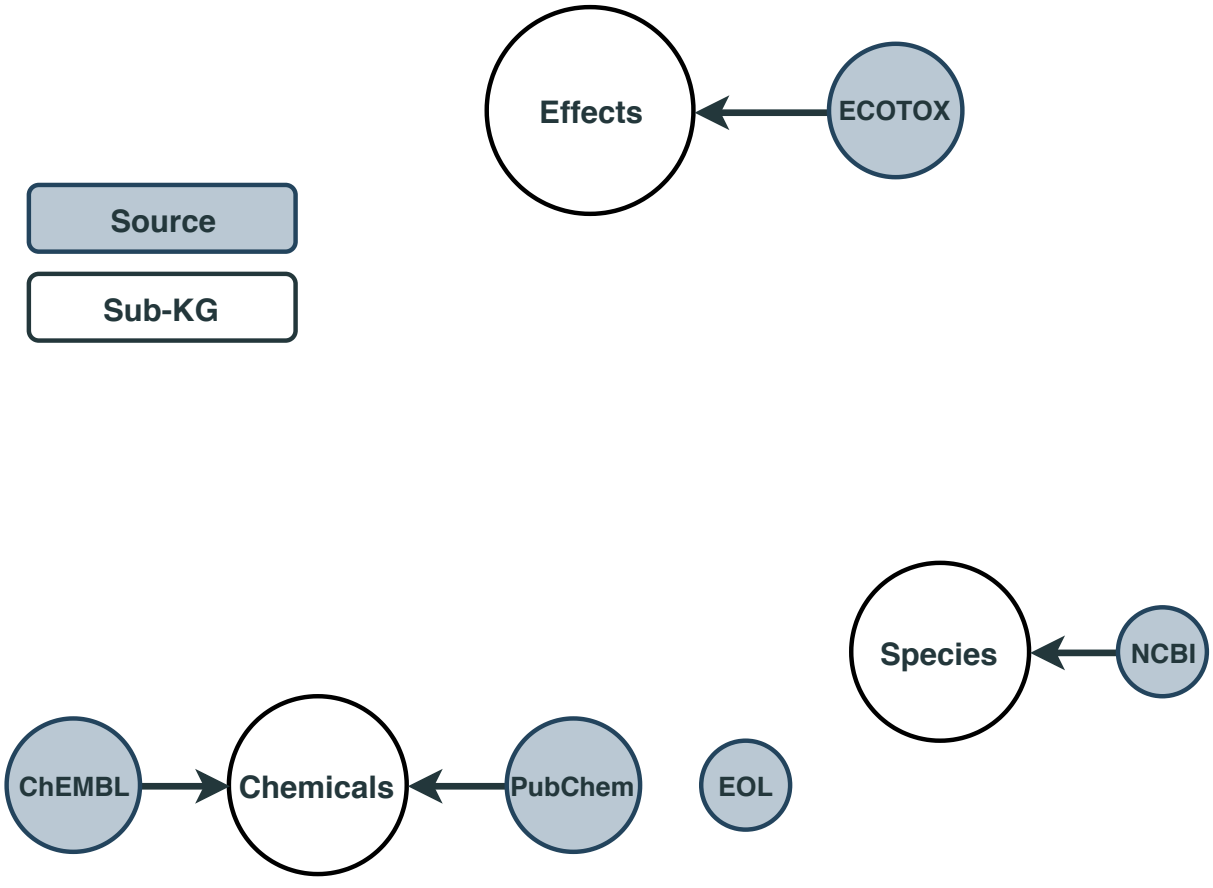
ECOTOX is the largest (public) source of effect data.

The TERA Knowledge Graph



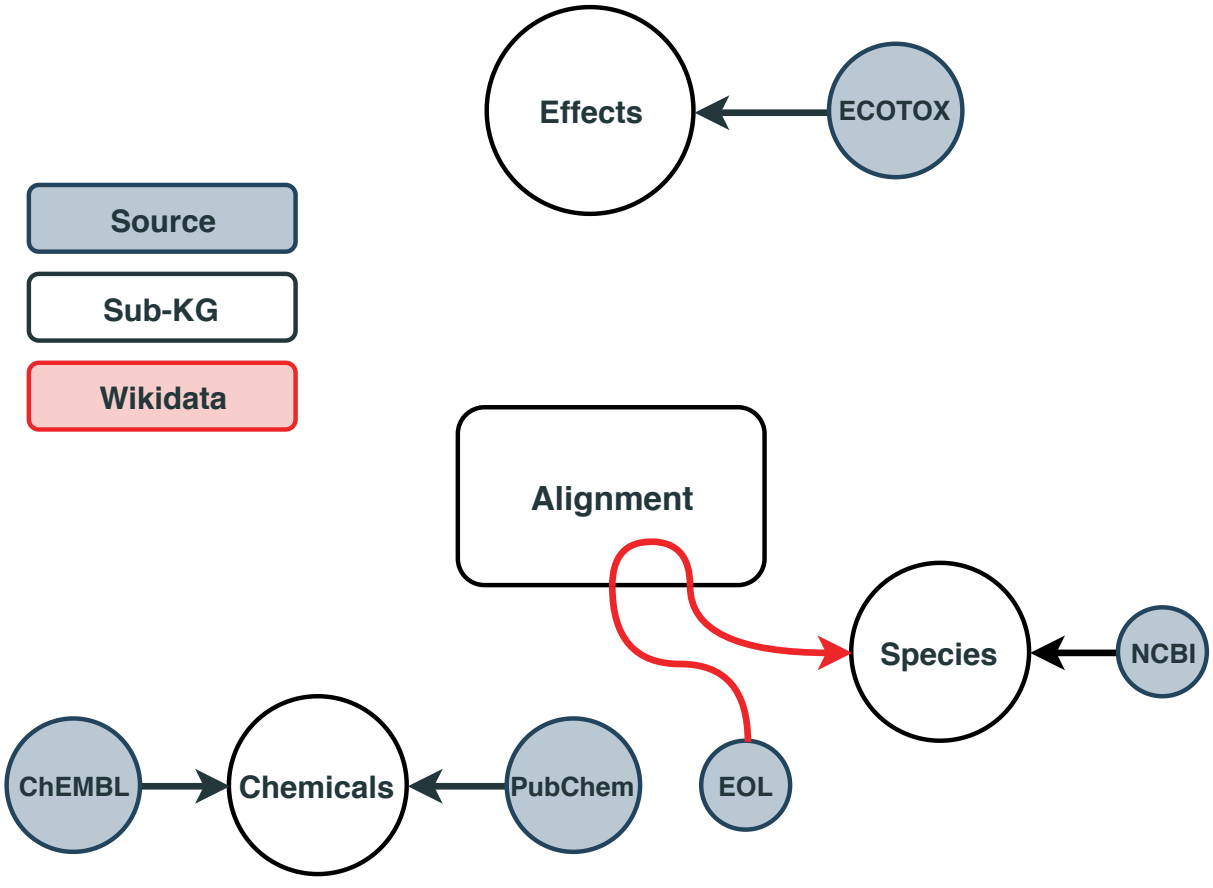
Importing the ChEMBL and PubChem knowledge graph.

The TERA Knowledge Graph



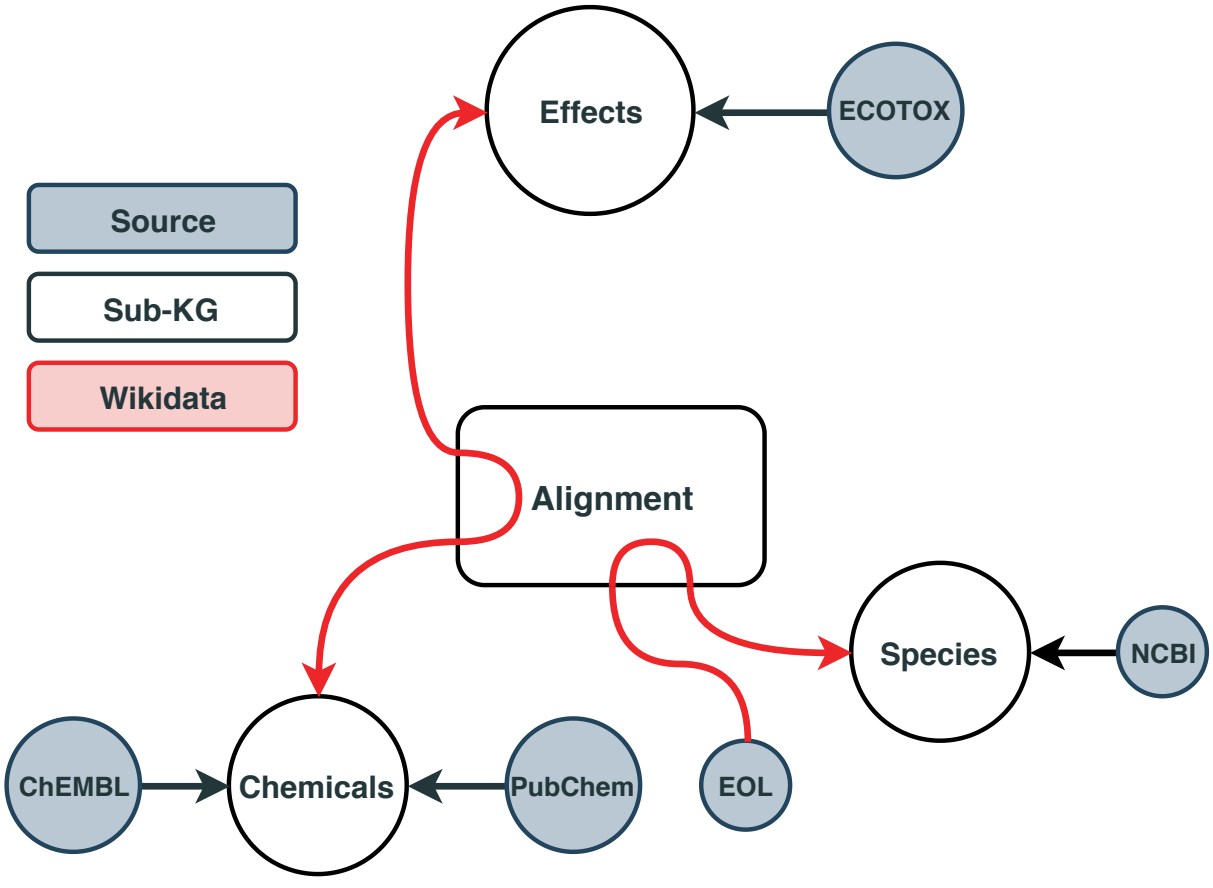
NCBI's tabular taxonomy and EOL traits are converted to a hierarchy.

The TERA Knowledge Graph



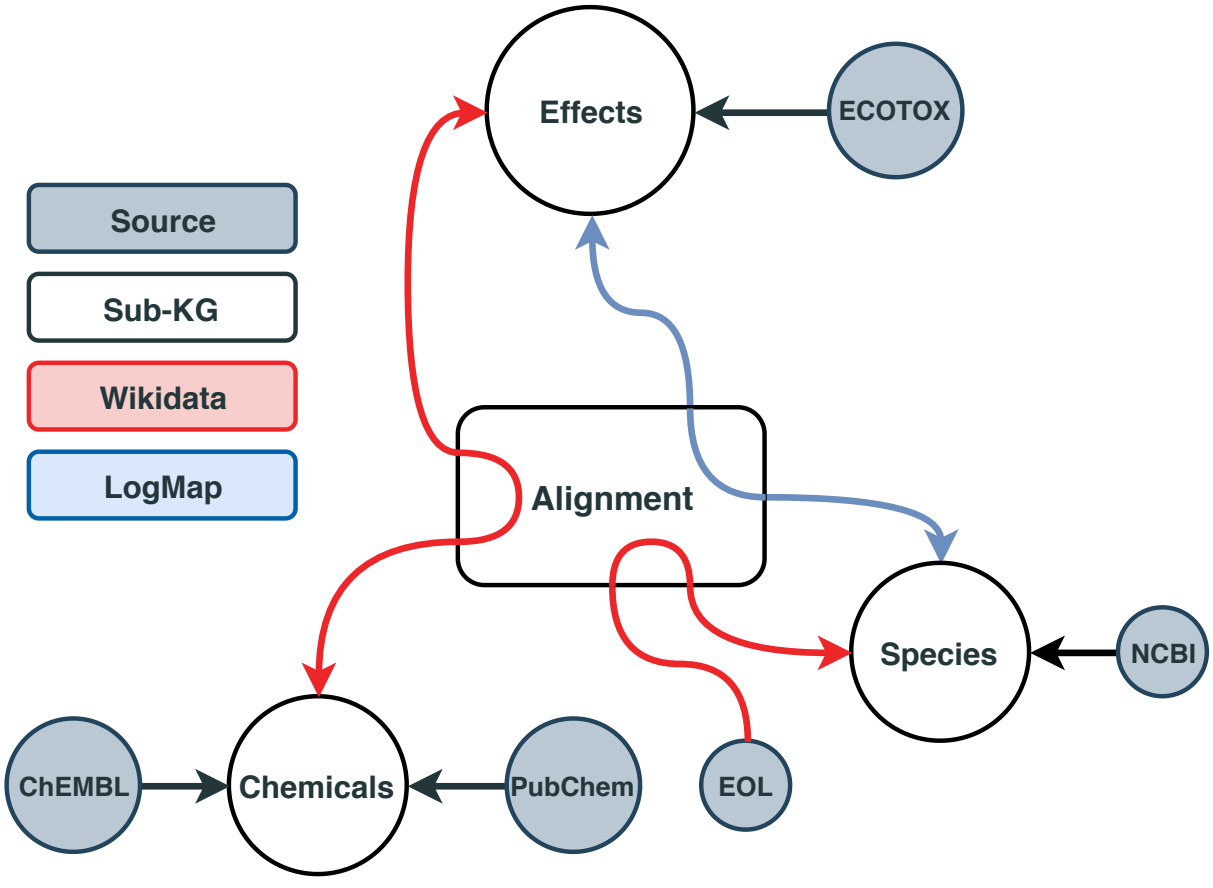
Encyclopedia of Life (EOL) is aligned to NCBI through Wikidata.

The TERA Knowledge Graph



Aligning proprietary chemical identifiers in ECOTOX to open identifiers in PubChem.

The TERA Knowledge Graph



Aligning taxonomies using ontology alignment tool LogMap.

Binary Effect Prediction Problem Definition

Binary Effect Prediction Problem Definition

C_1

C_2

C_3

Chemicals

Binary Effect Prediction Problem Definition

C_1

S_1

C_2

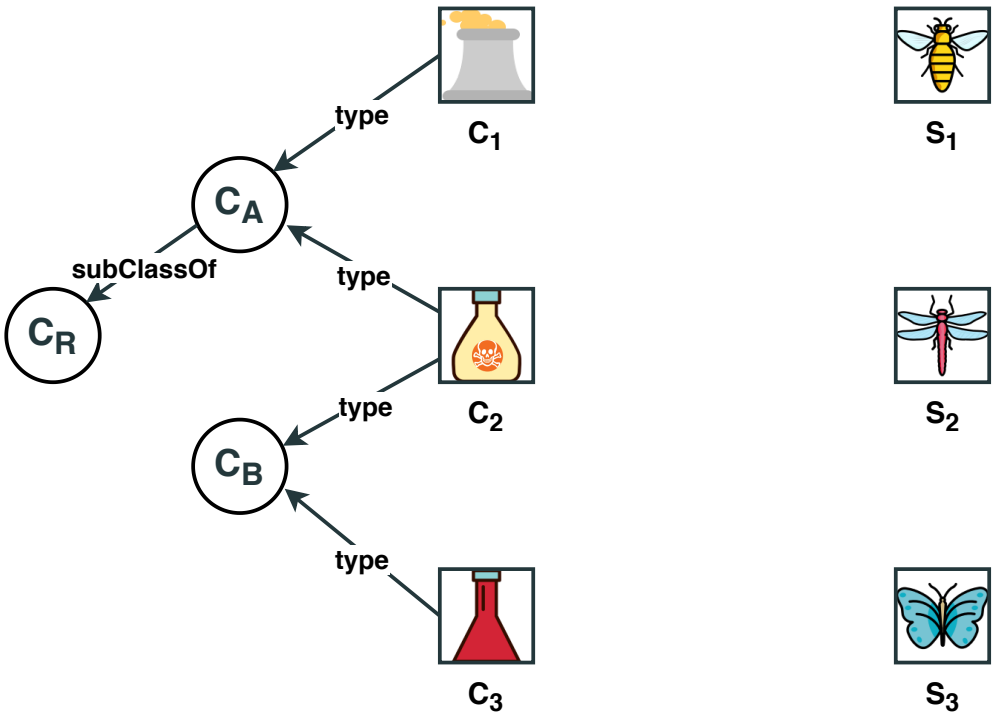
S_2

C_3

S_3

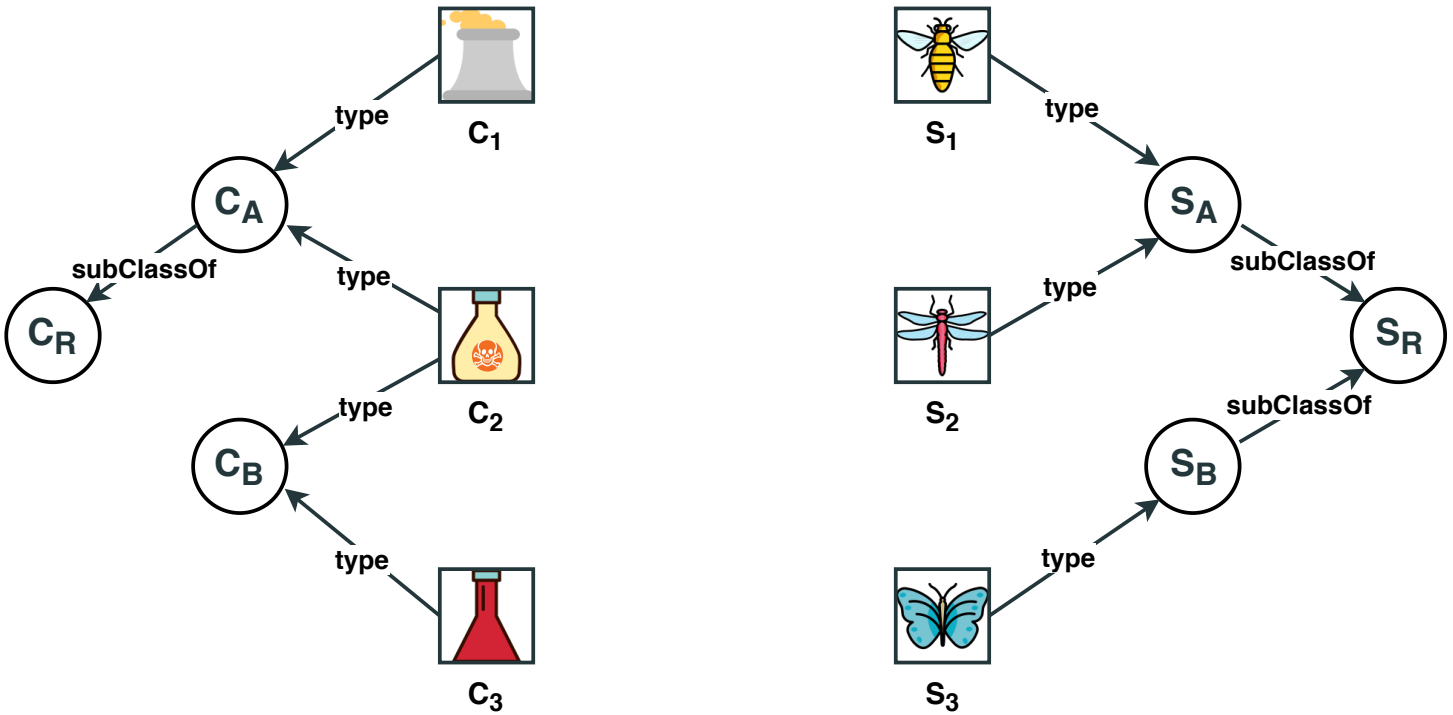
Species

Binary Effect Prediction Problem Definition



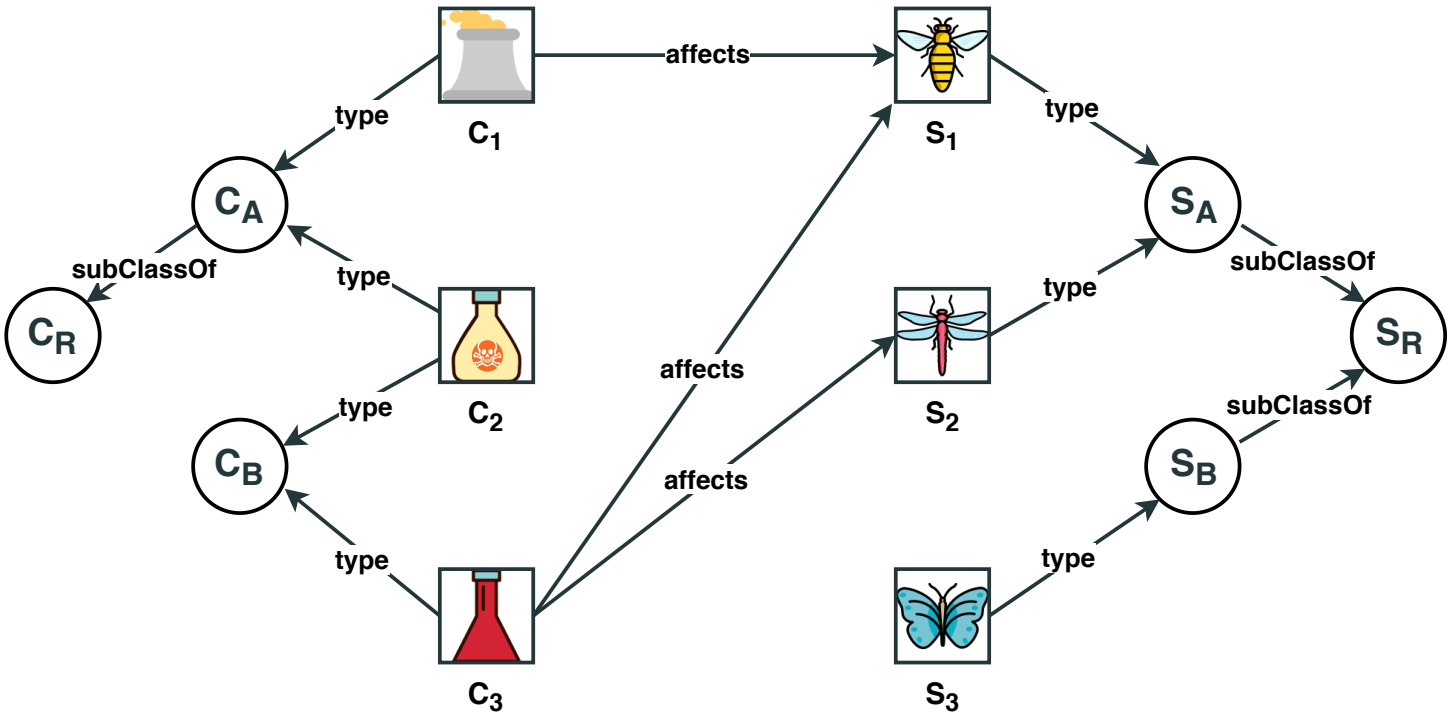
Chemical classification

Binary Effect Prediction Problem Definition



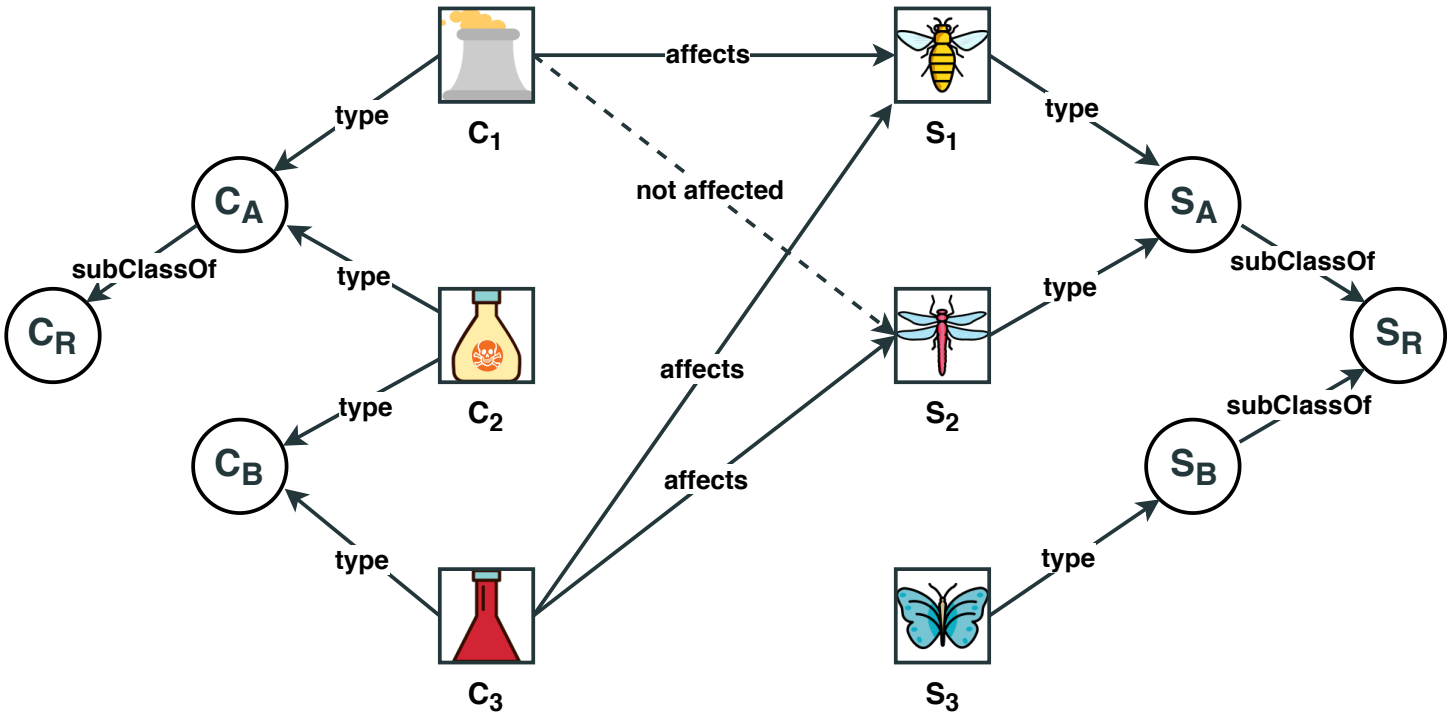
Taxonomy

Binary Effect Prediction Problem Definition



Positive samples

Binary Effect Prediction Problem Definition



Negative samples

Three models

In contrast to commonly used QSAR models, we aim at creating complete domain covering models.

Three models

In contrast to commonly used QSAR models, we aim at creating complete domain covering models.

Taxonomic distance baseline (BL)
Inherit effects from nearest neighbours.

Three models

In contrast to commonly used QSAR models, we aim at creating complete domain covering models.

Taxonomic distance baseline (BL)
Inherit effects from nearest neighbours.

Multi-layer perceptron (MLP)
The model use one-hot encodings of species and chemicals as input.

Three models

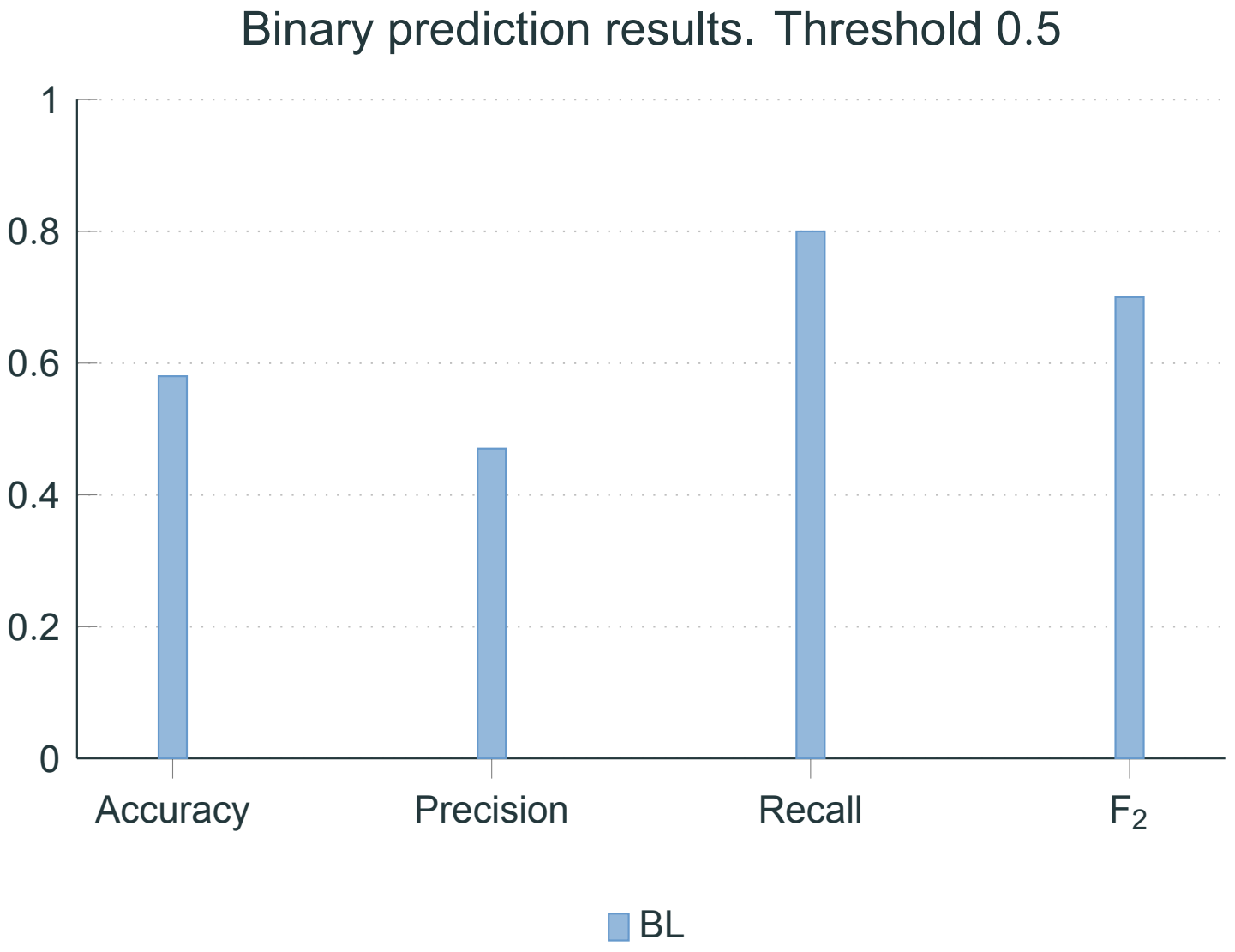
In contrast to commonly used QSAR models, we aim at creating complete domain covering models.

Taxonomic distance baseline (BL)
Inherit effects from nearest neighbours.

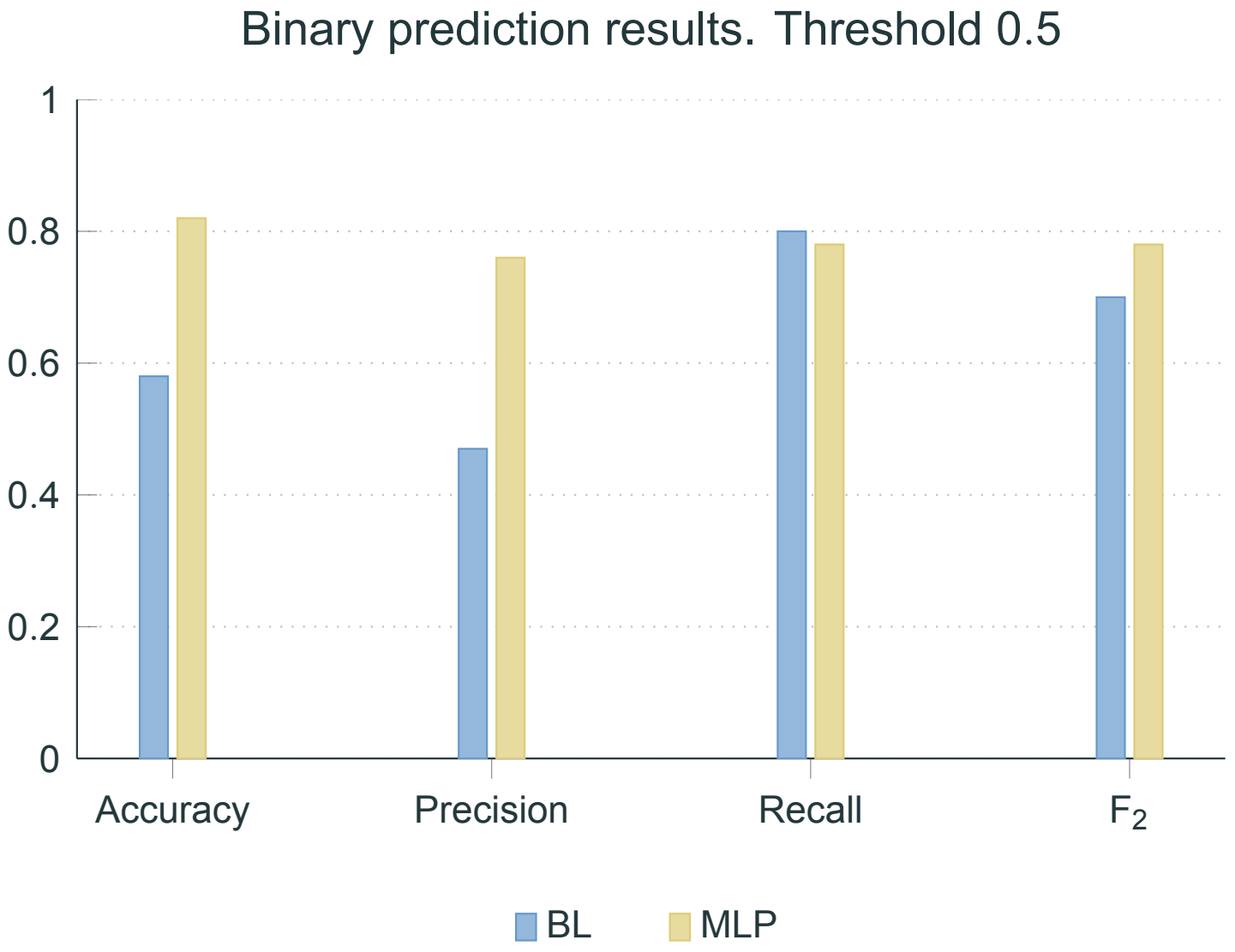
Multi-layer perceptron (MLP)
The model use one-hot encodings of species and chemicals as input.

Knowledge graph embedding (KGE)
Replacing the input layers in MLP with a KGE model (*e.g.*, TransE, DistMult, HolE).

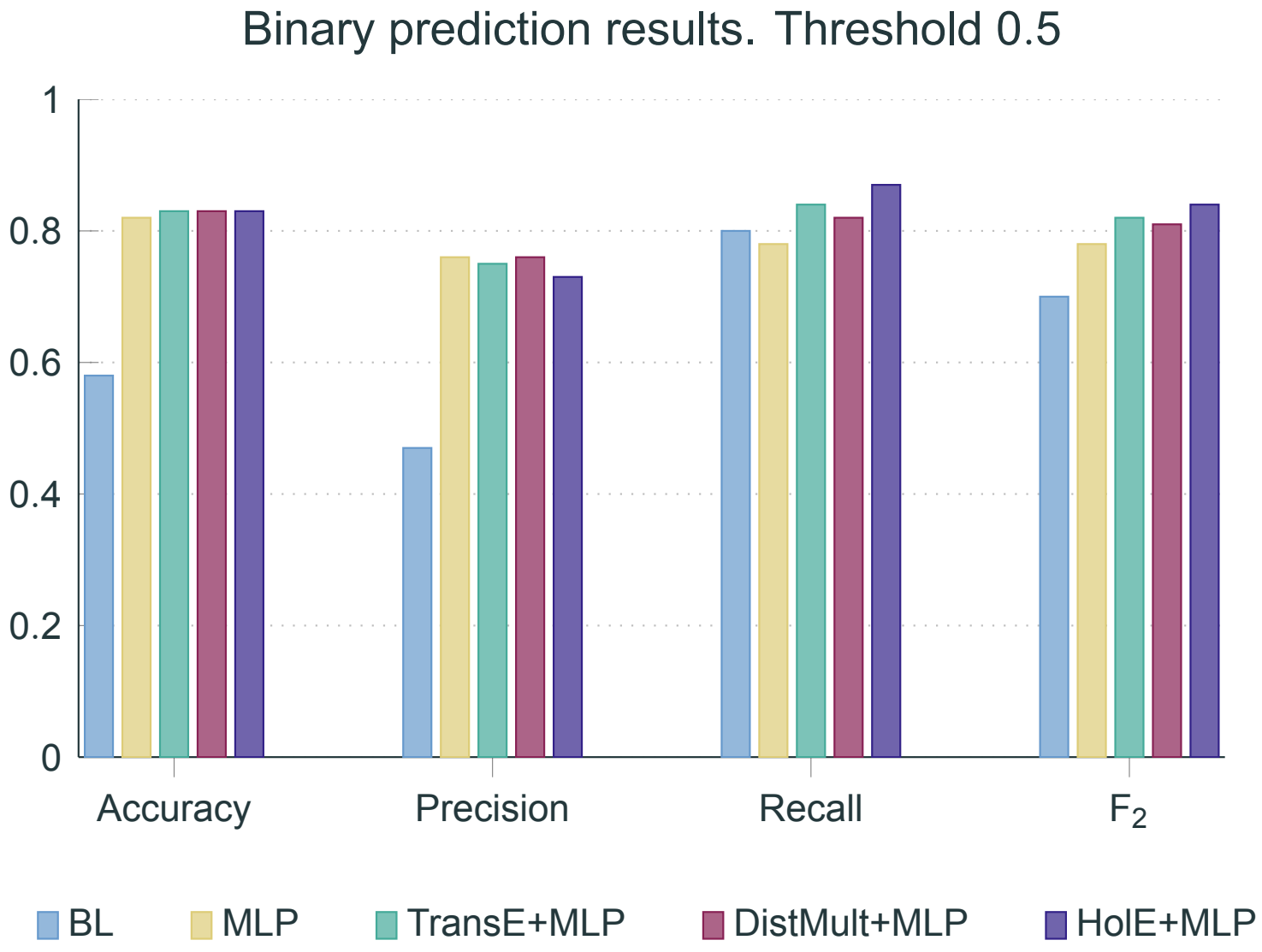
Preliminary Results



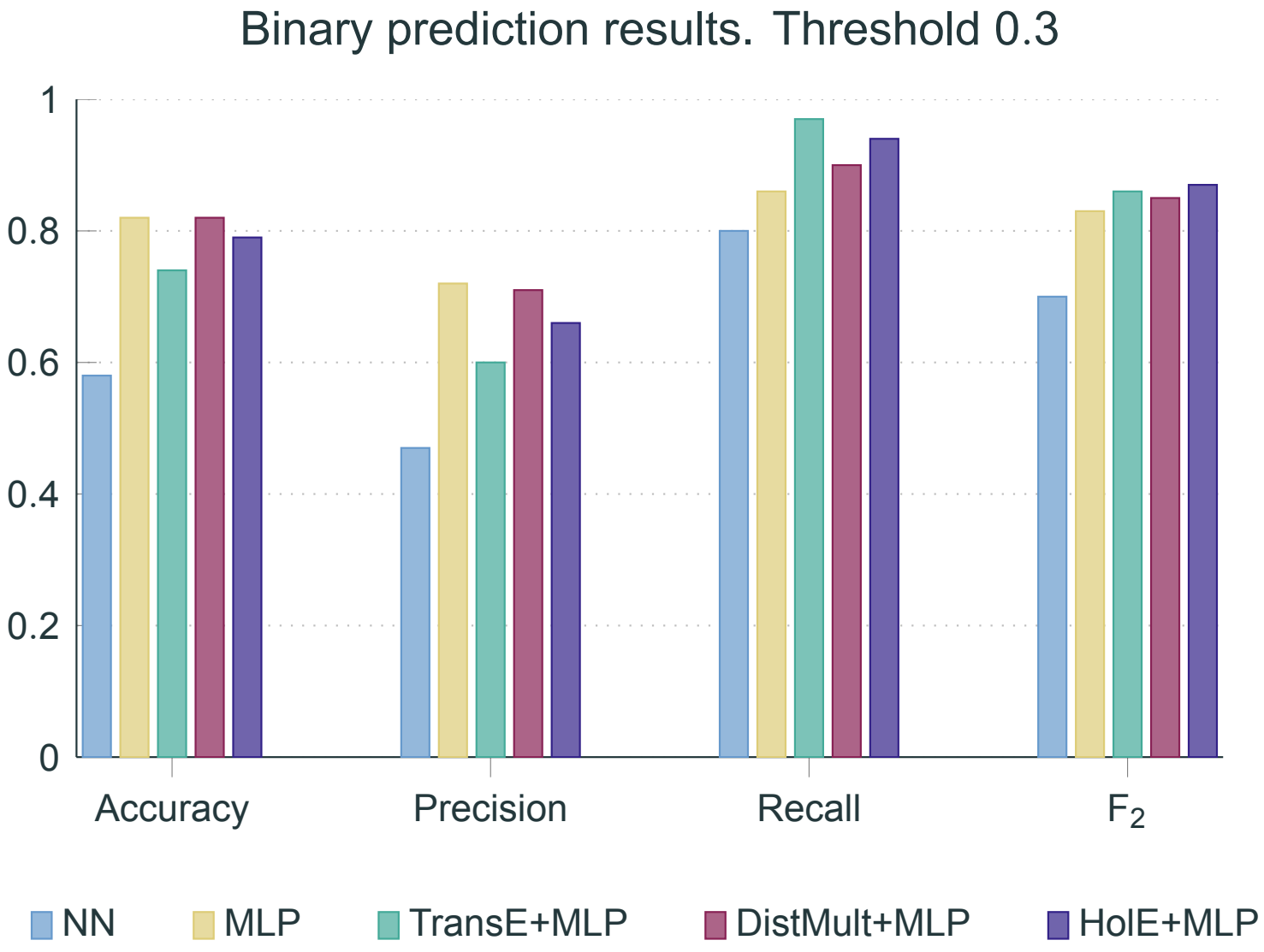
Preliminary Results



Preliminary Results



Preliminary Results



Completed and Future Work

Completed and Future Work

RQ1 ☑ Initial knowledge graph construction.

Completed and Future Work

RQ1 ☑ Initial knowledge graph construction.

RQ2 ☑ Shown that binary prediction of ecotoxicological effects can be improved with background knowledge.

Completed and Future Work

RQ1 ☑ Initial knowledge graph construction.

RQ2 ☑ Shown that binary prediction of ecotoxicological effects can be improved with background knowledge.

RQ1 ☐ Expand the knowledge graph with rules and axioms defined by domain researchers.

Completed and Future Work

RQ1 ☑ Initial knowledge graph construction.

RQ2 ☑ Shown that binary prediction of ecotoxicological effects can be improved with background knowledge.

RQ1 ☐ Expand the knowledge graph with rules and axioms defined by domain researchers.

RQ2 ☐ Move from binary labels to effect concentrations.

Completed and Future Work

RQ1 ☑ Initial knowledge graph construction.

RQ2 ☑ Shown that binary prediction of ecotoxicological effects can be improved with background knowledge.

RQ1 ☐ Expand the knowledge graph with rules and axioms defined by domain researchers.

RQ2 ☐ Move from binary labels to effect concentrations.

RQ2 ☐ Explore the use of more sophisticated models for embedding, *e.g.*, Graph Convolutional Networks.

Completed and Future Work

RQ1 ☑ Initial knowledge graph construction.

RQ2 ☑ Shown that binary prediction of ecotoxicological effects can be improved with background knowledge.

RQ1 ☐ Expand the knowledge graph with rules and axioms defined by domain researchers.

RQ2 ☐ Move from binary labels to effect concentrations.

RQ2 ☐ Explore the use of more sophisticated models for embedding, *e.g.*, Graph Convolutional Networks.

RQ1 ☐ Integration of work into ecological risk assessment pipeline.

Interested in more details?

Come to the presentation tomorrow at 12.20 in Lecture Theater 098.

ErikBMyklebust
ebm@niva.no