Swiss Institute of
Bioinformatics

ReconX Knowledge Graph in progress

Marco Pagni *et al.*

20 March 2024

Online

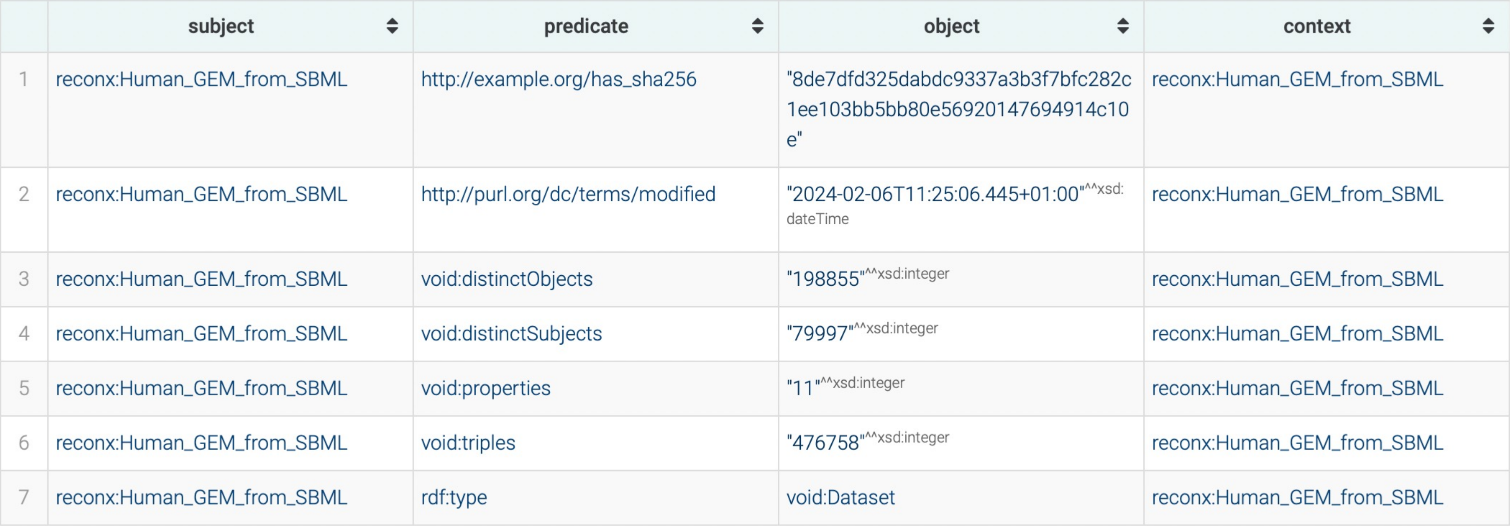
SWITZERLAND

https://github.com/sib-swiss/kgsteward

- **kgsteward** is a command-line tool to manage the content of a triplestore from a YAML configuration file
- written in Python… not to afraid the young generation ;-)
- it can upload RDF data from any URL or local files
- it can execute SPARQL updates
- it was developed using GraphDB as a triplestore. It has been demonstrated to work with RDF4J. Drivers for Virtuoso (maybe), qendpoint and Jena Fuseki could/should be written. Any help would be greatly appreciated. There is a need for a more muscular server than GraphDB Free which is limited one CPU.

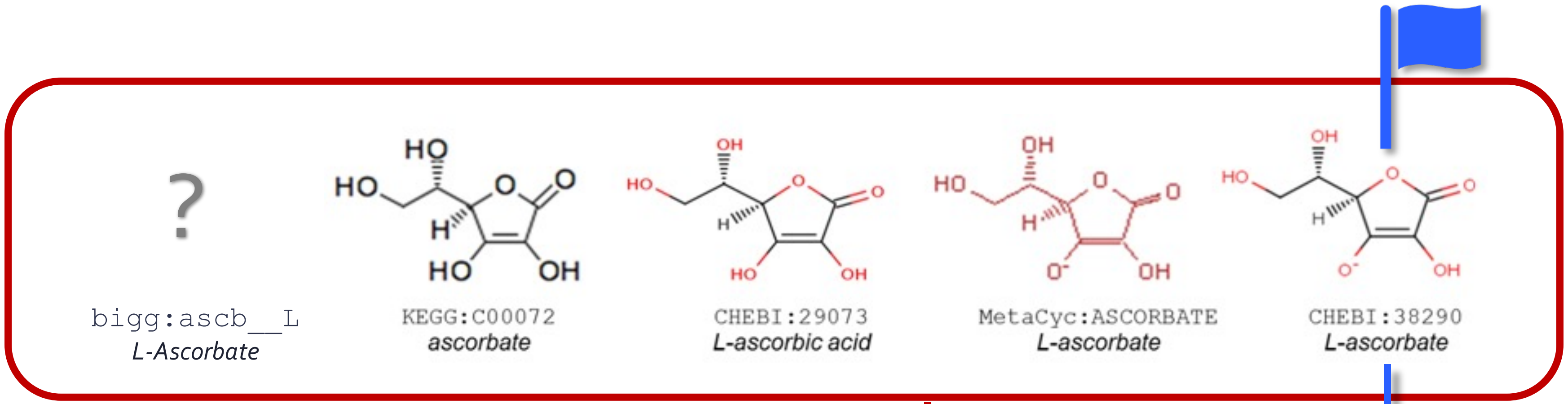
https://github.com/sib-swiss/kgsteward

- RDF data are organised in "dataset" , that are essentially the same as GraphDB/RDF4J *context* or *named graph* in SPARQL jargon.
- kgsteward inserts in the triplestore some minimal information about the sources of the dataset:
  - md5 checksum of local RDF files and local SPARQL files is the most important and permit to detect changes in the external data sources



	subject	predicate	object	context
1	reconx:Human_GEM_from_SBML	http://example.org/has_sha256	"8de7dfd325dabdc9337a3b3f7bfc282c1ee103bb5bb80e56920147694914c10e"	reconx:Human_GEM_from_SBML
2	reconx:Human_GEM_from_SBML	http://purl.org/dc/terms/modified	"2024-02-06T11:25:06.445+01:00"^^xsd:dateTime	reconx:Human_GEM_from_SBML
3	reconx:Human_GEM_from_SBML	void:distinctObjects	"198855"^^xsd:integer	reconx:Human_GEM_from_SBML
4	reconx:Human_GEM_from_SBML	void:distinctSubjects	"79997"^^xsd:integer	reconx:Human_GEM_from_SBML
5	reconx:Human_GEM_from_SBML	void:properties	"11"^^xsd:integer	reconx:Human_GEM_from_SBML
6	reconx:Human_GEM_from_SBML	void:triples	"476758"^^xsd:integer	reconx:Human_GEM_from_SBML
7	reconx:Human_GEM_from_SBML	rdf:type	void:Dataset	reconx:Human_GEM_from_SBML

Merging chemicals / selecting representative



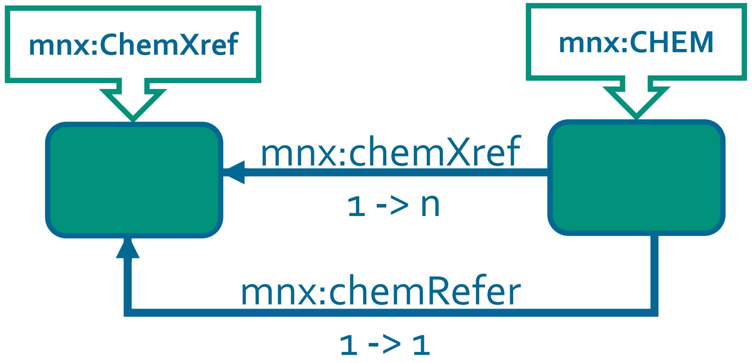
`bigg:ascb__L`
L-Ascorbate

`KEGG:C00072`
ascorbate

`CHEBI:29073`
L-ascorbic acid

`MetaCyc:ASCORBATE`
L-ascorbate

`CHEBI:38290`
L-ascorbate



`MNXM727871`

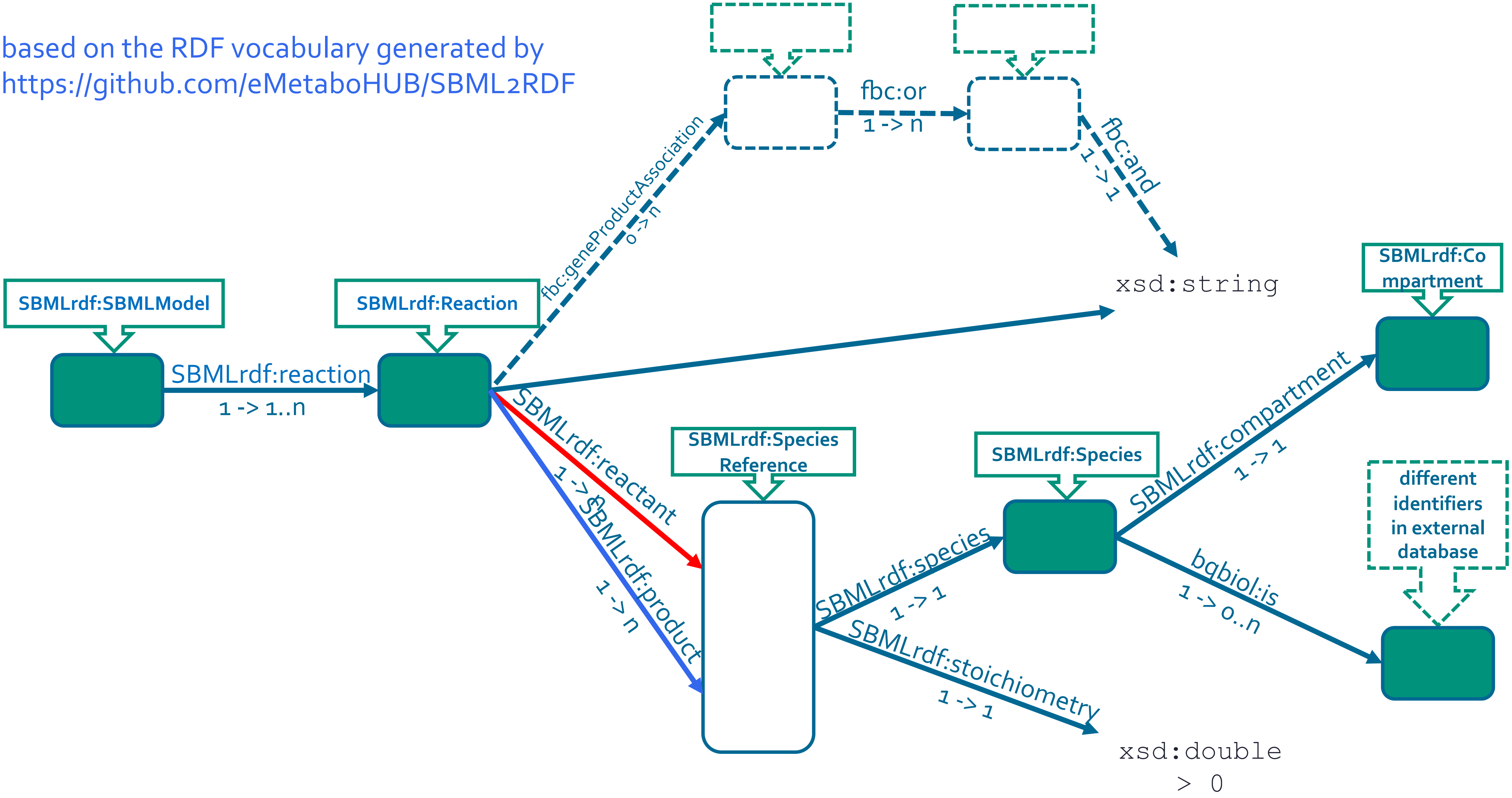
The MNXref identifier for this metabolite, *i.e.* an identifier for the set of molecules that are grouped together

`chebi:38290`

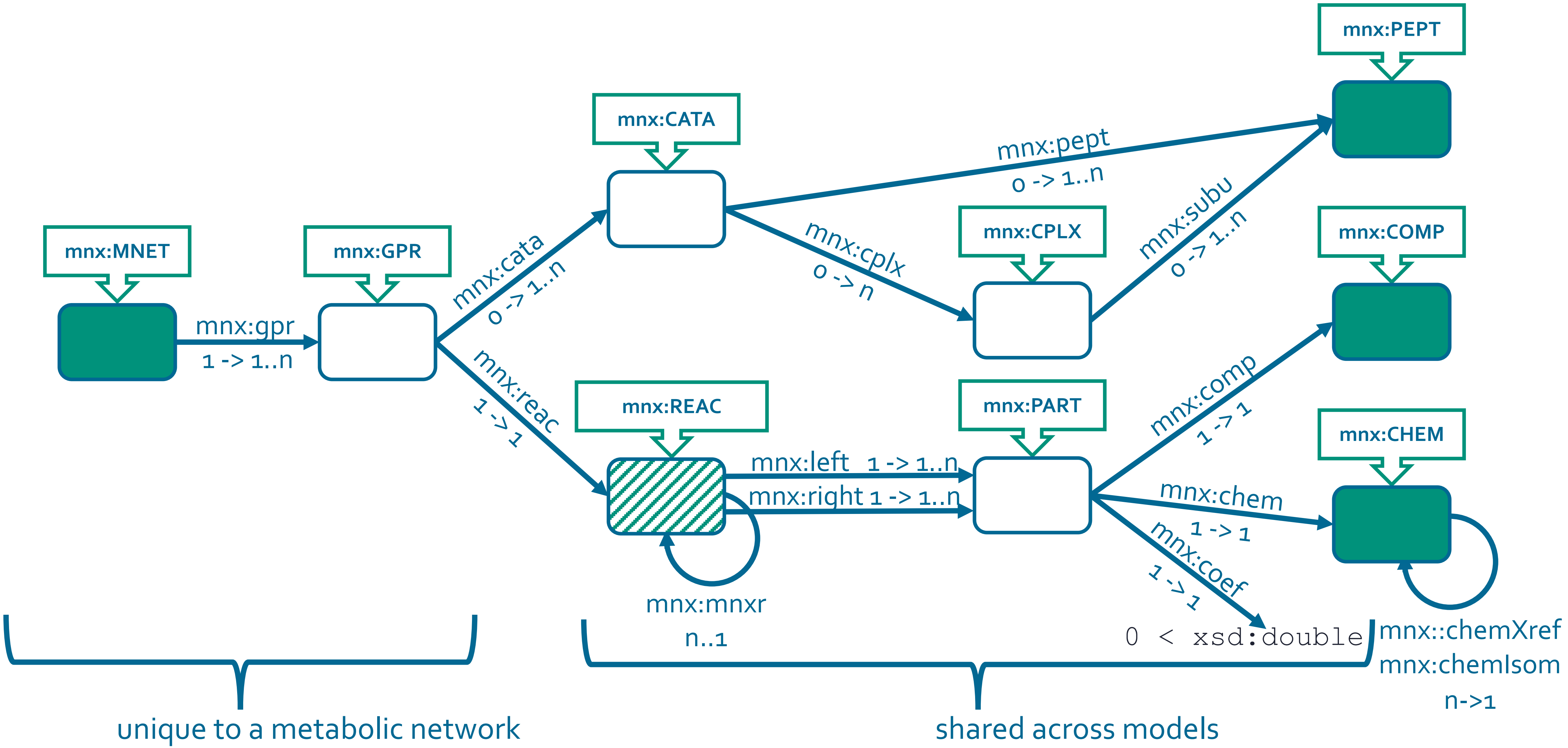
The **reference** (external) identifier that "best" represents this metabolite

Representation of reactions: **SBML** (simplified)

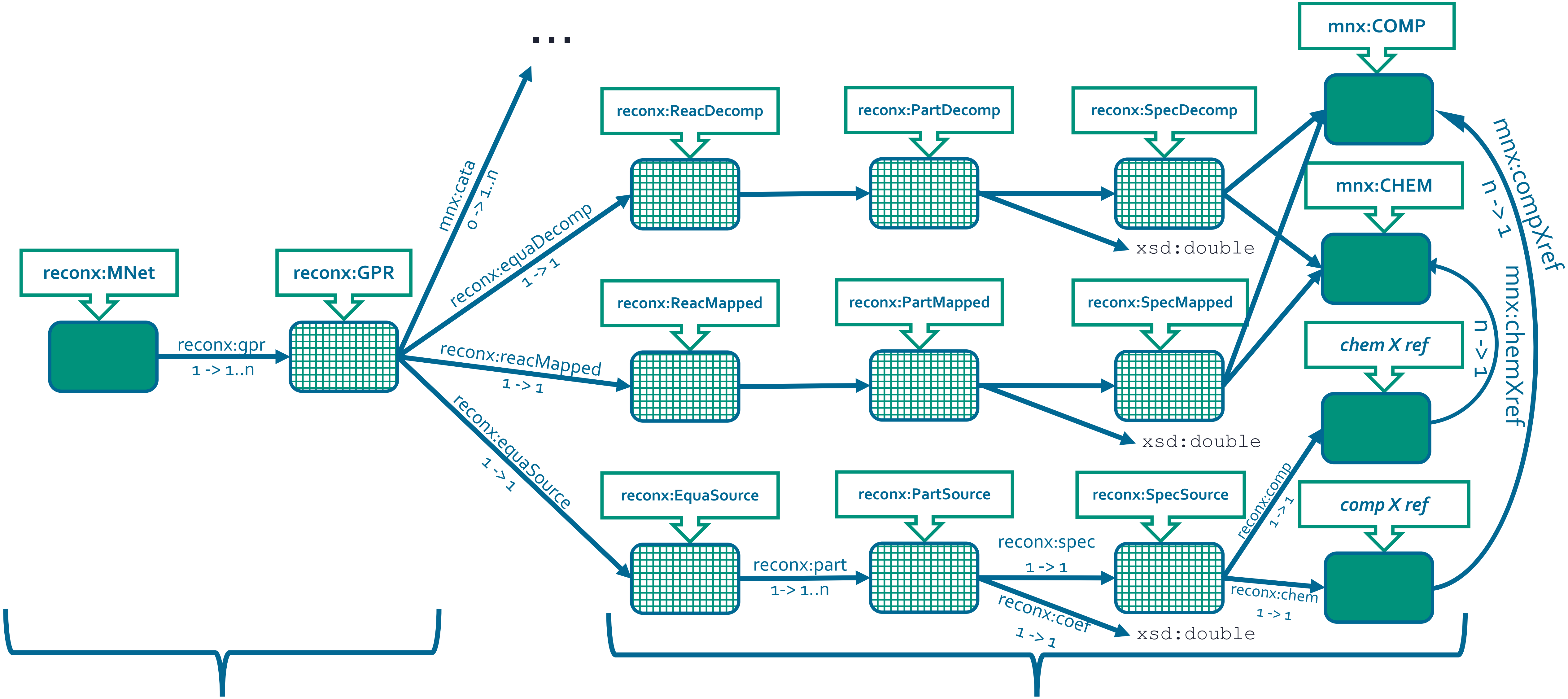
based on the RDF vocabulary generated by
https://github.com/eMetaboHUB/SBML2RDF



Representation of reactions: **MetaNetX**



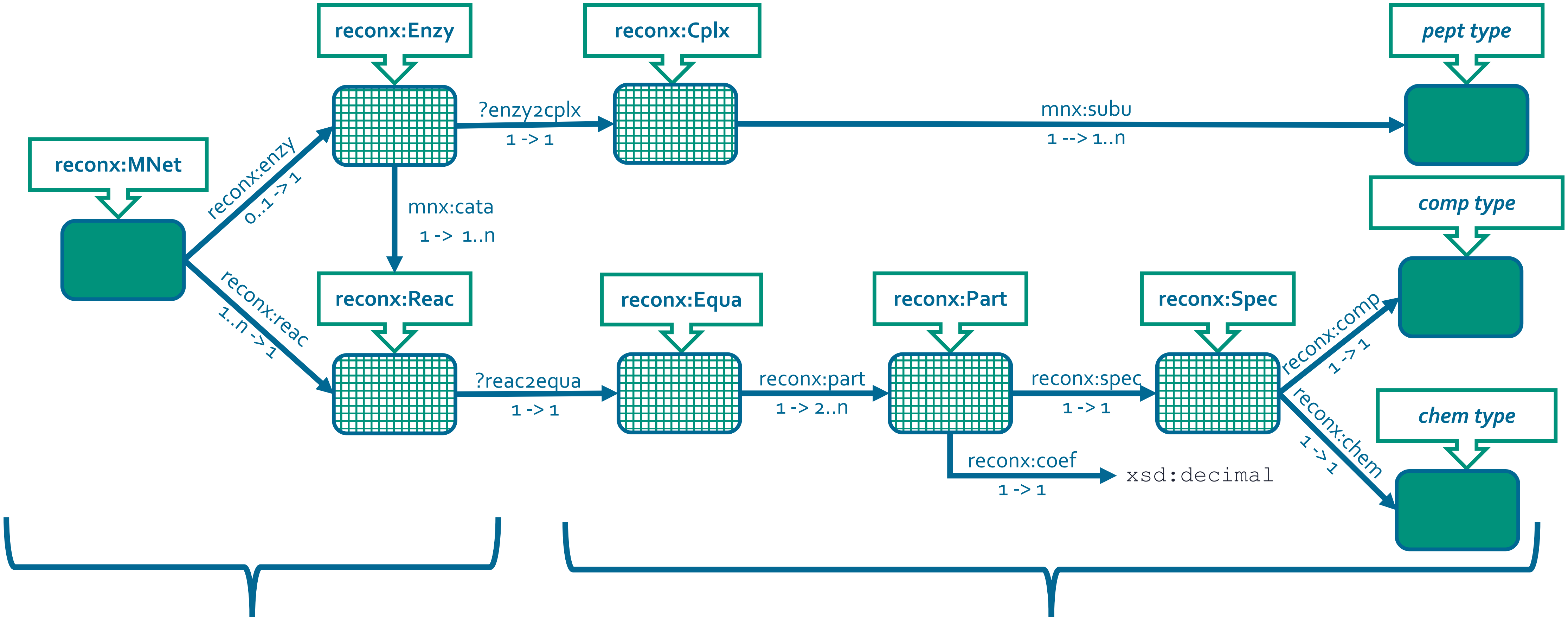
Integration of **ReconXKG** and **MetaNetX**



unique to a metabolic network

shared across metabolic networks

The core of **ReconXKG** (latest version)



unique to a metabolic network

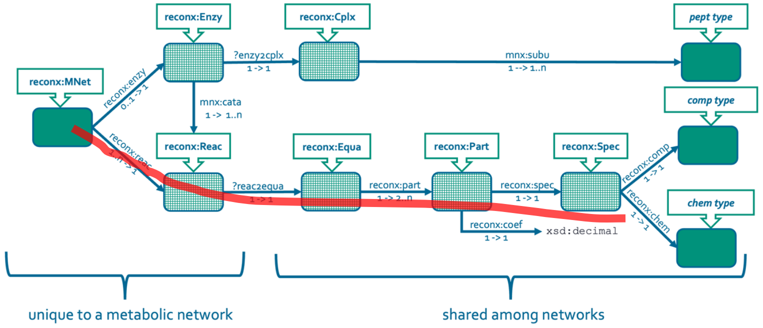
shared among networks

?enzy2cplx = reconx:cplxSource | reconx:cplxUP | reconx:cplxVMH

?reac2equa = reconx:equaSource | reconx:equaMNX | reconx:equaVMH | reconx:equaMNX_dc | reconx:equaVMH_dc

```
# The stoichiometric matrix of the E. coli core model.
# This is a sparse-matrix representation in which
# zero values are not shown. Rows and columns are
# identified by reac and spec IRIs, respectively
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX reconx: <https://reconx.vital-it.ch/kg/>

SELECT
        ( ?reac AS ?row )
        ( ?spec AS ?col )
        ( IF( ?flip, - ?coef, ?coef ) AS ?value )
WHERE{
    BIND( reconx:bigg_e_coli_core AS ?mnet ) # INPUT: any valid reconx mnet IRI
    BIND( reconx:equaSource AS ?reac2equa )  # INPUT: any valid equation "type"
    ?mnet reconx:reac ?reac .
    ?reac ?reac2equa        ?equa ;
          reconx:flipSource ?flip .
    ?equa reconx:part ?part .
    ?part reconx:spec ?spec ;
          reconx:coef ?coef .
}
ORDER BY ?reac ?spec
```



The five versions of biochemical equations

?reac2equa	Description
reconx:equaSource	• The original equation with no changes • Exchange reactions are lost in VMH recon
reconx:equaMNX	• Every metabolites is mapped to the MetaNetX best reference, which is most often a metabolite used by RHEA • Transported protons are converted to mnx:PMF • Balance protons are lost
reconx:equaVMH	• As reconx:equaMNX, but metabolites are mapped to a unique VMH identifier using MetaNetX, if feasible
reconx:equaMNX_dc	• De-compartimentalized version of reconx:equaMNX • Transport reactions are lost
reconx:equaVMH_dc	• De-compartimentalized version of reconx:equaVMH • Transport reactions are lost

Nota Bene: The comments in red will be addressed in future versions

The three versions of enzymatic complexes

?enzy2cplx	Description
reconx:cplxSource	• The original complex with no changes
reconx:cplxUP	• Genes/proteins are mapped to UniProt entries, possibly through NCBI geneid
reconx:cplxVMH	• Genes/proteins are mapped to VMH entries, possibly through NCBI geneid

Mapping between VMH and UniProt are not always one-to-one. Hence, the different version of enzymatic complexes are not necessarily strictly equivalent between resources.

VMH recon *versus RHEA*

what	in RHEA	in VMH	common	Description
reconx:equaMNX_dc	14799	5091	557	• De-compartimentalised chemical equations, reconciled through MetaNetX
chem	13156	4080	655	• Metabolites
pept	3483	3273	2281	• UniProt human proteins

Nota Bene:

- RHEA reactions are not limited to human, but those involving polymers are ignored.
- De-compartimentalisation is not (yet) performed ton transport reactions.
- Numbers for VMH genes are comparable and close to those for UniProt proteins.

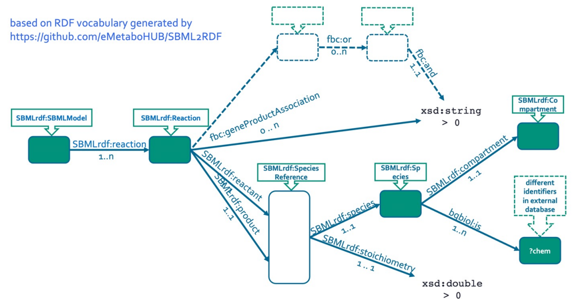
Challenges

Far too many of everything: labels, InChIs, names, cross references.

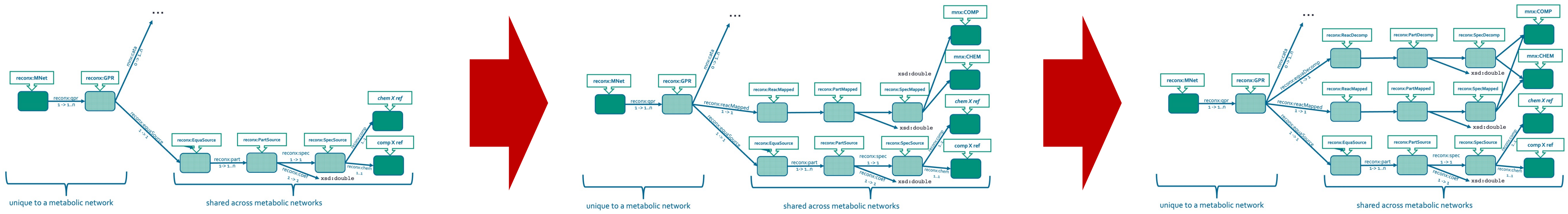
PoP implementation

Human-GEM.sbml

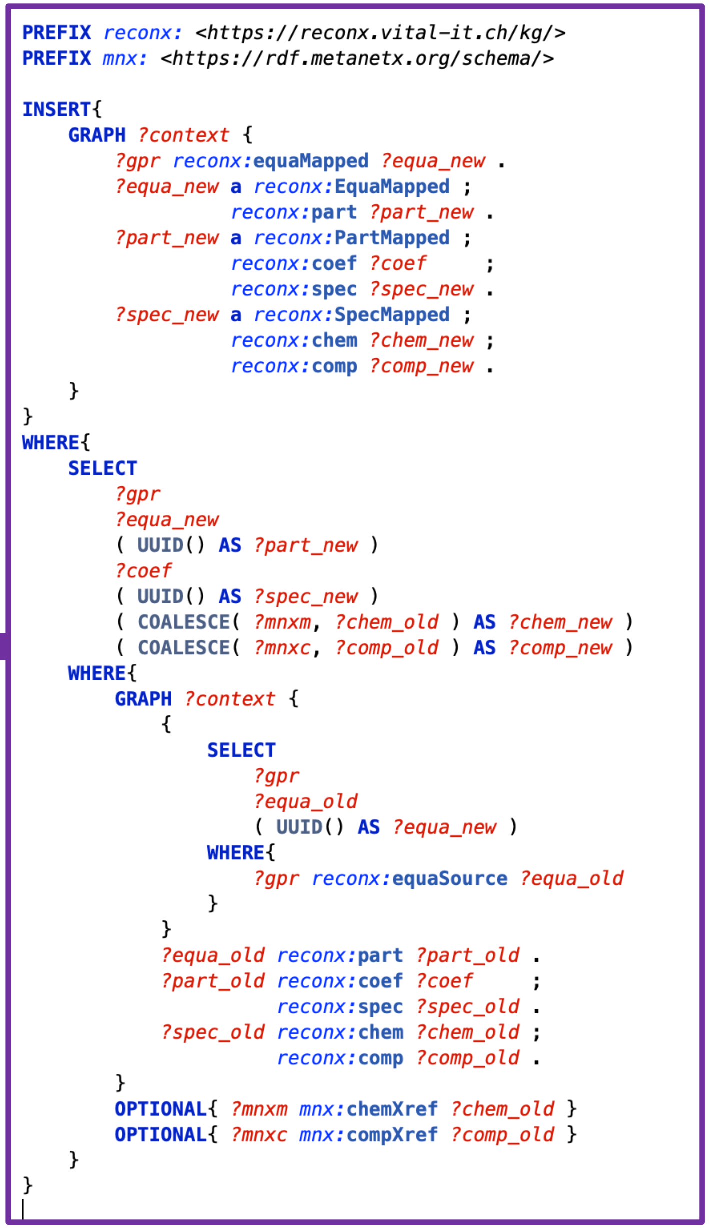
Human-GEM.ttl



```
server_url       : ${RECONXKG_URL}
username         : ${RECONXKG_USERNAME}
password         : ${RECONXKG_PASSWORD}
repository_id    : ReconXKG
server_config    : ${RECONXKG_DIR}/config/ReconXKG-config.ttl
dataset_base_IRI : http://example.org/context/
use_file_server  : true
graphs:
  - dataset: MNXref
    file:
      # - https://www.metanetx.org/ftp/beta/MNXref.ttl.gz
      - /Users/mpagni/gitlab.sib.swiss/MetaNetX/cooked/chemistry/ns/MNXref.ttl.gz
  - dataset: MNXref_schema
    url:
      - https://www.metanetx.org/ftp/beta/MNXref_schema.ttl
  - dataset: MNet_from_SBML
    # see recipe in data/model/SBML/README.md
    parent: MNXref
    file:
      - ${RECONXKG_DIR}/data/model/SBML/Human-GEM_fixed.ttl
    update:
      - sparql_update_file:
          - ${RECONXKG_DIR}/sparql/update/convert_SBML_to_MNet_raw.sparql
          - ${RECONXKG_DIR}/sparql/update/fix_MNet_external_reaction.sparql
#          - ${RECONXKG_DIR}/sparql/update/cleanup_SBML_to_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_source_to_mapped.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_mapped_to_decomp.sparql
#         - ${RECONXKG_DIR}/sparql/update/introduce_PMF_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_coef_sign_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_internal_IRI_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/complete_type_and_label_in_MNet.sparql
        replace:
          - "?context" : ${TARGET_GRAPH_CONTEXT}
```

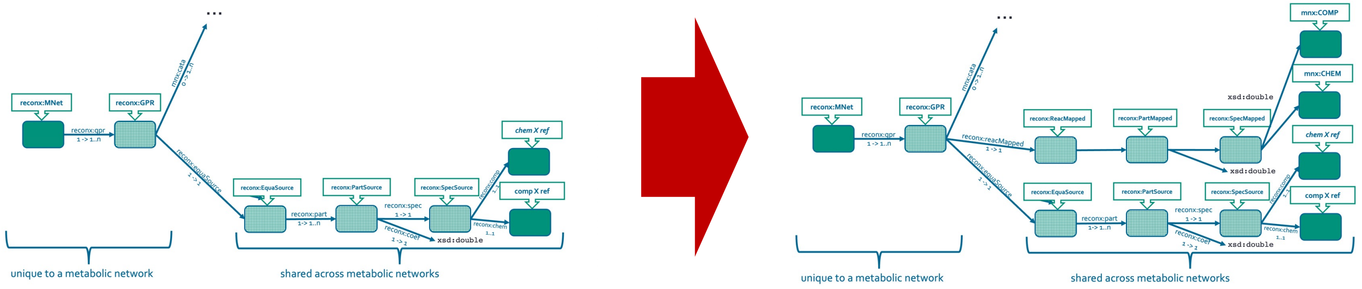


```
server_url        : ${RECONXKG_URL}
username          : ${RECONXKG_USERNAME}
password          : ${RECONXKG_PASSWORD}
repository_id     : ReconXKG
server_config     : ${RECONXKG_DIR}/config/ReconXKG-config.ttl
dataset_base_IRI  : http://example.org/context/
use_file_server   : true
graphs:
  - dataset: MNXref
    file:
      # - https://www.metanetx.org/ftp/beta/MNXref.ttl.gz
      - /Users/mpagni/gitlab.sib.swiss/MetaNetX/cooked/chemistry/ns/MNXref.ttl.gz
  - dataset: MNXref_schema
    url:
      - https://www.metanetx.org/ftp/beta/MNXref_schema.ttl
  - dataset: MNet_from_SBML
    # see recipe in data/model/SBML/README.md
    parent: MNXref
    file:
      - ${RECONXKG_DIR}/data/model/SBML/Human-GEM_fixed.ttl
    update:
      - sparql_update_file:
          - ${RECONXKG_DIR}/sparql/update/convert_SBML_to_MNet_raw.sparql
          - ${RECONXKG_DIR}/sparql/update/fix_MNet_external_reaction.sparql
#          - ${RECONXKG_DIR}/sparql/update/cleanup_SBML_to_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_source_to_mapped.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_mapped_to_decomp.sparql
#         - ${RECONXKG_DIR}/sparql/update/introduce_PMF_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_coef_sign_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_internal_IRI_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/complete_type_and_label_in_MNet.sparql
        replace:
          - "?context" : ${TARGET_GRAPH_CONTEXT}
```



```
PREFIX reconx: <https://reconx.vital-it.ch/kg/>
PREFIX mnx: <https://rdf.metanetx.org/schema/>

INSERT{
    GRAPH ?context {
        ?gpr reconx:equaMapped ?equa_new .
        ?equa_new a reconx:EquaMapped ;
                  reconx:part ?part_new .
        ?part_new a reconx:PartMapped ;
                  reconx:coef ?coef     ;
                  reconx:spec ?spec_new .
        ?spec_new a reconx:SpecMapped ;
                  reconx:chem ?chem_new ;
                  reconx:comp ?comp_new .
    }
}
WHERE{
    SELECT
        ?gpr
        ?equa_new
        ( UUID() AS ?part_new )
        ?coef
        ( UUID() AS ?spec_new )
        ( COALESCE( ?mnxm, ?chem_old ) AS ?chem_new )
        ( COALESCE( ?mnxc, ?comp_old ) AS ?comp_new )
    WHERE{
        GRAPH ?context {
            {
                SELECT
                    ?gpr
                    ?equa_old
                    ( UUID() AS ?equa_new )
                WHERE{
                    ?gpr reconx:equaSource ?equa_old
                }
            }
            ?equa_old reconx:part ?part_old .
            ?part_old reconx:coef ?coef      ;
                      reconx:spec ?spec_old .
            ?spec_old reconx:chem ?chem_old ;
                      reconx:comp ?comp_old .
        }
        OPTIONAL{ ?mnxm mnx:chemXref ?chem_old }
        OPTIONAL{ ?mnxc mnx:compXref ?comp_old }
    }
}
```



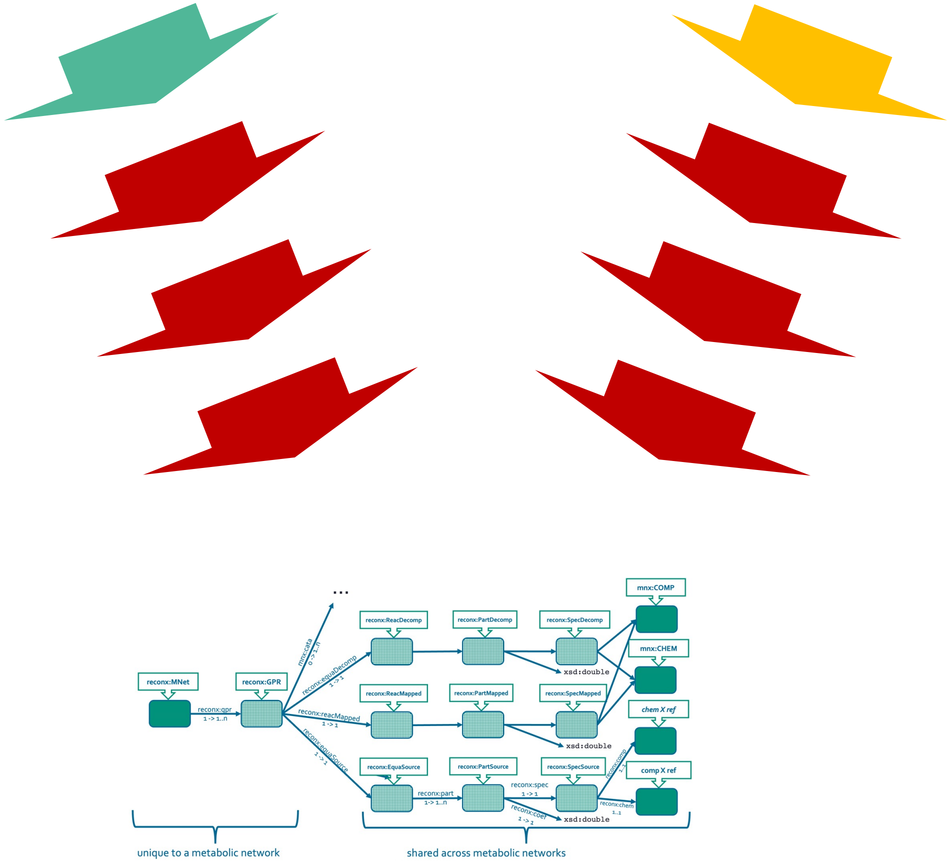
More PoP implementation RHEA + Recon3D

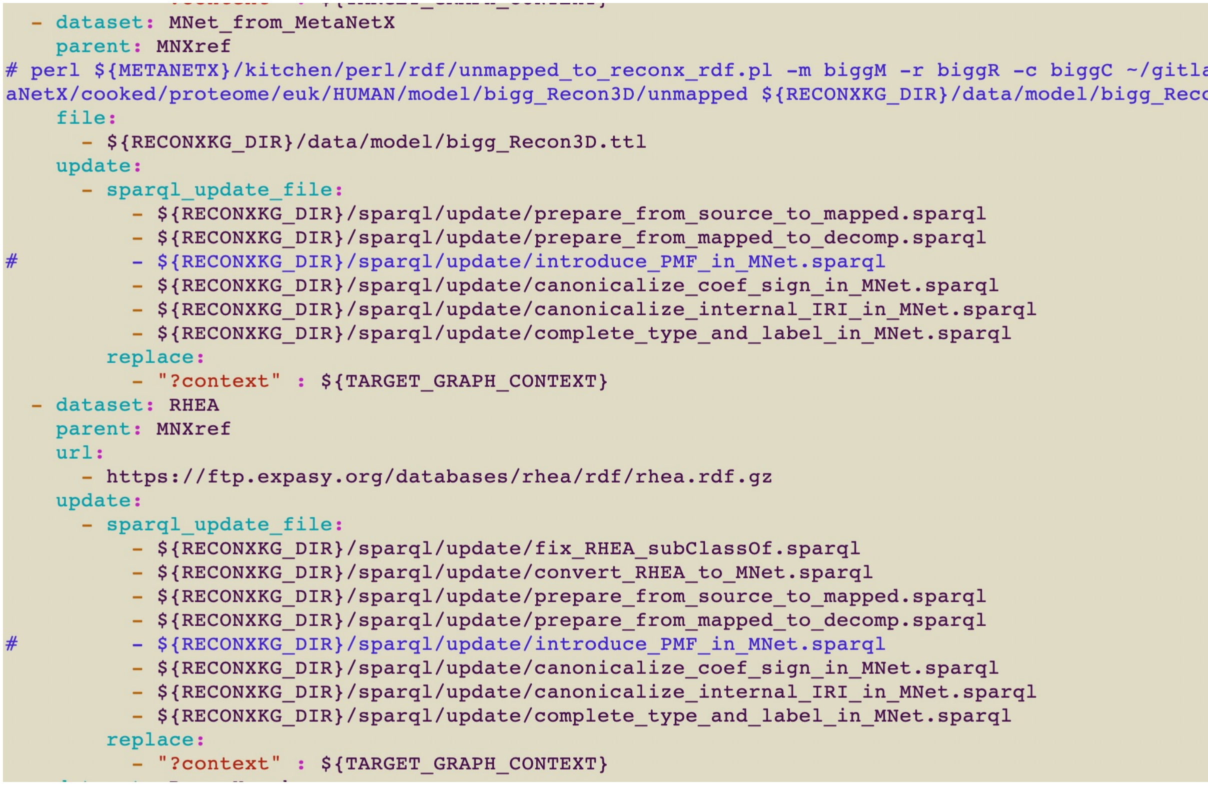
Recon3D

(MetaNetX file format)

RHEA

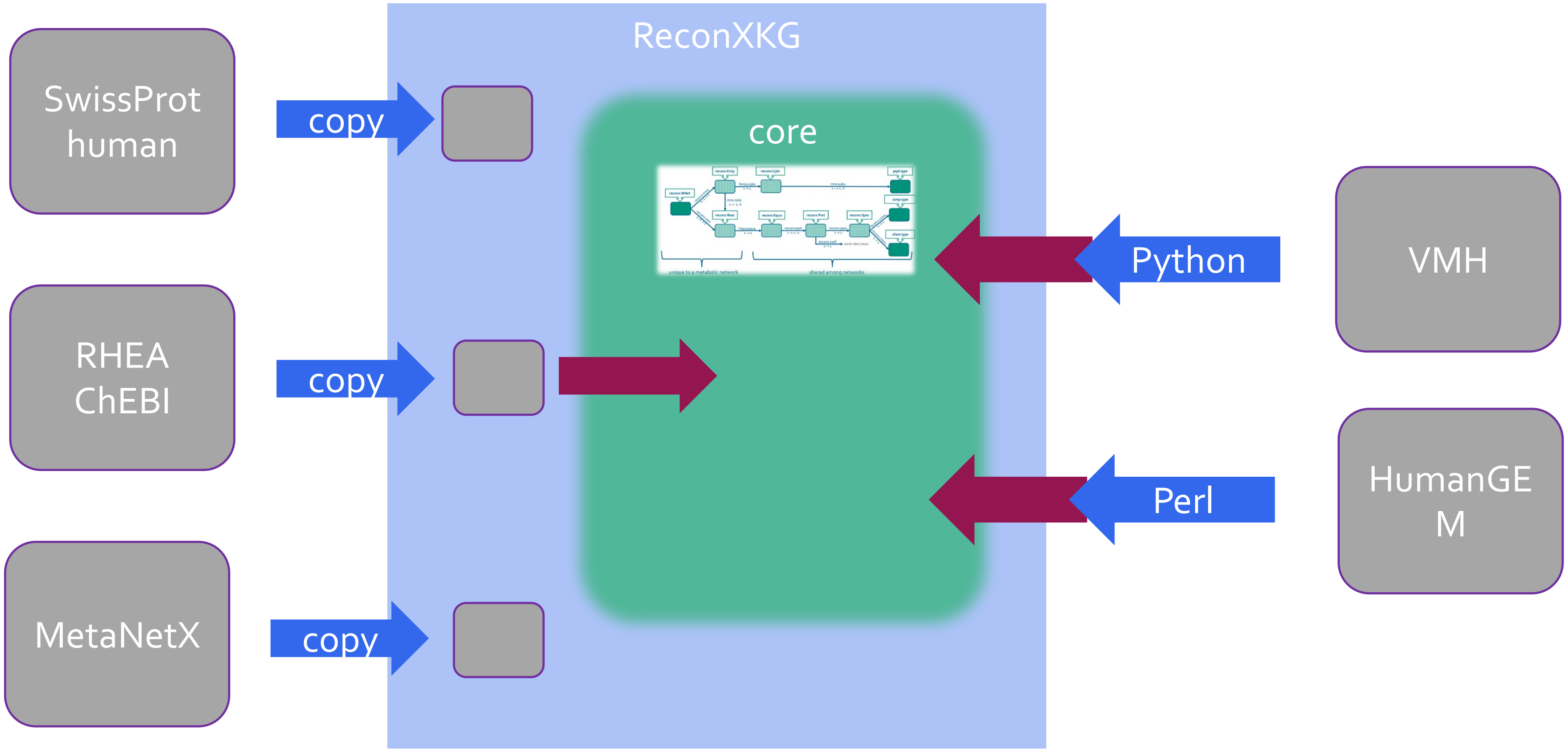
(RDF distributoin)





```
  - dataset: MNet_from_MetaNetX
    parent: MNXref
# perl ${METANETX}/kitchen/perl/rdf/unmapped_to_reconx_rdf.pl -m biggM -r biggR -c biggC ~/gitla
aNetX/cooked/proteome/euk/HUMAN/model/bigg_Recon3D/unmapped ${RECONXKG_DIR}/data/model/bigg_Reco
    file:
      - ${RECONXKG_DIR}/data/model/bigg_Recon3D.ttl
    update:
      - sparql_update_file:
          - ${RECONXKG_DIR}/sparql/update/prepare_from_source_to_mapped.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_mapped_to_decomp.sparql
#         - ${RECONXKG_DIR}/sparql/update/introduce_PMF_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_coef_sign_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_internal_IRI_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/complete_type_and_label_in_MNet.sparql
        replace:
          - "?context" : ${TARGET_GRAPH_CONTEXT}
  - dataset: RHEA
    parent: MNXref
    url:
      - https://ftp.expasy.org/databases/rhea/rdf/rhea.rdf.gz
    update:
      - sparql_update_file:
          - ${RECONXKG_DIR}/sparql/update/fix_RHEA_subClassOf.sparql
          - ${RECONXKG_DIR}/sparql/update/convert_RHEA_to_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_source_to_mapped.sparql
          - ${RECONXKG_DIR}/sparql/update/prepare_from_mapped_to_decomp.sparql
#         - ${RECONXKG_DIR}/sparql/update/introduce_PMF_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_coef_sign_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/canonicalize_internal_IRI_in_MNet.sparql
          - ${RECONXKG_DIR}/sparql/update/complete_type_and_label_in_MNet.sparql
        replace:
          - "?context" : ${TARGET_GRAPH_CONTEXT}
```

Data flows in ReconXKG



SwissProt human

copy

RHEA ChEBI

copy

MetaNetX

copy

ReconXKG

core

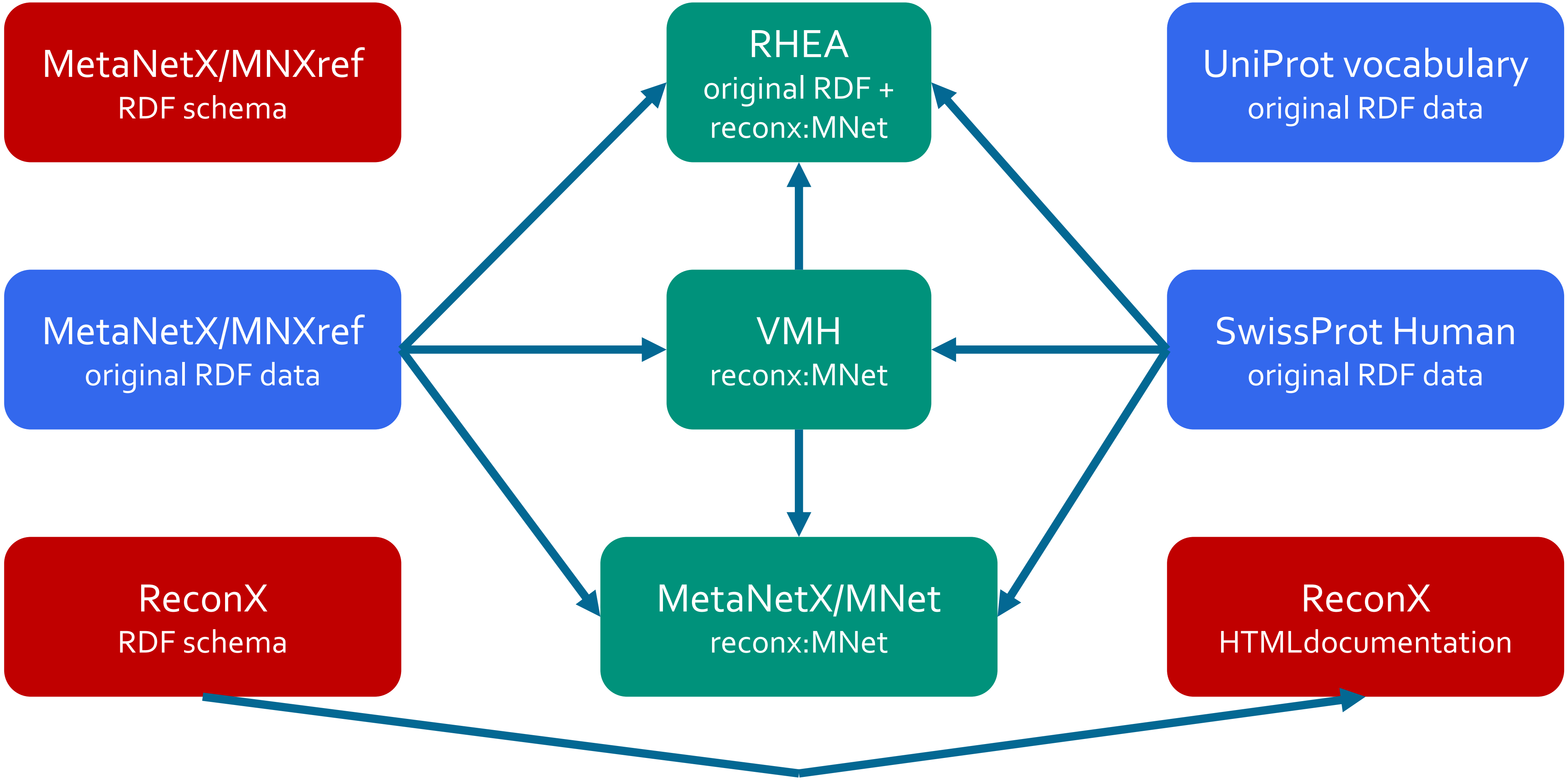
Python

VMH

Perl

HumanGEM

ReconX Knowledge Graph



MetaNetX/MNXref
RDF schema

RHEA
original RDF +
reconx:MNet

UniProt vocabulary
original RDF data

MetaNetX/MNXref
original RDF data

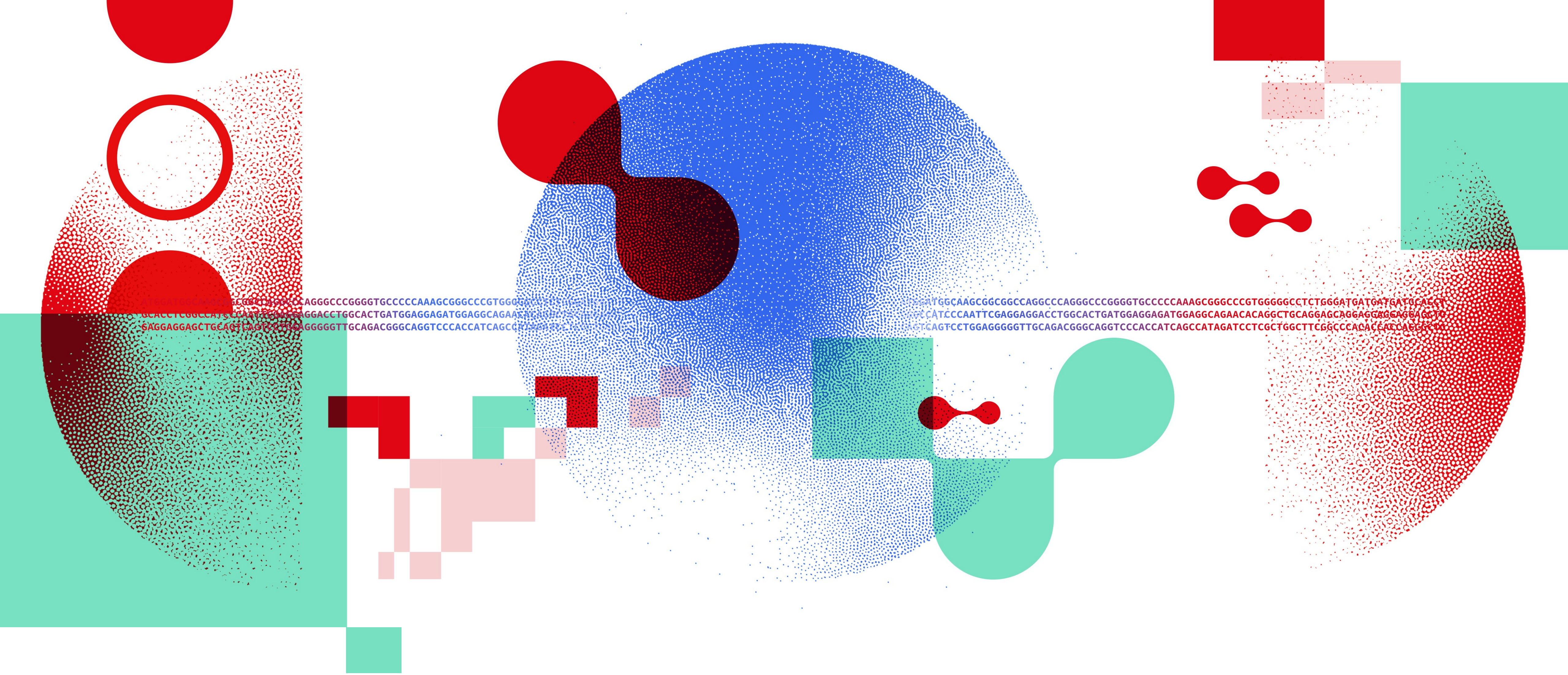
VMH
reconx:MNet

SwissProt Human
original RDF data

ReconX
RDF schema

MetaNetX/MNet
reconx:MNet

ReconX
HTMLdocumentation



elixir SWITZERLAND

SIB