

WikiPathways Summit 2018

The background of the slide features a dark blue field with various abstract, colorful lines and shapes. These include solid and dashed lines in orange, cyan, and magenta, some with arrowheads. There are also geometric shapes like a hexagon, circles, and rounded rectangles in the same color palette. A large, light gray arrow points downwards from the top title box towards the central text box.

Using WikiPathways with Its Resource Description Framework Format

Egon Willighagen
Maastricht University

Resource Description Framework

- Triples
 - *subject - predicate - object*
 - :someGene :hasIdentifier “ENSG12345678”
 - Resources: Internationalized Resource Identifiers
 - <http://identifiers.org/ensembl/ENSG12345678>
- Ontologies
 - Standardizing predicates, classes, etc
 - Common language → interoperability



WikiPathways RDF

RESEARCH ARTICLE

Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources

Andra Waagmeester¹, Martina Kutmon^{1,2}, Anders Riutta³, Ryan Miller¹, Egon L. Willighagen¹, Chris T. Evelo^{1,2*}, Alexander R. Pico³

1 Department of Bioinformatics - BiGCaT, Maastricht University, Maastricht, The Netherlands, **2** Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands, **3** Gladstone Institutes, San Francisco, California, United States of America

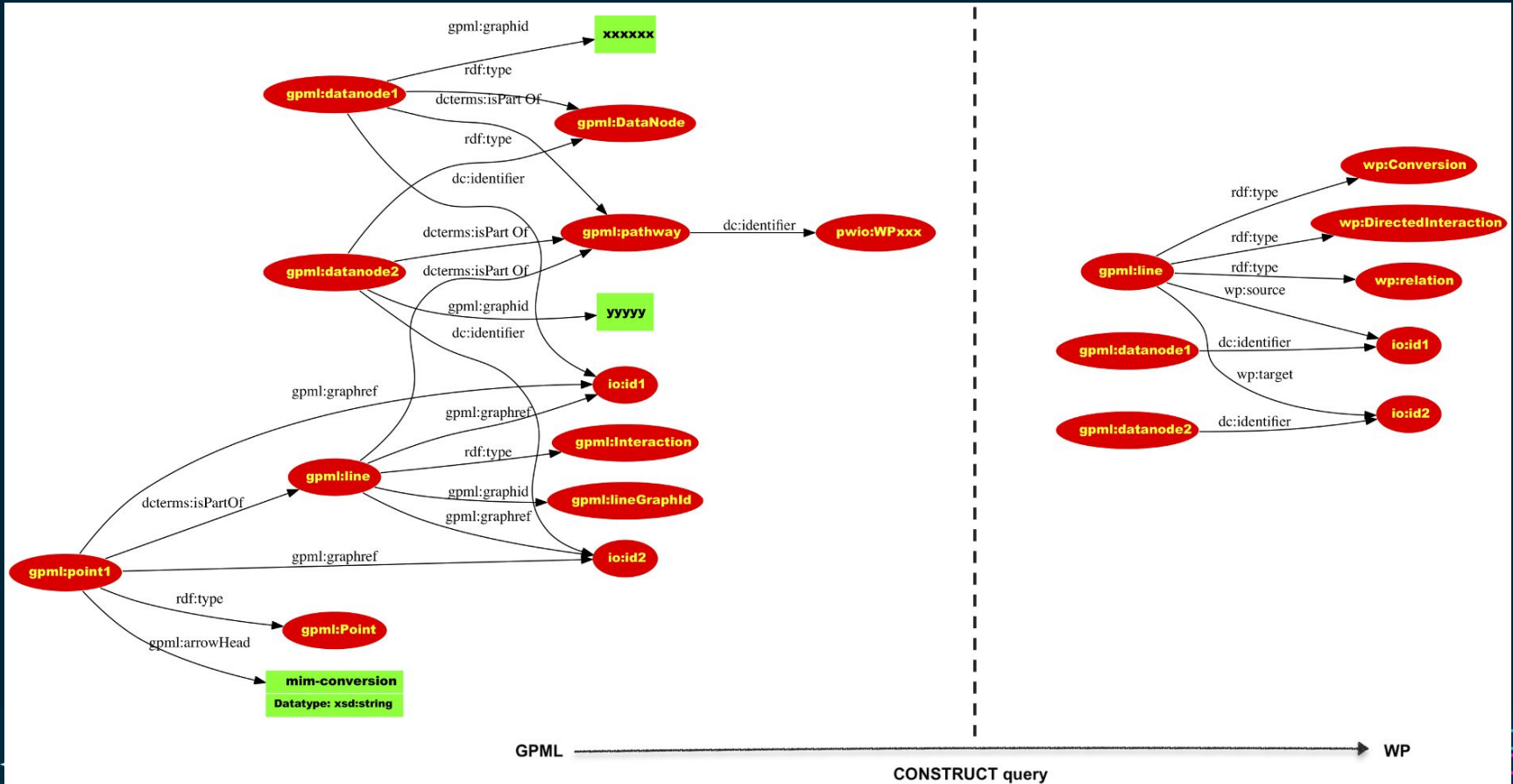
* chris.evelo@maastrichtuniversity.nl



Abstract

The diversity of online resources storing biological data in different formats provides a chal-

GPMLRDF → WPRDF



rdf.wikipathways.org

The WikiPathways Semantic Web Portal

[edit](#)

Welcome to the WikiPathways Semantic Web Portal!

This portal describes the Semantic Web features of the WikiPathways databases, such as the [Resource Description Framework](#) (RDF) translation, the ontology, and the new nanopublications.

The WikiPathways RDF is provided as part of the monthly releases and contains the [Curated](#) and [Reactome](#) pathways. The RDF is split in two parts, the GPMLRDF part which contains a direct translation of the content in the GPML files, and a WPRDF part which contains the biology represented in the GPML



The WikiPathways Vocabularies

The [WikiPathways vocabularies](#) are for the semantic information about the pathway, data nodes, and interactions and the GPML vocabulary is for the graphical information about how the pathway diagram is laid out and represented.

How to cite

If you use the RDF, vocabularies, or nanopublication, please cite the following paper:

- Waagmeester, A., Kutmon, M., Riutta, A., Miller, R., Willighagen, E. L., Evelo, C. T., Pico, A. R., Jun. 2016. Using the semantic web for rapid integration of WikiPathways with other biological online data resources. PLoS Comput Biol 12 (6), e1004989+. doi:[10.1371/journal.pcbi.1004989](https://doi.org/10.1371/journal.pcbi.1004989)

For the pathway content, please follow these [How to cite WikiPathways](#) instructions.

News

[edit](#)

2018-01-01 - The RDF is being used to expose [facts from WikiPathways as nanopublications](#).

2017-05-04 - This portal was created

SPARQL

[edit](#)

We provide [an \(experimental\) SPARQL endpoint](#) where data queries can be done.

WikiPathways Example SPARQL Queries

Example RDF

```
<http://identifiers.org/ncbigene/8792>  
  a          wp:DataNode , wp:Protein ;  
  rdfs:label  "RANK"^^xsd:string ;  
  dc:identifier <http://identifiers.org/ncbigene/8792> ;  
  dc:source   "Entrez Gene"^^xsd:string ;  
  dcterms:identifier "8792"^^xsd:string ;  
  dcterms:isPartOf  <http://rdf.wikipathways.org/Pathway/WP12_r97823/WP/Interaction/id6251ac9> ,  
                    <http://rdf.wikipathways.org/Pathway/WP12_r97823/WP/Interaction/idffa65a8b> ,  
                    <http://identifiers.org/wikipathways/WP12_r97823> ,  
                    <http://rdf.wikipathways.org/Pathway/WP12_r97823/WP/Interaction/f49c2> ;  
  wp:bdbEnsembl <http://identifiers.org/ensembl/ENSG00000141655> ;  
  wp:bdbEntrezGene <http://identifiers.org/ncbigene/8792> ;  
  wp:bdbHgncSymbol <http://identifiers.org/hgnc.symbol/TNFRSF11A> ;  
  wp:bdbUniprot   <http://identifiers.org/uniprot/Q9Y6Q6> ,  
                  <http://identifiers.org/uniprot/A0A1W2PR41> ;  
  wp:isAbout      <http://rdf.wikipathways.org/Pathway/WP12_r97823/DataNode/f00> .
```



SPARQL

queries

```
PREFIX wp:    <http://vocabularies.wikipathways.org/wp#>  
PREFIX rdfs:  <http://www.w3.org/2000/01/rdf-schema#>  
PREFIX dcterms: <http://purl.org/dc/terms/>
```

```
SELECT DISTINCT ?pathway (str(?label) as ?geneProduct)  
WHERE {  
    ?geneProduct a wp:GeneProduct .  
    ?geneProduct rdfs:label ?label .  
    ?geneProduct dcterms:isPartOf ?pathway .  
    ?pathway a wp:Pathway .  
  
    FILTER regex(str(?label), "CYP").  
}
```



SPARQL queries

4 Example queries

4.1 Metadata queries

4.2 Pathway oriented queries

- 4.2.1 Get the species currently in WikiPathways with their respective URI's
- 4.2.2 List pathways and their species
- 4.2.3 List the species captured in WikiPathways and the number of pathways per species
- 4.2.4 List all pathways for species "Mus musculus"
- 4.2.5 Get all pathways with a particular gene
- 4.2.6 Get all groups and complexes containing a particular gene
- 4.2.7 Get all the genes on a particular pathway
- 4.2.8 Count the number of pathways per ontology term
- 4.2.9 Get all pathways with a particular ontology term
- 4.2.10 Get all ontology terms for a particular pathway
- 4.2.11 Get all Reactome pathways
- 4.2.12 Get all pathways with Pubmed references
- 4.2.13 Get all pathways with a particular Pubmed reference
- 4.2.14 Get all pathways and the number of references per pathway
- 4.2.15 Get a full dump of all pathways and their pathway ontological terms

4.3 Data statistics oriented queries

- 4.3.1 Count the number of metabolites per species

4.4 Interaction oriented queries

- 4.4.1 Get all interactions for a particular datanode
- 4.4.2 Find all datanodes (GeneProducts, Metabolites, Pathways) that are connected to a particular datanode via any type of interaction (wp:Interaction).
- 4.4.3 Get all interactions for a particular pathway.
- 4.4.4 Get all interactions for a particular pathway and their participants.
- 4.4.5 Get all Interactions.
- 4.4.6 Get all Interactions for a species (Homo sapiens).
- 4.4.7 Get downstream adjacent nodes from a source.
- 4.4.8 Get upstream adjacent nodes from a target.

4.5 Datasource oriented queries

- 4.5.1 Get all datasources currently captured in WikiPathways
- 4.5.2 Get the number of entries per datasource in WikiPathways
- 4.5.3 Count the identifiers per data source
- 4.5.4 Count the identifiers per data source and order them from high to low
- 4.5.5 Return all compounds annotated with the "ChEMBL compound" as data source and the pathways they are in

4.6 Curators oriented queries

- 4.6.1 Get the pathway with the erroneous data source "null"
- 4.6.2 Get all geneproducts that lack either a DataSource or an Identifier
- 4.6.3 Get entities with more than one Identifier
- 4.6.4 PubChem-compound 1004
- 4.6.5 Outdated HMDB identifiers
- 4.6.6 Metabolites not classified as such

SPARQL queries

4.1 Metadata queries

4.2 Pathway oriented queries

4.2.1 Get the species currently in WikiPathways with their respective URI's

4.2.2 List pathways and their species

4.2.3 List the species captured in WikiPathways and the number of pathways per species

4.2.4 List all pathways for species "Mus musculus"

4.2.5 Get all pathways with a particular gene

4.2.6 Get all groups and complexes containing a particular gene

4.2.7 Get all the genes on a particular pathway

4.2.8 Count the number of pathways per ontology term

4.2.9 Get all pathways with a particular ontology term

4.2.10 Get all ontology terms for a particular pathway

4.2.11 Get all Reactome pathways

4.2.12 Get all pathways with Pubmed references

4.2.13 Get all pathways with a particular Pubmed reference

4.2.14 Get all pathways and the number of references per pathway

4.2.15 Get a full dump of all pathways and their pathway ontological terms

4.3 Data statistics oriented queries

4.3.1 Count the number of metabolites per species

4.4 Interaction oriented queries



Federated SPARQL queries

```
SELECT ?pathway ?target ?assay ?smiles WHERE {  
  { SELECT DISTINCT ?pathway ?uniprot  
    iri( bif:concat("http://bio2rdf.org/uniprot:",  
      bif:regexp_substr('http://identifiers.org/uniprot/(.*)',?uniprot, 1))  
    ) as ?chembluniprot  
  WHERE {  
    ?s ?p ?uniprot ; dcterms:isPartOf ?pathway .  
    FILTER regex(?uniprot, "uniprot")  
  }  
}  
SERVICE <http://rdf.farmbio.uu.se/chembl/sparql/> {  
  ?target owl:sameAs ?chembluniprot .  
  ?score chembl:forTarget ?target .  
  ?assay chembl:hasTargetScore ?score .  
  ?activity chembl:onAssay ?assay ;  
    chembl:forMolecule ?molecule .  
  ?molecule bo:smiles ?smiles .  
}  
}
```



Computer-Assisted Curation

```
SELECT DISTINCT ?homepage ?label ?identifier
WHERE {
```

```
  ?gene dc:source "Ensembl"^^xsd:string ;
        rdfs:label ?label ;
        dcterms:identifier ?identifier ;
        dcterms:isPartOf ?pathway ;
        a wp:GeneProduct .
```

```
  FILTER (!strStarts(?identifier, "ENSRN"))
```

```
  ?pathway wp:organism <http://purl.obolibrary.org/obo/NCBITaxon_10116> ;
        foaf:page ?homepage .
```

```
}
```

Test Result

6 failures (±0)

594 tests (±0)

Took 2 min 26 sec.

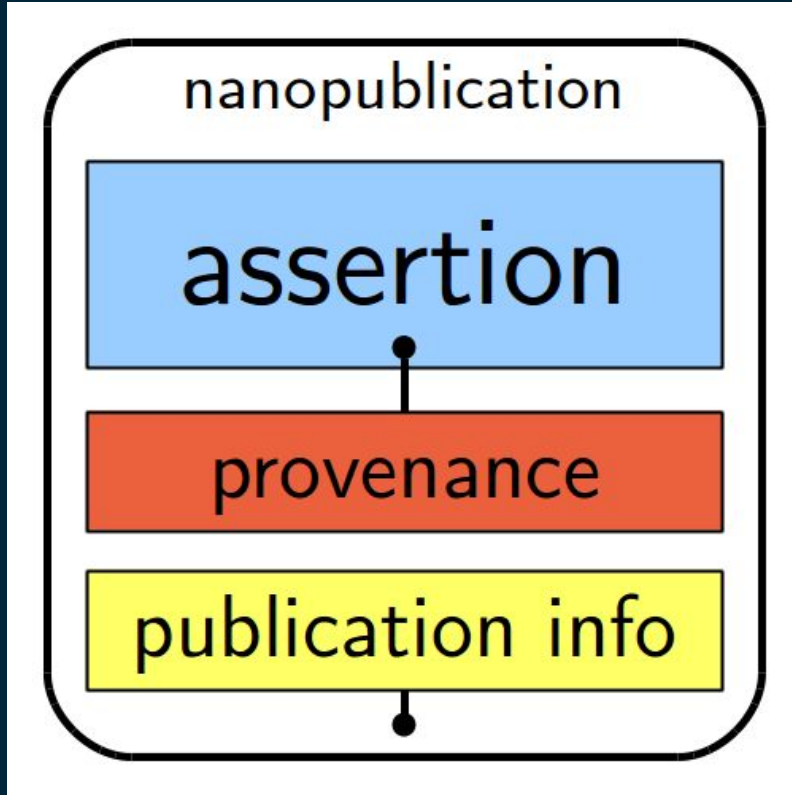
[add description](#)

All Failed Tests

| Test Name | Duration | Age |
|---|----------|-----|
| nl.unimaas.bigcat.wikipathways.curator.EnsemblGenes.outdatedIdentifiers | 0.93 sec | 1 |
| nl.unimaas.bigcat.wikipathways.curator.EnsemblGenes.outdatedIdentifiers | 0.54 sec | 1 |
| nl.unimaas.bigcat.wikipathways.curator.EnsemblGenes.outdatedIdentifiers | 0.53 sec | 1 |
| nl.unimaas.bigcat.wikipathways.curator.EnsemblGenes.outdatedIdentifiers | 0.33 sec | 1 |
| nl.unimaas.bigcat.wikipathways.curator.EnsemblGenes.outdatedIdentifiers | 36 ms | 1 |
| nl.unimaas.bigcat.wikipathways.curator.EnsemblGenes.outdatedIdentifiers | 0.27 sec | 1 |



Nanopublications



Groth P, Gibson A, Velterop J. The anatomy of a nanopublication. *Information Services & Use*. 2010 1;30(1-2):51-6.



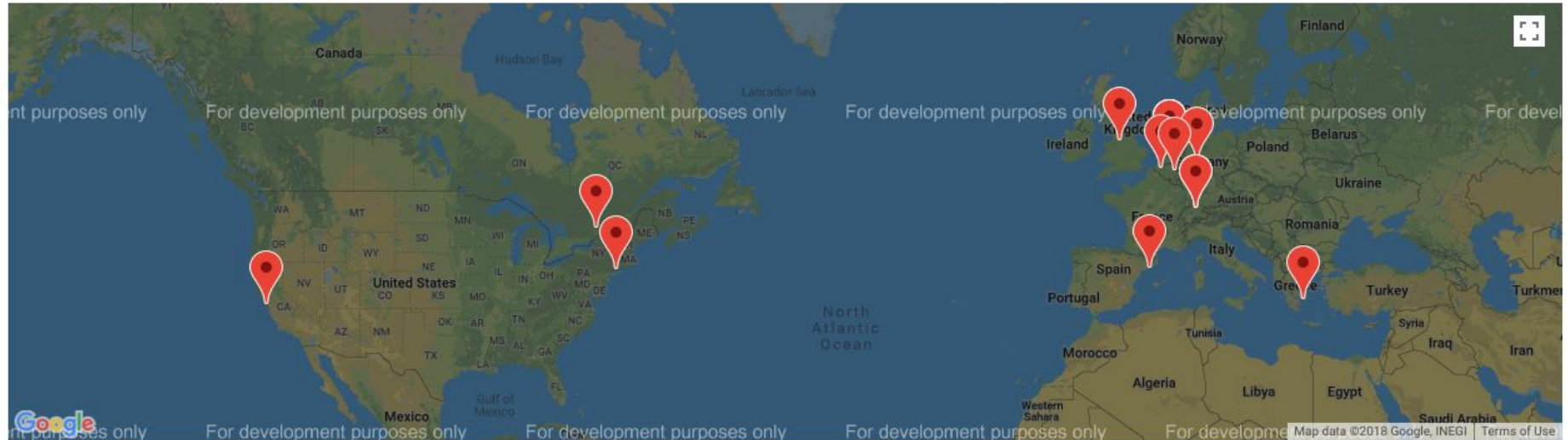
WikiPathways Nanopublications

```
sub:assertion {  
  chebi:16670 a wp:Metabolite . wormbase:ZC376.7 a wp:Protein .  
  interaction:id482a1806 wp:organismName "Caenorhabditis elegans" ;  
    wp:source wormbase:ZC376.7 ; wp:target chebi:16670 ; a wp:Interaction .  
}  
sub:provenance {  
  wp:WP525_r84475 dc:title "Mitochondrial Unfolded-Protein Response (UPRmt)"@en .  
  sub:assertion has-source: pmid:22700657 , wp:WP525_r84475 .  
}  
sub:publicationInfo {  
  sub:activity a prov:Activity ; prov:atLocation wd:Q1137652 ;  
  prov:used <https://github.com/wikipathways/nanopublications> .  
  this: dcterms:created "2018-10-10"^^xsd:date ; dcterms:rights cczero: ;  
  prov:wasDerivedFrom wp:WP525_r84475 ; prov:wasGeneratedBy sub:activity .  
}
```



Nanopub Monitor

16 nanopub server instances running on 12 distinct physical servers containing at least 10822105 unique nanopubs



| URL | Status | OK Ratio | Resp Time (Dist) | Last Seen OK | NP Count | IP Address | Server Location |
|---|--------|-----------|------------------|---------------------|----------|----------------|--------------------------|
| http://sprout038.sprout.yale.edu/nanopub-server/ | OK | 99.78035% | 300 ms (6212 km) | 2018-10-10 06:55:13 | 10822105 | 128.36.124.48 | New Haven, United States |
| http://np.inn.ac/ | OK | 95.38783% | 7 ms (0 km) | 2018-10-10 06:55:12 | 10822105 | 129.132.255.27 | Zurich, Switzerland |
| http://npx1.inn.ac/ | OK | 95.39681% | 4 ms (0 km) | 2018-10-10 06:55:11 | 676640 | 129.132.255.27 | Zurich, Switzerland |
| http://app.tkuhn.eculture.labs.vu.nl/nanopub-server-1/ | OK | 99.37364% | 36 ms (612 km) | 2018-10-10 06:55:11 | 2705299 | 130.37.193.11 | Amsterdam, Netherlands |
| http://app.tkuhn.eculture.labs.vu.nl/nanopub-server-2/ | OK | 99.36939% | 60 ms (612 km) | 2018-10-10 06:55:10 | 2706199 | 130.37.193.11 | Amsterdam, Netherlands |



Conclusions / Outlook

- Yet another output format
 - SPARQL for querying
 - Curation
 - Nanopublications
- Outlook: WPRDF 2.0
 - More FAIR (e.g. more useful IRIs)
 - More biology (e.g. taking advantage of newer GPML)
 - ...
- github.com/wikipathways/GPML2RDF



Acknowledgments

- WikiPathways RDF
 - Andra Waagmeester
 - Gladstone Institutes: Alex Pico et al.
 - Maastricht University: Chris Evelo et al.
- Nanopublications
 - Tobias Kuhn (VU Amsterdam)

