

New Paper: “Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources”

Egon Willighagen 

Published June 25, 2016

Citation

Willighagen, E. (2016). New Paper: “Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources”. In *chem-bla-ics*. *chem-bla-ics*. <https://doi.org/10.59350/me1j9-t5g38>

Keywords

Wikipathways, Curation, Sparql, Rdf, Wikidata

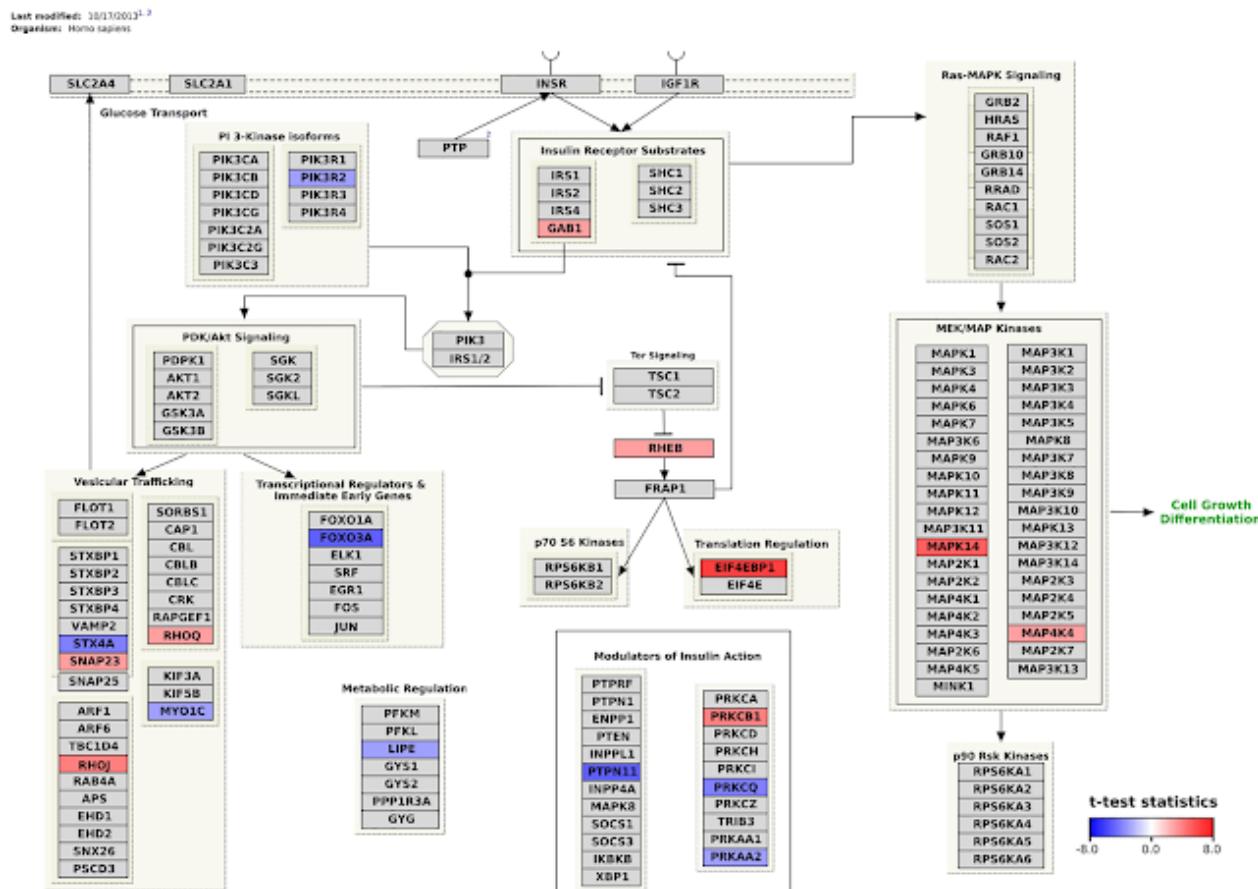
Abstract

Andra Waagmeester published a paper on his work on a semantic web version of the WikiPathways (doi:10.1371/journal.pcbi.1004989). The paper outlines the design decisions, shows the SPARQL endpoint, and several examples SPARQL queries. These include federates queries, like a mashup with DisGeNET (doi:10.1093/database/bav028) and EMBL-EBI’s Expression Atlas.

Copyright

Copyright © Egon Willighagen 2016. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Andra Waagmeester published a paper on his work on a semantic web version of the WikiPathways (doi:10.1371/journal.pcbi.1004989). The paper outlines the design decisions, shows the SPARQL endpoint, and several examples SPARQL queries. These include federates queries, like a mashup with DisGeNET (doi:10.1093/database/bav028) and EMBL-EBI's Expression Atlas. That results in nice visualisations like this:



If you have the relevant information in the pathway, these pathways can help a lot in helping understanding of what is biologically going on. And, of course, used for exactly that a lot.

Press release

Because press releases have become an interesting tool in knowledge dissemination, I wanted to learn what it involved to get one out. This involved the people as [PLOS Computational Biology](#) and the press offices of the Gladstone Institutes and our Maastricht University ([press release 1](#), [press release 2 EN/NL](#)). There is already one thing I learned in retrospect, and I am pissed with myself that I did not think of this: you should always have a graphics supporting your story. I have been doing this for a long time in my blog now (sometimes I still forget), but did not think of that in the press release. The press release was picked up by three outlets, though all basically as we presented it to them (thanks to [Altmetric.com](#)):

chem-bla-ics



So far, Altmetric has seen 3 news stories from 3 outlets.



Easy integration of biological knowledge improves understanding of diseases

Science Daily, 23 Jun 2016

Researchers have improved the integration of disparate sources and types of data which will advance scientists' understanding...



Easy integration of biological knowledge improves understanding of diseases

MedicalXpress, 23 Jun 2016

Researchers have improved the integration of disparate sources and types of data which will advance scientists' understanding...



Easy integration of biological knowledge improves understanding of diseases

Health Medicinet, 23 Jun 2016

Researchers have improved the integration of disparate sources and types of data which will advance scientists' understanding...

SPARQL

But what makes me appreciate this piece of work, and WikiPathways itself, is how it creates a central hub of biological knowledge. Pathway databases capture knowledge not easily embedded in generally structured (relational) databases. As such, expressing this in the RDF format seems simple enough. The thing I really love about this approach, is that your queries become machine readable stories, particularly when you start using human readable variants of SPARQL for this. And you can [share these queries with the online scientific community with, for example, myExperiment](#).

There are two applications how I have used SPARQL on WikiPathways data for metabolomics: 1. curation; 2. statistics. Data analysis is harder, because in the RDF world resources scientific lenses are needed to accommodate for the chemical structural-temporal complexity of metabolites. For curation, we have long used SPARQL for unit tests to support the curation of WikiPathways. Moreover, I have manually used the SPARQL endpoint to find curation tasks. But now that the paper is out, I can blog about this more. For now, [many examples SPARQL queries can be found in the WikiPathways wiki](#). It features several queries showing statistics, but also some for curation. This is an example query I use to improve the interoperability of WikiPathways with [Wikidata](#) (also for [BridgeDb](#)):

```
SELECT DISTINCT ?metabolite WHERE {  
  ?metabolite a wp:Metabolite .  
  OPTIONAL { ?metabolite wp:bdbWikidata ?wikidata . }  
  FILTER (!BOUND(?wikidata))  
}
```

Feel free to give this query a go at sparql.wikipathways.org!

Triptych

This paper completes a nice triptych of three papers about WikiPathways in the past 6 months. Thanks to whole community and [the very many contributors!](#) All three papers are linked below.