Check for updates

SOFTWARE TOOL ARTICLE

REVISED Explicit interaction information from WikiPathways in RDF facilitates drug discovery in the Open PHACTS Discovery Platform [version 2; peer review: 2 approved]

Ryan A. Miller [1], Peter Woollard [2], Egon L. Willighagen [1], Daniela Digles[3], Martina Kutmon [1,4], Antonis Loizou[5], Andra Waagmeester [1,6], Stefan Senger[2], Chris T. Evelo [1,4,7]

[1]Department of Bioinformatics (BiGCaT), Maastricht University, Maastricht, The Netherlands
[2]GlaxoSmithKline (GSK), Stevenage, UK
[3]Pharmacoinformatics Research Group, Department of Pharmaceutical Chemistry, University of Vienna, Vienna, Austria
[4]Maastricht Center for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands
[5]Blue Sky IT, Amsterdam, The Netherlands
[6]Micelio, Antwerp, Belgium
[7]Open PHACTS Foundation, Science Park, Cambridge, UK

Abstract

Open PHACTS is a pre-competitive project to answer scientific questions developed recently by the pharmaceutical industry. Having high quality biological interaction information in the Open PHACTS Discovery Platform is needed to answer multiple pathway related questions. To address this, updated WikiPathways data has been added to the platform. This data includes information about biological interactions, such as stimulation and inhibition. The platform's Application Programming Interface (API) was extended with appropriate calls to reference these interactions. These new methods of the Open PHACTS API are available now.

Keywords

Open PHACTS, drug discovery, semantic, bioinformatics, WikiPathways, pathway database, API

Open Peer Review

Reviewer Status ✔ ✔

	Invited Reviewers	
	1	**2**
version 2 (revision) 12 Oct 2018	✔ report	✔ report
version 1 17 Jan 2018	? report	? report

1. **Augustin Luna** [iD], Dana-Farber Cancer Institute, Boston, USA

2. **Yi-An Chen** [iD], National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Ryan A. Miller (ryan.miller@maastrichtuniversity.nl)

Author roles: Miller RA: Writing – Original Draft Preparation; **Woollard P**: Conceptualization; **Willighagen EL**: Conceptualization, Software, Supervision; **Digles D**: Data Curation; **Kutmon M**: Conceptualization; **Loizou A**: Software; **Waagmeester A**: Conceptualization; **Senger S**: Conceptualization; **Evelo CT**: Conceptualization, Supervision

How to cite this article: Miller RA, Woollard P, Willighagen EL *et al.* **Explicit interaction information from WikiPathways in RDF facilitates drug discovery in the Open PHACTS Discovery Platform [version 2; peer review: 2 approved]** F1000Research 2018, **7**:75 https://doi.org/10.12688/f1000research.13197.2

First published: 17 Jan 2018, **7**:75 https://doi.org/10.12688/f1000research.13197.1

Introduction

Targeting proteins to ideally restore normal biological processes is a common starting point in drug discovery[1]. The Open PHACTS Discovery Platform (OPDP) was designed to help identify protein targets and information about their associations with each other[2–4]. The OPDP supports target identification and validation by including target-target interactions from WikiPathways[5–7]. Of these interaction networks, proteins sharing a downstream path allows investigation of alternative drug target combinations. Even the knowledge of which biological pathways participate in disease-related processes provides insight in the pathway topology between the targets. The importance and need of providing access to interaction information for real-world research questions was outlined in a recent Open PHACTS paper[8].

The Open PHACTS project was born out of the desire to integrate pharmacological data from multiple precompetitive sources to efficiently address scientific questions that cannot be answered with single data sources[8]. It integrates data using linked data approaches[3] from chemical and biological sources such as ChEBI, ChEMBL, UniProt, and WikiPathways[6]. However, the OPDP did not previously include calls to access specific up- and downstream interaction effects. This information is needed for questions related to drug repositioning and repurposing. Up- or downstream targets may be interesting alternatives with similar therapeutic effect to targets, for which it is particularly hard to develop a drug agent. Thus, finding a target that has already been drugged or is more drug tractable will be advantageous. Here we describe how to identify alternative targets in the same cellular pathway using OPDP against the WikiPathways data.

Methods

Implementation

The WikiPathways Resource Description Framework data (WPRDF) is released as part of the monthly releases[5]. The native format for WikiPathways is Graphical Pathway Markup Language (GPML) based on the eXtensible Markup Language (XML) standard. The RDF export is transformed from the original GPML. In the RDF representation we use two distinct controlled vocabularies, to distinguish between the graphical notation of a pathway and the biological meanings expressed in the pathway. This is done to allow integration with other pathway repositories which use other graphical notations or none. The WikiPathways RDF also includes details about directed and undirected interactions. Directed biochemical interactions capture the source and target which are depicted as an arrow in simple pathway drawings. WikiPathways adds biological meaning to interactions with Molecular Interaction Map (MIM) interaction types, like inhibitions, enzyme catalyzed reactions, and stimulations[9], as well as Systems Biology Graphical Notation (SBGN) interactions[10]. Reactome pathways in WikiPathways use SBGN interactions[11,12]. However, because MIM and SBGN use different drawing styles, we normalize their inhibition types into a common inhibition type, defined by the WikiPathways ontology (https://vocabularies.wikipathways.org/wp).

The WikiPathways basic drawing tools also contain generic arrows and T-bar annotations that give the user the ability to create basic diagrams without the semantic meaning of MIM or SBGN notations. The interactions connecting these nodes are captured, but the only explicit information is that it is a directed interaction from a source to a target. To handle more complicated enzyme reaction drawings, where there is not a single line that directly connects targets in a cascade of enzymatic reactions, a query was developed that recognizes these types of reactions. However, this is not implemented in the current Open PHACTS Application Programming Interface (API).

Version 2.1 of the OPDP API contains three new calls for interactions and their pathways. The first call, */pathway/getInteractions*, returns all interactions involved in a pathway. To use this feature, the user specifies a pathway URI and OPDP returns its interactions including information about direction and the connected entities. The direction information is relayed as a starting node having a *wp:source* annotation, while the end of the interaction has the *wp:target* annotation. In its simplest form, this means that if gene product A is interacting with a gene product B, then we have *wp:source* for product A and *wp:target* for product B. However, the presented new methods also support interactions with multiple sources and targets for more complex interactions that are more accurately represented this way.

The second added call, */pathways/interactions/byEntity*, returns the direction of the interactions involving this entity. An entity is specified by a URI and can be a metabolite, protein, gene product, or RNA. API options allow the user to select only upstream or only downstream interactions. If a direction is not specified in the call, all the adjacent interactions will be retrieved regardless of their direction. The results also specify the interaction type (e.g. inhibition, stimulation, conversion). Vocabularies.wikipathways.org also identifies catalysis and binding events as well as a more generic directedInteraction in

the case where the type of the interaction is not identified. This ability to select the interaction direction is specifically what allows users to answer scientific questions around upstream and downstream effects, such as those defined by Open PHACTS. The third API call is */pathways/interactions/byEntity/count* which is a helper function that returns the number of interactions for a target.

Operation

The OPDP API calls are backed by SPARQL searches against the loaded WikiPathways RDF. The query parameters that are required or optional are given in the documentation of Open PHACTS (https://dev.openphacts.org/docs/2.1). As in previous versions, the API uses HTTP GET to call methods and needs a (free) application ID and key (see https://dev.openphacts.org/signup)[3].

To ensure multiple URI schemes can be used to identify genes, proteins, and metabolites, the Open PHACTS platform uses an Identifier Mapping Service (IMS)[6]. This ensures that people can use Ensembl, NCBI Gene, and others for genes, UniProt, Ensembl, etc. for proteins, and HMDB, ChEBI, CAS registry number, and PubChem for metabolites. Furthermore, it supports identifiers.org formatted URIs, further simplifying entering identifiers[13].

Example queries

We are demonstrating the platform with three example calls. All the API calls require use of an application ID and an application key. This key and ID can be acquired by creating a free Open PHACTS account. The first example is an application to the PI3K/AKT pathway for cell growth regulation which contain important targets for cancer treatment[14]. The AKT protein has a central role and usefully shows the API call's ability to return connected elements with the */pathways/interactions/byEntity* and the */pathway/getInteractions* calls. The API calls can help aid drug discovery by taking a target, in this case AKT, and easily identify other connected proteins that could potentially be used as drug targets with a common downstream effect.

Figure 1 shows the web interface of the API call that returns the connectivity of the AKT2 target to both upstream or downstream proteins or gene products. This method allows the user to identify connections to other targets in the pathway. The results of that API call (Figure 2) show the AKT2 interaction with microRNA. A helper method (Figure 3): */pathways/interactions/byEntity/count* is also included. It returns the number of all interactions in which an entity is participates. This helps the user get a sense of the prevalence of the queried entity with interactions in pathways found on WikiPathways. An example result for this query can be found in Supplementary Figure 1.

The other call implemented, */pathway/getInteractions* (Figure 4), demonstrates an API call to return all interactions in the MicroRNAs in cardiomyocyte hypertrophy pathway[15]. This pathway has interaction details for AKT, mTOR, and PI3K, which are all important targets in cancer research[16]. For each interaction the participants are given and whether it is a directed

Figure 1. Parameters (bottom) and *curl* command (top) for the GET */pathways/interactions/byEntity* call. The GET portion tells the API to retrieve data with the associated call. It takes an entity URI, the Ensembl ID for AKT2, and returns a list interactions for AKT2. The obligatory parameters are shown in bold. Entity IDs that are acceptable for queries include Ensembl, Entrez Gene, and UniProt for genes, proteins, and RNAs. For metabolites the ID sources HMDB, ChEBI, and ChemSpider, for example, are acceptable entity IDs

Example query results for */pathways/interactions/byEntity* **call for AKT2**

```
"items": [
  {
    "_about": "http://rdf.wikipathways.org/Pathway/WP1544_r75258/WP/Interaction/id28bfdd47",
    "isPartOf": {
      "_about": "http://identifiers.org/wikipathways/WP1544",
      "title_en": "MicroRNAs in cardiomyocyte hypertrophy",
      "title": "MicroRNAs in cardiomyocyte hypertrophy",
      "inDataset": "http://www.wikipathways.org",
      "latest_version": "http://identifiers.org/wikipathways/WP1544_r75258",
      "pathway_organism": {
        "_about": "http://purl.obolibrary.org/obo/NCBITaxon_9606",
        "inDataset": "http://www.wikipathways.org",
        "label": "Homo sapiens"
      }
    },
    "inDataset": "http://www.wikipathways.org",
    "source": {
      "_about": "http://identifiers.org/ensembl/ENSG00000207875",
      "inDataset": "http://www.wikipathways.org",
      "type": [
        "http://vocabularies.wikipathways.org/wp#GeneProduct",
        "http://vocabularies.wikipathways.org/wp#Rna"
      ]
    },
    "target": {
      "_about": "http://identifiers.org/ncbigene/208",
      "inDataset": "http://www.wikipathways.org",
      "type": [
        "http://vocabularies.wikipathways.org/wp#Protein",
        "http://vocabularies.wikipathways.org/wp#GeneProduct"
      ]
    },
    "type": [
      "http://vocabularies.wikipathways.org/wp#DirectedInteraction",
      "http://vocabularies.wikipathways.org/wp#Inhibition"
    ]
  }
}
```

Figure 2. Result in the JSON format of the AKT2 query from Figure 1. The participants of the interaction are directed from source (hsa-let7b) to target (AKT2). It also shows the type of interaction (inhibition), and the biological types of the interaction participants.

Example input for */pathways/interactions/byEntity/count* **call for AKT2**

Curl

```
curl -X GET --header "Accept: application/json" "https://beta.openphacts.org/2.1/pathways/interactions/byEntity/count?
uri=http%3A%2F%2Fidentifiers.org%2Fensembl%2FENSG00000105221&app_id=0a081d11&app_key=df2facbe3d5cee743dc500a1589e53bf"
```

Request URL

```
https://beta.openphacts.org/2.1/pathways/interactions/byEntity/count?
uri=http%3A%2F%2Fidentifiers.org%2Fensembl%2FENSG00000105221&app_id=0a081d11&app_key=df2facbe3d5cee743dc500a1589e53bf
```

Parameters

Parameter	Value	Description	Parameter Type	Data Type
uri	http://identifiers.org/ensembl/ENSG0000010	An entity URI.	query	string
app_id	0a081d11	Your access application id	query	string
app_key	df2facbe3d5cee743dc500a1589e53bf	Your access application key	query	string
pathway_organism		The rdfs:label for the pathway organism (URL encode). e.g.: Homo sapiens.	query	string
direction	▼	One of: 'up', 'down'. 'up' returns only Interactions where the specifies entity is the target, and 'down' returns only Interactions where the specified entity is the source.	query	string
interaction_type	▼	The type of interaction to return. One of: 'directed', 'undirected'.	query	string

Figure 3. Parameters (bottom) and *curl* command (top) for the GET */pathways/interactions/byEntity/count* call. It takes a URI for an entity, in this case the Ensembl ID for AKT2 and returns a count of the interactions to which this gene product is involved. Only the entity URI, app ID, and app key are required fields. Optional parameters are pathway organism, direction, or type of interaction.

Example input for */pathway/getInteractions* call for MicroRNAs in cardiomyocyte hypertrophy pathway

Figure 4. Parameters (bottom) and *curl* command (top) for the /*pathway*/*getInteractions* call. It is intended to take the pathway URI from WikiPathways and return a list of interaction involved in that particular pathway. Pathway URI, app ID, and app key are the only required values for this call.

or undirected interaction. An example result for this query can be seen in Supplementary Figure 2.

Example workflows
In order to demonstrate the basic use of the introduced API methods, we developed two workflows, available in the Supplementary Material. One uses Python to return a file with the results in a table and the other uses a HTML webpage using the ops.js JavaScript client library[17]. More involved workflows have been developed for KNIME and Pipeline Pilot[18,19].

The Python script example uses the Open PHACTS */pathway/getInteraction* API call and prompts the user to enter a WikiPathways pathway number that they wish to query, such as 1544 for WikiPathways pathway WP1544. Invocation of the API call with the pathway identifier returns information about the directed interactions that are involved with the pathway. The information that is returned is the interaction ID used by WikiPathways, the interaction type, and URIs for the source and target of the interaction. In order to convert the URIs into something more readable, a SPARQL query is then executed to get labels, from the WikiPathways SPARQL endpoint, for the source and target of the interaction. The results are written to a file with the interaction ID, interaction type, URIs for the source and target, as well as alias IDs, the *curl* for the API call, the pathway ID used, and a number of interactions returned.

The second example uses a HTML5 webpage and the ops.js JavaScript client library to retrieve interactions for a particular gene, using the URI for the gene's Ensembl identifier and the */pathways/interactions/byEntity* API method. The ops.js library passes the returned JSON with interaction information to a callback function, where the interacting source and target are extracted and the interacting entity determined. For each interacting entity, which may be a protein, RNA, or small compound, a call to the */pathways/interactions/byEntity/count* method is made to return the number of interaction that entity has.

Summary
While the calls identified here are simple calls, workflow tools make it possible to take advantage of the integrative nature of the OPDP to make API calls in succession. Two such workflow tools that work with the OPDP are KNIME and Pipeline Pilot. With these tools, it is possible to perform a directional query of a target and identify alternative targets that can then be queried against the chemistry calls to identify active compounds for these alternative targets. The client libraries ops.js, ops4j, and ropenphacts also support Open PHACTS and the interaction calls for pathways. This allows users to perform API calls to the OPDP using their preferred language or platform, such as JavaScript, Java, or R.

The addition of interactions with direction information allows OPDP to answering more of the pre-defined scientific questions[2]. The directional information allows the user to explore how proteins and gene products are connected with one another and easily access this information. This is illustrated in the example queries using the cancer target AKT.

Software availability
Online service: https://dev.openphacts.org/docs/2.1

Latest source code is available at: https://github.com/openphacts/OPS_LinkedDataApi

Archived source code of discussed version: https://doi.org/10.5281/zenodo.1068252[20]

License: Apache License 2.0

Grant information

Acknowledgments

Supplementary material
Supplementary File 1. Contains two additional figures that show example API call output and contains the workflows to demonstrate the use of the APIs.

Click here to access the data

References

1. Schreiber SL: **Target-oriented and diversity-oriented organic synthesis in drug discovery.** *Science.* 2000; **287**(5460): 1964–1969.
   PubMed Abstract | Publisher Full Text

2. Azzaoui K, Jacoby E, Senger S, *et al.*: **Scientific competency questions as the basis for semantically enriched open pharmacological space development.** *Drug Discov Today.* 2013; **18**(17–18): 843–852.
   PubMed Abstract | Publisher Full Text

3. Williams AJ, Harland L, Groth P, *et al.*: **Open PHACTS: semantic interoperability for drug discovery.** *Drug Discov Today.* 2012; **17**(21–22): 1188–1198.
   PubMed Abstract | Publisher Full Text

4. Digles D, Zdrazil B, Neefs JM, *et al.*: **Open PHACTS computational protocols for *in silico* target validation of cellular phenotypic screens: knowing the knowns.** *Medchemcomm.* 2016; **7**(6): 1237–1244.
   PubMed Abstract | Publisher Full Text | Free Full Text

5. Waagmeester A, Kutmon M, Riutta A, *et al.*: **Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources.** *PLoS Comput Biol.* 2016; **12**(6): e1004989.
   PubMed Abstract | Publisher Full Text | Free Full Text

6. Gray AJG, Groth P, Loizou A, *et al.*: **Applying linked data approaches to pharmacology: Architectural decisions and implementation.** *Semant Web.* 2014; **5**(2): 101–113.
   Publisher Full Text

7. Kelder T, Pico AR, Hanspers K, *et al.*: **Mining biological pathways using WikiPathways web services.** *PLoS One.* 2009; **4**(7): e6447.
   PubMed Abstract | Publisher Full Text | Free Full Text

8. Chichester C, Digles D, Siebes R, *et al.*: **Drug discovery FAQs: workflows for answering multidomain drug discovery questions.** *Drug Discov Today.* 2015; **20**(4): 399–405.
   PubMed Abstract | Publisher Full Text

9. Luna A, Karac EI, Sunshine M, *et al.*: **A formal mim specification and tools for the common exchange of mim diagrams: an xml-based format, an api, and a validation method.** *BMC Bioinformatics.* 2011; **12**(1): 167.
   PubMed Abstract | Publisher Full Text | Free Full Text

10. Le Novère NL, Hucka M, Mi H, *et al.*: **The systems biology graphical notation.**
    Nat Biotechnol. 2009; **27**(8): 735–741.
    PubMed Abstract | Publisher Full Text

11. Kutmon M, Riutta A, Nunes N, *et al.*: **WikiPathways: capturing the full diversity of pathway knowledge.** *Nucleic Acids Res.* 2016; **44**(D1): D488–D494.
    PubMed Abstract | Publisher Full Text | Free Full Text

12. Croft D, Mundo AF, Haw R, *et al.*: **The Reactome pathway knowledgebase.** *Nucleic Acids Res.* 2014; **42**(Database issue): D472–D477.
    PubMed Abstract | Publisher Full Text | Free Full Text

13. Juty N, Le Novère N, Hermjakob H, *et al.*: **Towards the collaborative curation of the registry underlying Identifiers.org.** *Database (Oxford).* 2013; **2013**: bat017.
    PubMed Abstract | Publisher Full Text | Free Full Text

14. Vanhaesebroeck B, Stephens L, Hawkins P: **PI3K signalling: the path to discovery and understanding.** *Nat Rev Mol Cell Biol.* 2012; **13**(3): 195–203.
    PubMed Abstract | Publisher Full Text

15. Levels M, Hanspers K, Kutmon M, *et al.*: **Micrornas in cardiomyocyte hypertrophy (Homo sapiens).** 2017.
    Reference Source

16. Li H, Zeng J, Shen K: **PI3K/AKT/mTOR signaling pathway as a therapeutic target for ovarian cancer.** *Arch Gynecol Obstet.* 2014; **290**(6): 1067–1078.
    PubMed Abstract | Publisher Full Text

17. Dunlop I, Willighagen E, Elblood AW, *et al.*: **ops.js: Ops.js 7.1.0 for Open PHACTS 2.2 api** *Zenodo.* 2018.
    Publisher Full Text

18. Berthold MR, Cebron N, Dill F, *et al.*: **KNIME: The Konstanz Information Miner.** In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007).* Springer, 2007; 319–326.
    Publisher Full Text

19. Ratnam J, Zdrazil B, Digles D, *et al.*: **The application of the open pharmacological concepts triple store (Open PHACTS) to support drug discovery research.** *PLoS One.* 2014; **9**(12): e115460.
    PubMed Abstract | Publisher Full Text | Free Full Text

20. fundatureanu-sever, Kerber R, Soiland-Reyes S, *et al.*: **openphacts/OPS_ LinkedDataApi: Open PHACTS Linked Data API 2.1.0 (Version 2.1.0).** *Zenodo.* 2016.
    http://www.doi.org/10.5281/zenodo.1068252

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 23 October 2018

https://doi.org/10.5256/f1000research.17945.r39411

✔ **Augustin Luna** 🆔

Dana-Farber Cancer Institute, Boston, MA, USA

The authors satisfactorily addressed my concerns.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 19 October 2018

https://doi.org/10.5256/f1000research.17945.r39412

✔ **Yi-An Chen** 🆔

National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan

The author has addressed all the issues. I have no further comments.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 03 April 2018

? **Yi-An Chen** 🆔

National Institutes of Biomedical Innovation, Health and Nutrition, Osaka, Japan

The authors add new features to the Open PHACTS APIs which facilitate querying reactions in WikiPathways.

Some comments:
1. In the first example, the query for AKT2 in the figure 1 uses http://identifiers.org/ensembl/ENSG00000105221 but the result in the figure 2 shows http://identifiers.org/ncbigene/208. I guess the API should be able to accept different commonly used identifiers, however, either the article or the web interface mention about the acceptable identifiers. In addition, It would be nice if the input of the query accepts simple identifiers, instead of constructing the full URI.

2. The result in figure 2 is difficult to interpret. It would be nice to expand the query to get the basic information for those interacting entities. For example, the gene symbol or the compound name.

3. The examples in figure 3 and figure 4 didn't show the query results.

4. Simple queries in the examples may be difficult to show the usefulness of the new implemented APIs. A more sophisticated application which contains either bulk query or pipeline query should be helpful for the readers to understand the demand of using the APIs.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 11 Oct 2018

Ryan Miller, Maastricht University, Maastricht, The Netherlands

1. The Open PHACTS API includes the Identifier Mapping Service component which allows use of many identifier schemes, as long as suitable link sets are available (at http://data.openphacts.org/1.5/ims/linksets/). These sets include many of the popular gene, protein, and metabolite identifier:for genes, proteins, and RNAs Ensembl, Entrez Gene, UniProt, and others; for metabolites the ID sources can include HMDB, ChEBI, PubChem, CAS registry numbers, and ChemSpider.

2. Retrieving specific gene symbol or compound names aside from the identifier URI, resource, and identifier can be part of a workflow in workflow tools like KNIME or Pipeline Pilot. We did create a simple workflow to address this. In this example, we use the /pathway/getInteractions API call to return the directed interactions for a pathway and the resulting IDs for the interaction and then returned the more human readable labels for the URIs.

3. Additional example output figures were added to the supplementary materials section to reflect the addition of example query results for the remaining calls.

4. We added an Example Workflows section describing two simple workflows. More complex applications and workflows have been recently published for KNIME and Pipeline Pilot, which we added as reference 19.

Competing Interests: No competing interests were disclosed.

Reviewer Report 24 January 2018

https://doi.org/10.5256/f1000research.14318.r29933

?

Augustin Luna (iD)

Dana-Farber Cancer Institute, Boston, MA, USA

The tool provides additional features to the Open PHACTS platform to simplify the access of interaction information from WikiPathways. Some comments:

- The authors go into some detail about the MIM/SBGN representations used by WikiPathways. To my understanding, WikiPathways makes use of the format GPML as a data storage format. Some description of the relationship of GPML and the WPRDF format would be nice for readers to understand the data processing involved.

- Looking at the code repository, the API interface seems to be built using Swagger. Is there the plan to provide supported client packages to access the API from various programming languages? Are users encouraged to produce their own using the swagger-codegen (https://github.com/swagger-api/swagger-codegen) project? Have the authors had any experience with these auto-generated packages in the context of their API?

- The sentence "However, the new methods also support interactions with multiple sources and targets.", are these the methods that are elsewhere described as not implemented currently?

- What entity IDs are acceptable? The examples show ENSEMBL is this the only acceptable ID source? Does the documentation provide any examples for ID conversion from much more common gene symbols, for example?

- I assume that when "up" or "down" interactions are not specified for return then both are returned is this correct?

- Are the example interactions (i.e. inhibition, stimulation, conversion) the only ones available in the API? If not, is there a listing of the interaction types with description of the interaction type.

- It would be more clear to just state the call function name rather than "the first and third calls" and "the api call" in the example queries section.

- Additional examples of the result output for all described function calls would be helpful for the described API calls.

- Additional examples that highlight the integrative nature of Open PHACTS would be nice. For example, to show how results of the currently described API calls can paired the other data available, such as the CHEBI data from the platform.

Is the rationale for developing the new software tool clearly explained?

Yes

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow

replication of the software development and its use by others?

Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 11 Oct 2018

Ryan Miller, Maastricht University, Maastricht, The Netherlands

1. **WikiPathways is represented using GPML (Graphical Pathway Markup Language), which is a modified XML format. The WikiPathways RDF is divided into two divisions, the GPML and the WPRDF. The GPML side is the portion that is responsible for the drawing and graphical elements. The WPRDF side is there to describe the semantic elements of the RDF. We have updated the Methods section and clarified this in the first "Implementation" paragraph.**
2. **The API is REST-like and dedicated clients are not required. However, a few client libraries are already available for various languages and platforms: JavaScript (ops.js), Java (ops4j), ropenphacts, and KNIME nodes. Personal experience with auto-generated packages has been negative, but based on code generation in general, not the Open PHACTS API. Supported libraries are referenced in the summary sections along with the workflow environments.**
3. **We intended to clarify that this simple example referred to single-source, single-target interactions, but that the framework also supports interactions with more than one source and/or target. Such complex interactions occur frequently. The last sentence of the third paragraph under Methods: Implementation has been updated to make this more clear.**
4. **The Open PHACTS API includes the Identifier Mapping Service component which allows use of many identifier schemes, as long as suitable link sets are available (at http://data.openphacts.org/1.5/ims/linksets/). These sets include many of the popular gene, protein, and metabolite identifier:for genes, proteins, and RNAs Ensembl, Entrez Gene, UniProt, and others; for metabolites the ID sources can include HMDB, ChEBI, PubChem, CAS registry numbers, and ChemSpider. The caption of Figure 1 has been extended and a paragraph added to Operation section to reflect this.**
5. **Yes, you are correct: if the user does not specify the directions for the**

interactions, all the immediate interactions are retrieved regardless of their direction. This can be found in the last paragraph of the Methods section and updated here to clarify.

6. **Vocabularies.wikipathways.org also identifies catalysis and binding events as well as a more generic directedInteraction in the case where the type of the interaction is not identified. This can be found in the last paragraph of the Methods: Implementation.**

7. **The text in the under Example Queries section has been updated to use the names of the API calls used.**

8. **Additional figures that show example outputs for the remaining figures can be found in the supplemental info section.**

9. **Workflow tools make it possible to take advantage of the integrative nature of the OPDP to make API calls in succession. With these tools, it is possible to perform a directional query of a target and identify alternative targets that can then be queried against the chemistry calls to identify active compounds for these alternative targets. An Example Workflows section has been added describing two simple workflows and the code for the workflows can be found in the supplementary information.**

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias

- You can publish traditional articles, null/negative results, case reports, data notes and more

- The peer review process is transparent and collaborative

- Your article is indexed in PubMed after passing peer review

- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research