

Biological pathway abstractions

Citation for published version (APA):

Waagmeester, A. S. (2024). *Biological pathway abstractions: from two-dimensional drawings to multidimensional linked data*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20240116aw>

Document status and date:

Published: 01/01/2024

DOI:

[10.26481/dis.20240116aw](https://doi.org/10.26481/dis.20240116aw)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

This thesis explored methods to extract biological knowledge for pathway curation efficiently. While many platforms host scientific knowledge, scientific literature remains the central pillar of existing scientific knowledge. Scientific literature, for quite some time now, has been growing at an ever-increasing pace. With existing access constraints, this growth poses challenges for researchers to obtain knowledge efficiently.

The use case in this thesis is pathway curation. Pathways are like biological maps that show how different parts of a living organism work together. In this thesis, I explore other methods to facilitate and structure knowledge so that a pathway curator has unrestricted access to existing knowledge next to access to scientific knowledge.

In Chapter 2, we explore biological pathways and how the field of bioinformatics helps in pathway curation. Bioinformatics plays a role in organizing biological knowledge in formats that serve automatic pipelines. We also mention that pathways evolved from mere illustrations in discussions among peers towards full-fledged models that can be used in research pipelines to process and assess findings. Chapter 3 describes a text-mining pipeline that semi-automatically identifies potential pathway parts. These parts were presented to pathway curators, who validated their relevance and extended the scrutinized pathway with novel features. Text mining leads to structured data from loosely structured. Other approaches from the field of Bioinformatics lead to a myriad of structured data formats. Chapter 4 presents a representation of pathway knowledge on a framework that, next to the data, also formally captures its semantics. This framework makes rapid integration with other structured data sources possible. Chapter 5 follows up on the linked-data concept introduced in Chapter 4. In that chapter, we describe how an online linked-data platform, called Wikidata, is used as a hub in linking similarly structured linked data resources, as the one described in chapter 4. Finally, Chapter 6 describes a protocol to document the linked data. This is because while the knowledge is structured and formalized as linked data, the underlying schema of linked data

remains implicit. By using formal language describing linked-data schemas, curators and data owners alike can formally describe expectations from the data and what is offered by the data. This allows for rapid validation of expectations.

The thesis started its exploration from loosely structured scientific knowledge towards a structured representation of data and expectations.