Maastricht University

NUTRIM

NUTRIM School of Nutrition and Translational Research in Metabolism

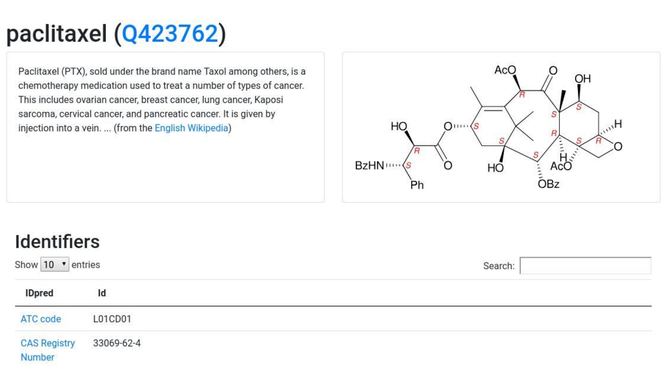
Wikidata and Scholia as a hub linking chemical knowledge

Egon Willighagen[A], Denise Slenter[A], Daniel Mietchen[B], Chris Evelo[A,C], Finn Nielsen[D]

[A] Department of Bioinformatics - BiGCaT, Maastricht University, The Netherlands,
[B] Data Science Institute, University of Virginia, Charlottesville, Virginia, USA,
[C] Maastricht Centre for Systems Biology - MaCSBio, Maastricht University, The Netherlands,
[D] Cognitive Systems, DTU Compute, Technical University of Denmark, Denmark

Introduction

Making chemical databases more FAIR (findable, accessible, interoperable, and reusable) benefits computational chemistry and cheminformatics. We here discuss Wikidata, a young sister project of Wikipedia, with one key difference: it is a machine readable database, making it far more useful for interoperability of molecular databases in systems biology [1,2]. Thanks to the WikiProject Chemistry community on Wikidata, there is a growing amount of information about chemical compounds.



Methods

Scholia is a Python/Flask-based server system that creates webpages using a template approach [5]. It defines templates for concepts around knowledge exchange, such as publications, journals, publishers, but also topics. It uses SPARQL queries against the Wikidata Query Service (WDQS, query.wikidata.org) and visualizes the data in various forms. Furthermore, we used a combination of Bioclipse (bioclipse.net) and QuickStatements to add missing chemical compounds for biological pathways from WikiPathways [6]. Where needed, new Wikidata properties were proposed.

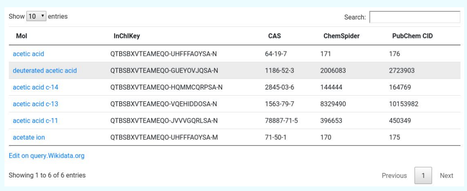
Results

We here introduce our contributions to the WikiProject Chemistry to support FAIR-ification of open chemical knowledge. For example, we proposed new Wikidata properties to annotate compounds with external database identifiers for the EPA CompTox Dashboard [3], the SPLASH [4], and MetaboLights. We also introduced a Scholia extension [5], visualizing data about chemicals and chemical classes:

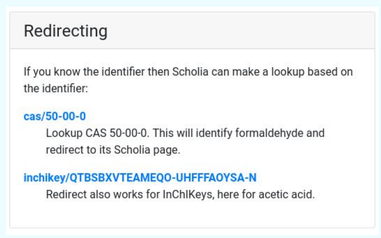
https://tools.wmflabs.org/scholia/

Provenance: "stated in"

Related compounds



Lookup by identifier

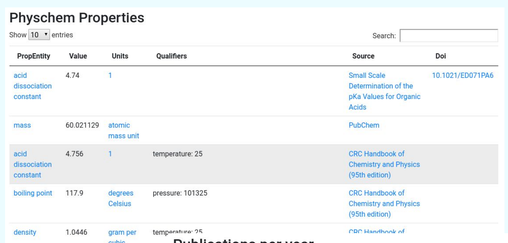


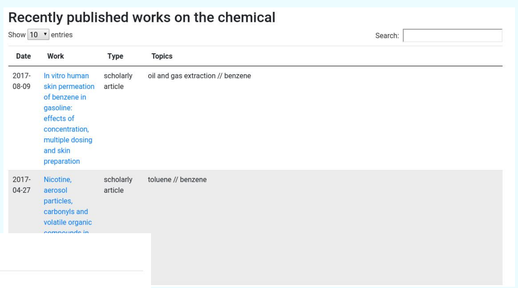
Identifiers

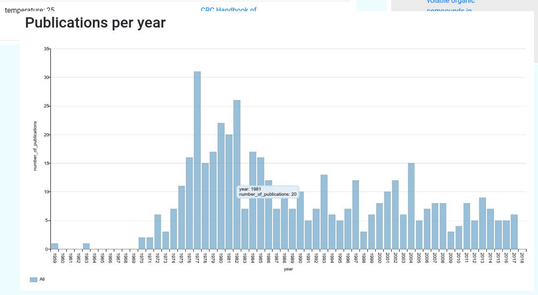
IDpred	IDpredLabel	count
wd:P235	InChIKey	152393
wd:P233	canonical SMILES	152233
wd:P234	InChI	149944
wd:P662	PubChem CID	145798
wd:P661	ChemSpider ID	125510
wd:P2017	isomeric SMILES	84844
wd:P683	ChEBI ID	84011
wd:P231	CAS Registry Number	72475
wd:P652	UNII	59293
wd:P592	ChEMBL ID	49622
wd:P3117	DSSTOX substance identifier	36373
wd:P232	EC ID	20335
wd:P1579	Beilstein Registry Number	19083
wd:P665	KEGG ID	15065
wd:P2566	ECHA InfoCard ID	12362
wd:P715	Drugbank ID	7786
wd:P595	Guide to Pharmacology Ligand ID	5950
wd:P2057	HMDB ID	5705
wd:P2064	KNApSAcK ID	4272

Identifier mappings are made available via BridgeDb.

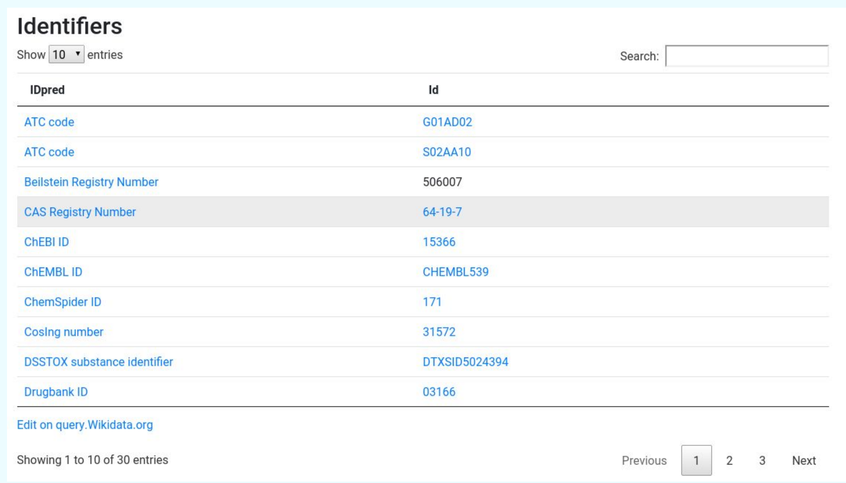
Literature-backed (PhysChem) Facts







Linking Databases



References

[1] Enabling Open Science: Wikidata for Research (Wiki4R), Research Ideas and Outcomes, 1, **2015**, doi:10.3897/RIO.1.E7573 [2] WikiGenomes: an open Web application for community consumption and curation of gene annotation data in Wikidata, Database, 2017:1, **2017**, doi:10.1101/102046 [3] The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, Journal of Cheminformatics, 9(1), **2017**, doi:10.1186/S13321-017-0247-6 [4] SPLASH, a hashed identifier for mass spectra, Nature Biotechnology, 34(11), **2016**, doi:10.1038/NBT.3689 [5] Scholia, Scientometrics and Wikidata, The Semantic Web: ESWC 2017 Satellite Events, **2017**, doi:10.1007/978-3-319-70407-4_36 [6]WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research", Nucleic Acids Research, 46(D1), **2018**, doi:10.1093/NAR/GKX1064

Funding Scholia has received funding from the Alfred P. Sloan Foundation under grant number G-2019-11458.

Correspondence to:
Dr Egon Willighagen
ORCID: 0000-0001-7542-0286
@egonwillighagen

Dept of Bioinformatics - BiGCaT, NUTRIM, Maastricht University
http://www.bigcat.unimaas.nl/
@bigcat_UM

WIKIDATA