

Beilstein journals contain Bioschemas

Egon Willighagen 

Published February 13, 2025

Citation

Willighagen, E. (2025). Beilstein journals contain Bioschemas. In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/yyjnz-n5j48>

Keywords

Bioschemas, Rdf, Chemistry, Beilstein

Abstract

Two weeks ago, the Beilstein Institute announced Bioschemas support in their journals:

Copyright

Copyright © Egon Willighagen 2025. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Two weeks ago, the [Beilstein Institute announced Bioschemas support in their journals](#):

We streamline the discoverability of your research by incorporating machine-readable chemical information into many of our published articles. This includes the conversion of chemical structures from submitted ChemDraw files to InChI strings and validating them using open-source tools.

The idea is far from new and has been around for two decades. But the [two Beilstein journals](#) (both [diamond Open Access](#)), actually integrated into their active publishing model. That has been trialed and put in action before. For example, there was (is?) [Project Prospect](#) (2007), [chemical structure annotation in Nature Chemistry](#) (2009), [SMILES in the ACS Journal of Medicinal Chemistry](#) (2014) (doi:[10.1021/jm5002056](#)), and [FAIR chemical structures in the Journal of Cheminformatics](#) (2021) (doi:[10.1186/s13321-021-00520-4](#)).

But this announcement is a new step. I like how validation of the chemical structures is part of the approach, and I like how they use the [Bioschemas](#) extension of [schema.org](#). The last because they use two Bioschemas types/profiles that contributed to or initiated, respectively: [MolecularEntity](#) and [ChemicalSubstance](#).

First stop for me is to check the schema.org annotation with a validation tool, like [Google's Rich Results Test](#). That gives an idea how they may have their search engine pick it up. The test article I was given on LinkedIn is [Xiao et al.'s Molecular diversity of the reactions of MBH carbonates of isatins and various nucleophiles](#) (doi:[10.3762/bjoc.21.21](#)) in the [Beilstein Journal of Organic Chemistry](#), and we indeed [see the schema.org annotation show up](#):

id	https://www.beilstein-journals.org/bjoc/articles/21/1860-5397-21-21-CAFIBKBZWJFZCW-FXBPSFAMSA-N
about	
hasBioChemEntityPart	
smiles	<chem>C1=CC=C(C=C1)CN2C3=C(C=CC=C3)/C(=C(\C#N)/CNC4=C(C=CC=C4)Cl)/C2=O</chem>
inChI	InChI=1S/C24H18CIN3O/c25-19-9-6-10-20(13-19)27-15-18(14-26)23-21-11-4-5-12-22(21)28(24(23)29)16-17-7-2-1-3-8-17/h1-13,27H,15-16H2/b23-18-
type	MolecularEntity
molecularFormula	C24H18CIN3O
inChIKey	FOQOVOLYYARWPA-NKFKGCMQSA-N
type	ChemicalSubstance
id	https://www.beilstein-journals.org/bjoc/articles/21/1860-5397-21-21-FOQOVOLYYARWPA-NKFKGCMQSA-N
about	
hasBioChemEntityPart	

chem-bla-ics

And because of the use of open standards, extracting the information is not so hard with, for example here, Bacting ([doi:10.21105/joss.02558](https://doi.org/10.21105/joss.02558)), based on a 2022 script from the NanoSafety Cluster projects NanoCommons and SbD4Nano:

```
@Grab(group='io.github.egonw.bacting', module='managers-rdf', version='1.0.4')
@Grab(group='io.github.egonw.bacting', module='managers-ui', version='1.0.4')
@Grab(group='io.github.egonw.bacting', module='net.bioclipse.managers.jsoup', version='1.0.4')

bioclipse = new net.bioclipse.managers.BioclipseManager(".");
rdf = new net.bioclipse.managers.RDFManager(".");
jsoup = new net.bioclipse.managers.JSoupManager(".");

articles = [
    args[0]
]

kg = rdf.createInMemoryStore()

for (article in articles) {
    htmlContent = bioclipse.download(article)

    htmlDom = jsoup.parseString(htmlContent)

    // application/ld+json

    bioschemasSections = jsoup.select(htmlDom, "script[type='application/ld+json']");

    for (section in bioschemasSections) {
        bioschemasJSON = section.html()
        rdf.importFromString(kg, bioschemasJSON, "JSON-LD")
    }
}

turtle = rdf.asTurtle(kg);

println "#" + rdf.size(kg) + " triples detected in the JSON-LD"
// println turtle

sparql = """
PREFIX schema: <http://schema.org/>
SELECT ?entity ?inchikey ?smiles WHERE {
    ?entity a schema:MolecularEntity .
    OPTIONAL { ?entity schema:inChIKey ?inchikey }
}
```

chem-bla-ics

```
OPTIONAL { ?entity schema:smiles ?smiles }  
}
```

```
results = rdf.sparql(kg, sparql)
```

```
for (i=1;i<=results.rowCount;i++) {  
  println "${results.get(i, "inchikey")}\t${results.get(i, "smiles")}"  
}
```

The output is a simple table:

MGAPJMNPGGTFHJ - JEIPZWNWSA - N	CN1C(=O)/C(=C/2\C3=CC(=CC=C3N(CC4=CC=CC=C4)C2=O)C1)/C(=P
XEWMQVUVGAHESA - UHFFFAOYSA - N	CC1=CC=C(C=C1)NC2=C(C3C4=CC(=CC=C4N(CC5=CC=CC=C5)C3=O)C)C
UVTJORFYHPGJDZ - PYCFMQQDSA - N	CCCCN1C2=CC=C(C)C=C2/C(=C(\C#N)/CNC3=CC=C(C)C=C3)/C1=O
ILWGDUYVQRAMMG - PGMHBOJBSA - N	CCCCN1C2=CC=C(C)C=C2/C(=C(\C#N)/CNC3=CC=C(C=C3)C1)/C1=O
CAFIBKBZWFZCW - FXBPSFAMSA - N	CCCCN1C2=CC=C(C)C=C2/C(=C(\C#N)/CNC3=CC=CC=C3)/C1=O
UOJSFLANMVIMBV - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NC4=CC=C(C=C4)C1
VNJBGTGXAGHCSO - OAPYJULQSA - N	COC(=O)/C(=C\1/C2=C(C=CC=C2)N(CC3=CC=CC=C3)C1=O)/C=P(C4=C
KJXQRAKSOANQTJ - GFMRDNFCSA - N	CC1=CC=C(C=C1)NC/C(=C\2/C3=C(C=CC=C3)N(CC4=CC=CC=C4)C2=O)
IGEBJMZDOPBFGF - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NC4=CC=CC=C4)C1=
SSANVPNESOMKOM - AWQADKOQSA - N	C1=CC=C(C=C1)CN2C3=CC=C(C=C3/C(=C(/C#N)\C=P(C4=CC=CC=C4)
GEHWHSHQSIQZKL - NVQSTNCTSA - N	CCCCN1C2=CC=C(C=C2/C(=C\3/C(=P(C4=CC=CC=C4)(C5=CC=CC=C5)C
PALRSQOHFLRWDH - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NC4=CC=C(C=C4)OC
KBFODZMDS AFLFR - UHFFFAOYSA - N	CN1C(=O)C(=C(C1=O)NC2=CC(=CC=C2)C1)C3C4=CC(=CC=C4N(CC5=CC
JCGAVVZYXDJPBU - GFMRDNFCSA - N	CC1=C(C=CC=C1)NC/C(=C\2/C3=C(C=CC=C3)N(CC4=CC=CC=C4)C2=O)
DZFPDPCDEQGLPLY - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NC4=CC=C(C)C=C4)
XMRNJCUJYOXYXJU - DAFNUICNSA - N	CC1=CC=C(C=C1)NC/C(=C\2/C3=CC(=CC=C3N(CC4=CC=CC=C4)C2=O)C
SSDSNBBHEUUKGI - UHFFFAOYSA - N	CC1=CC=C2C(=C1)C(C3=C(C(=O)N(C)C3=O)N(C)C4=CC=CC=C4)C(=O)
USFYPRDMNXMWPO - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NC4=CC=C(C=C4)Br
XYHTWFULRHTEAG - MUGXBBEHS A - N	CCCCN1C2=CC=C(C)C=C2/C(=C(/C#N)\C=P(C3=CC=CC=C3)(C4=CC=CC
XALDZIBHNNIVAM - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NC4=C(C=CC=C4)O)
TUTWQHBRQPMLME - OAPYJULQSA - N	COC(=O)/C(=C\1/C2=CC(=CC=C2N(CC3=CC=CC=C3)C1=O)C1)/C=P(C4
IYEHFTMZZMIPRU - UHFFFAOYSA - N	CC1=CC=C(C=C1)NC2=C(C3C4=CC(=CC=C4N(CC5=CC=CC=C5)C3=O)C1)
KBSDGNPLIPXCEX - UHFFFAOYSA - N	CCCCN1C2=CC=C(C)C=C2C(C3=C(C(=O)N(C)C3=O)NCC4=CC=CC=C4)C1
BQGIUMITIGHBSD - UHFFFAOYSA - N	CCCCNC1=C(C2C3=CC(=CC=C3N(CC4=CC=CC=C4)C2=O)C)C(=O)N(C)C1
PNSOLOPHIVUPOZ - MNPDQUGUSA - N	CCCCNC/C(=C\1/C2=CC(=CC=C2N(CCCC)C1=O)C)/C#N
HLTBKJRJOIZCMJ - PYCFMQQDSA - N	CCCCN1C2=CC=C(C)C=C2/C(=C(\C#N)/CN(C)C3=CC=CC=C3)/C1=O
FFLHFLUBMRBQTB - UHFFFAOYSA - N	CCCCN1C2=CC=C(C=C2C(C3=C(C(=O)N(C)C3=O)NC4=CC=C(C)C=C4)C1
FOQOVOLYYARWPA - NKFKGCMQSA - N	C1=CC=C(C=C1)CN2C3=C(C=CC=C3)/C(=C(\C#N)/CNC4=CC(=CC=C4)C
KLEPCAQFOXJLNV - UHFFFAOYSA - N	CC1=C(C=CC=C1)NC2=C(C3C4=CC(=CC=C4N(CC5=CC=CC=C5)C3=O)C1)

That also made me realize that there are not chemical names in the annotation. That would be really useful to move things forward. Then again, PubChem will likely just generate the IUPAC name, since they have access to such software anyway. They have teamed up with PubChem

chem-bla-ics

which will index it, but I will be interested in seeing how to use this for `main subject` annotation in [Wikidata](#).

A final note for now, the model they use is annotate the article with chemical substances (`ChemicalSubstance`) with (one or more?) molecular entities (`MolecularEntity`). That is a model that scales well to their other journal, the [Beilstein Journal of Nanotechnology](#). But scraping that is for another post.