



VIA  
PANORAMICA

Revista de Estudios Anglo-Americanos  
A Journal of Anglo-American Studies

# TOWARDS THE MAELSTROM THE DIGITAL HUMANITIES IN ACTION

1ST CETAPS MEETING ON  
DIGITAL HUMANITIES  
DOSSIER

THEMATIC SECTION

EXTRA



# An Approach to Linked Open Data: From Wikidata to authority files data

NILS GEIBLER

Cologne Center for eHumanities

---

**ABSTRACT:** The workshop consisted of three main parts, starting with a brief introduction to Linked Open Data by recapitulating the main stages of its development and explaining some core principles and technologies. The second part focused on the approach used by the FID Philosophie<sup>i</sup> to combine Wikidata and authority files. Finally, the third part was done hands-on with a Jupyter notebook in Google Colab, going through the aforementioned approach step by step using Python.

**KEYWORDS:** Linked Open Data, FID Philosophie, Wikidata, Python

**RESUMO:** O workshop consistiu em três partes principais, começando com uma breve introdução ao Linked Open Data, recapitulando as principais etapas do seu desenvolvimento e explicando alguns princípios e tecnologias fundamentais. A segunda parte concentrou-se na abordagem usada por FID Philosophie para combinar Wikidata e authority files. Por fim, a terceira parte foi dedicada a aplicações práticas da abordagem teórica indicada com recurso ao Jupyter notebook no Google Colab e a Python.

**PALAVRAS-CHAVE:** Linked Open Data, FID Philosophie, Wikidata, Python

## A BRIEF HISTORY AND INTRODUCTION TO LINKED OPEN DATA (LOD)

The World Wide Web (WWW) was born in 1991, linking (web) sites all over the world.<sup>2</sup> Shortly after, the first vision of a more “meaningful” (i.e., add machine-readable descriptions) web emerged and was written down (1994-2000): the Semantic Web.<sup>3</sup> In 2006 and 2010, the principles of Linked Data (LD) and Linked Open Data (LOD) were stated,<sup>4</sup> which would perceive resources not only as websites but emphasise a more abstract view, also encouraging open publishing. With the Google Knowledge Graph, finally a practical and economically successful example entered the stage in 2012.<sup>5</sup> This seems to have given way to thinking of the Semantic Web not as “one big thing” but as something that could be handled and created in portions, independently, and that could be reconnected at will – by applying the principles and technologies of LOD. One of the core technologies of LOD is ontologies, which provide classes and properties for connecting entities within such graphs with respect to a certain domain or field and also may provide means to point to similar or same entities in other graphs.<sup>6</sup>

## THE *FID PHILOSOPHIE'S* APPROACH TO LINKED OPEN DATA

The *Fachinformationsdienst Philosophie* (FID, Engl. Specialised Information Service) is a cooperative (infrastructure) project of the University and City Library of Cologne (UCL) and the Cologne Center for eHumanities (CCeH). It provides researchers (from the discipline of philosophy) with several services to facilitate their research. Among those services is the PhilFinder.<sup>7</sup> It acts as a central hub to connect the different services of the FID and provide a quick entry into literature research. This is accomplished by creating websites depending on profiles, e.g. for persons, and by combining biographical data and available resources from the FID’s specialised library as well as the broader catalogue of the library network in one place. As the FID itself does not have the resources and means to store data or build a database on its own, the idea was to build upon existing open resources. This resulted in using Wikidata as

a hub to get basic data regarding as many entities as possible and become able to easily add new entities.<sup>8</sup> Wikidata also provides specific identifier properties that link to representations of the same entities in other resources.<sup>9</sup> One of these resources is the Gemeinsame Normdatei (GND, Engl. Integrated Authority File)<sup>10</sup> of the Deutsche Nationalbibliothek (DNB, Engl. German National Library), which is used as an authority file in the German-speaking area – thus, GND identifiers are reused in several databases and projects. It also provides authority data regarding works and persons (mainly German-centric). So, one may perceive the GND as a more specialised knowledge base than Wikidata, which may provide us with less but better data. The FID combines both knowledge bases in a way by prioritising the GND over Wikidata for overlapping data, i.e. data fields that are present in both, and using Wikidata only as a fallback. This approach could also be used for combining data from other knowledge bases. The Python implementation presented in the workshop already includes the Virtual International Authority File (VIAF)<sup>11</sup> and the Library of Congress Name Authority File (LCNAF)<sup>12</sup> as an approach to include international authority files. Thus, creating different profiles with different prioritisation of resources may be used to set different accentuations for cultural and language areas – in contrast to prioritising the GND for the German-speaking sphere.

For clarification: The FID does not request raw GND data, but uses the lobid-GND-API<sup>13</sup> to get already refined data via a more accessible format.

## THE HANDS-ON PART OF THE WORKSHOP<sup>14</sup>

A recent presentation of the FID's approach in the network of FIDs brought up the idea of implementing PhilFinder's core component ("Nexus") in Python. This would facilitate prototyping and enable a more comprehensive and focused point of reference for sharing and discussion. The workshop has been an ideal opportunity to put this into action: by implementing it in Python as well as in presenting the approach. An established way of presenting Python code is the use of Jupyter notebooks with combined cells of code and

explanatory, documentary notes. The notebook is divided into four main parts: a setup workflow that is used to install the dependencies and hide more complex code, a first workflow that explains the basic components and suggests modifications (i.e., setting and altering queried triples, used properties, resource priorities), a second workflow that uses an implementation of the so-called “Nexus” class (which implements a prioritised aggregation of data sources with Wikidata as a data hub), and a third workflow that only queries Wikidata without using another resource. All workflows provide a visualisation (timeline for names and dates) and a download of the obtained data in the end and start with querying Wikidata via SPARQL<sup>15</sup> to get a specified subset.

The workshop concluded with the participants formulating their own queries to Wikidata based on their research interests. They learnt to modify the code and queries accordingly and to search for the appropriate Wikidata classes and properties.

## END NOTES

<sup>1</sup> <https://philportal.de>.

<sup>2</sup> Cf. <https://www.britannica.com/topic/Semantic-Web> and Berners-Lee and Fischetti, Weaving the Web: the WWW was first conceptualised in 1989.

<sup>3</sup> Ibid, see also Hitzler, ‘A Review of the Semantic Web Field’.

<sup>4</sup> Cf. <https://www.w3.org/DesignIssues/LinkedData.html>.

<sup>5</sup> Cf. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.

<sup>6</sup> Noy, ‘Ontology Development 101’.

<sup>7</sup> [philportal.de/philfinder](https://philportal.de/philfinder).

<sup>8</sup> Cf. a comprehensive blog post by Joachim Neubert from 2017, which documents a workflow on reconciliation with other sources and enrichment of Wikidata: <https://zbw.eu/labs/en/blog/wikidata-as-authority-linking-hub-connecting-repec-and-gnd-researcher-identifiers.html>.

<sup>9</sup> Cf. a list of all external identifiers: <https://www.wikidata.org/wiki/Special:ListProperties/external-id>.

<sup>10</sup> <https://www.dnb.de/EN/gnd>.

<sup>11</sup> <https://viaf.org/>.

<sup>12</sup> <https://id.loc.gov/authorities/names.html>.

<sup>13</sup> <https://lobid.org/gnd/api>.

<sup>14</sup> See <https://github.com/pbd84/LOD-workshop-20241029> for both the slides and the Jupyter notebook.

<sup>15</sup> A query language for graph data, cf. <https://www.w3.org/TR/rdf-sparql-query/> and <https://www.w3.org/TR/sparql11-overview/>.

---

## SOURCES

(All web sources were last accessed on 25.11.2024 and archived via <https://web.archive.org/>.)

- Berners-Lee, Tim. 'Is Your Linked Open Data 5 Star?', 2010. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, Tim. 'Linked Data', 2006. <https://www.w3.org/DesignIssues/LinkedData.html>.
- German National Library. 'The Integrated Authority File (GND)'. Accessed 25 November 2024. <https://www.dnb.de/EN/gnd>.
- Google. 'Introducing the Knowledge Graph: Things, Not Strings', 16 May 2012. <https://blog.google/products/search/introducing-knowledge-graph-things-not/>.
- hbz. 'Lobid-Gnd'. Accessed 25 November 2024. <https://lobid.org/gnd/api>.
- Hitzler, Pascal. 'A Review of the Semantic Web Field'. *Communications of the ACM* 64, no. 2 (25 January 2021): 76–83. <https://doi.org/10.1145/3397512>.
- Hosch, W. L. 'Semantic Web'. In *Encyclopedia Britannica*. Accessed 25 November 2024. <https://www.britannica.com/topic/Semantic-Web>.
- Library of Congress. 'Library of Congress Name Authority File'. Accessed 25 November 2024. <https://id.loc.gov/authorities/names.html>.
- Neubert, Joachim. 'Wikidata as Authority Linking Hub: Connecting RePEc and GND Researcher Identifiers'. *ZBW Labs* (blog), 30 November 2017. <https://zbw.eu/labs/en/blog/wikidata-as-authority-linking-hub-connecting-repec-and-gnd-researcher-identifiers.html>.
- Noy, Natasha. 'Ontology Development 101: A Guide to Creating Your First Ontology', 2001. [https://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](https://protege.stanford.edu/publications/ontology_development/ontology101.pdf).
- OCLC. 'Virtual International Authority File'. Accessed 25 November 2024. <https://viaf.org/>.

## NILS GEIBLER

Nils Geißler studied information processing and philosophy with a BA at the University of Cologne (currently: MA information processing - with participation in the a.r.t.e.s. Research Master program) and has been working as a student/research assistant at the CCeH since 2014 in various projects (mainly: Capitularia, Modeling between Digital and Humanities: Thinking in Practice, NFDI: Text+).

---

## HOW TO CITE

Geißler, Nils (2024). "Workshop 1: An Approach to Linked Open Data: From Wikidata to Authority Files Data". *VIA PANORAMICA: Revista de Estudos Anglo-Americanos*, vol. 13, n.º 2, 2024, pp. 93-97. Web: <http://ojs.letras.up.pt/>. DOI: 10.21747/2182-9934/via13\_2w1