

# FAIR in the SSH scholarly communication breakout session



Erzsébet Tóth-Czifra  
Open Science Fair,  
Porto, 17.09.2019

# Welcome!

## Session outline

Who we are and how each of us can  
contribute to a FAIR scholarly  
communication ecosystem?



Who is missing?

What does FAIR mean in the context of scholarly communication in the SSH?

Expectations vs. challenges



# Case studies



# Persona 1

*I am a post-doctoral researcher working in the field of spatial humanities. My biggest struggle with publication is that in many cases I feel like I have to do double work: I build exciting databases and interactive maps on the top of them but these does not 'count' as publications so I also need to write papers about them. However, due to the long delay in publication, my works become outdated by the time of getting published.*

## Persona 2

*We are running a small Open Access journal dedicated to the history of science in Hungary. We want to make our journal more attractive for scholars, but the problem is, we are not indexed by Google Scholar (let alone Web of Science or Scopus) and we feel that our work remains basically invisible in the digital horizon and also in terms of research evaluation.*

# Persona 3

*We recently launched a new preprint service dedicated to cultural studies. How could we strengthen our user base?*

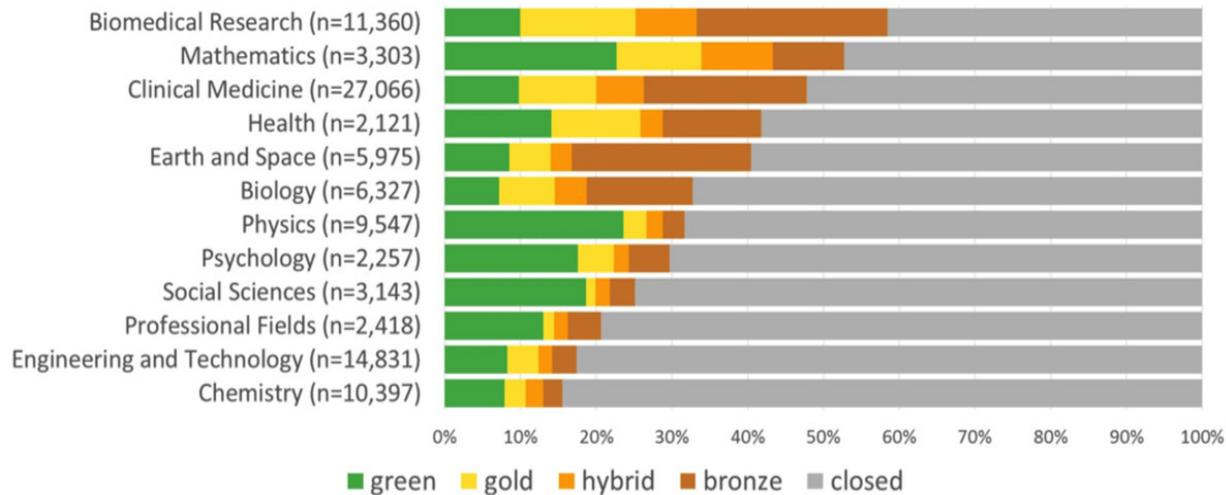
# Persona 4

*I am working as a subject librarian in arts and humanities and recently, a historian scholar contacted me with the following problem: she selected a high-prestige OA journal to publish their project results but it charges 1500 CHF publication fee which is more than her annual budget for such expenses. How can I help her?*

A real use case

Some of the issues we  
are facing

# Yes, we have a visibility problem



## FUNDER EFFECT

Even within the same discipline, access varies greatly by funder. Of chemistry papers supported by the NIH, 81% were open access\*; 24% of NSF's chemistry papers were.

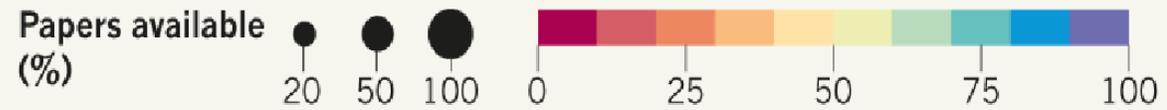


Figure 4: Percentage of different access types of a random sample of WoS articles and reviews with a DOI published between 2009 and 2015 per NSF discipline (excluding Arts and Humanities).

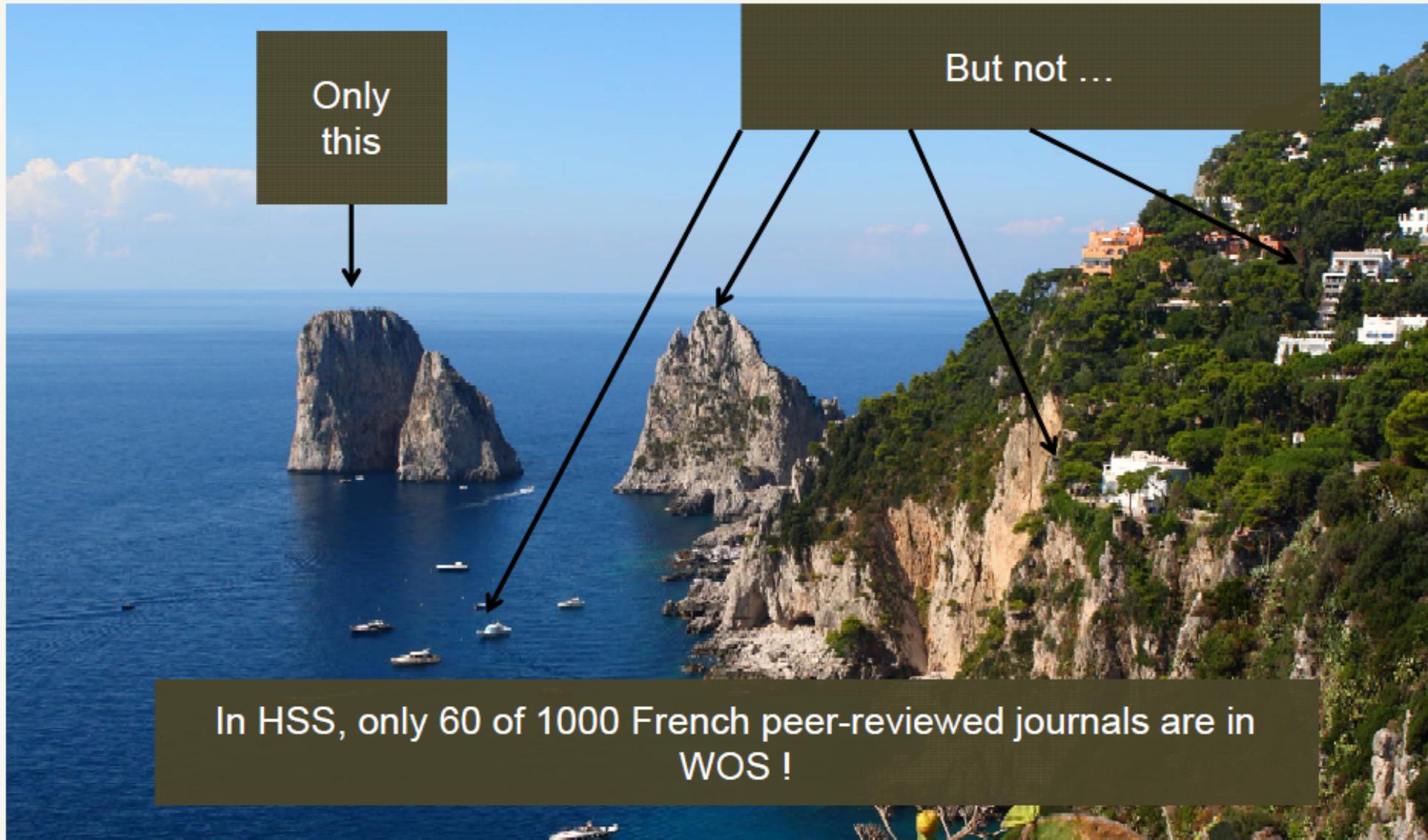
(Piwowar et al. 2018) <https://doi.org/10.7717/peerj.4375>

Larivière & Sugimoto (2018)



<https://www.nature.com/articles/d41586-018-07101-w>

In fact, the WOS is an island ignoring the majority of journals



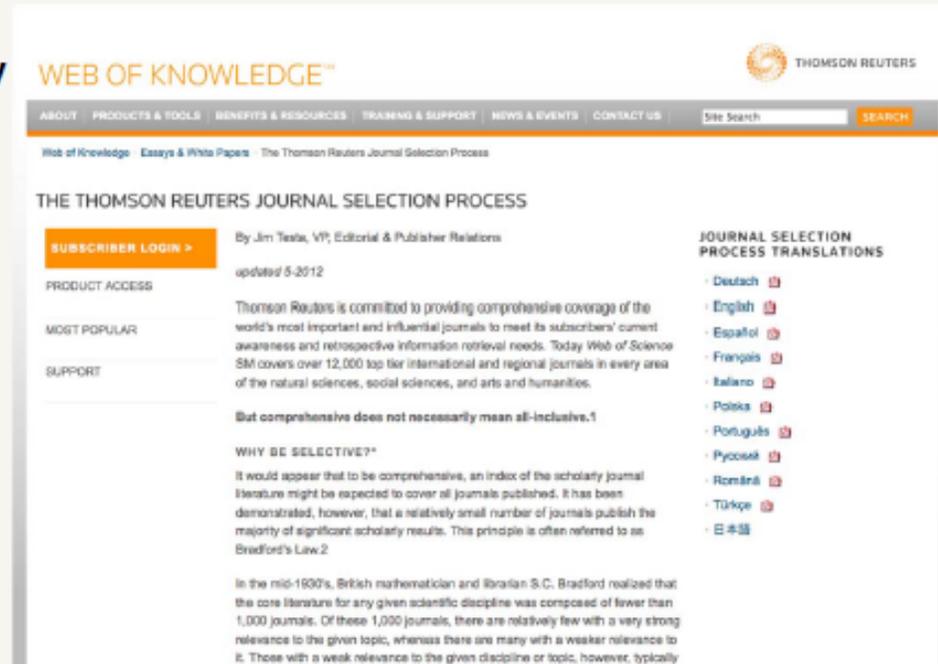
Source: Marin Dacos:  
Digital Humanities in  
Open Access.  
[https://openaccess.  
mpg.de/1528691/Ses  
sion-3-Dacos.pdf](https://openaccess.mpg.de/1528691/Session-3-Dacos.pdf)

# The advent of English-only ?

Most of the time, Humanities and Social sciences write in their native language, not primarily in English.

The WOS seems not to worry about that :

**“English is the universal language of science. For this reason Thomson Reuters focuses on journals that publish full text in English, or at very least, bibliographic information in English.”**



The screenshot shows the Thomson Reuters Web of Knowledge website. The header includes the logo and navigation links: ABOUT, PRODUCTS & TOOLS, BENEFITS & RESOURCES, TRAINING & SUPPORT, NEWS & EVENTS, CONTACT US, and a search bar. The main content area is titled "THE THOMSON REUTERS JOURNAL SELECTION PROCESS" and is written by Jim Teala, VP, Editorial & Publisher Relations, updated 5-2012. The text explains that Thomson Reuters is committed to providing comprehensive coverage of the world's most important and influential journals. It notes that today Web of Science SM covers over 12,000 top tier international and regional journals in every area of the natural sciences, social sciences, and arts and humanities. A key point is that "But comprehensive does not necessarily mean all-inclusive." The page also includes a "WHY BE SELECTIVE?" section, which states that it would appear that to be comprehensive, an index of the scholarly journal literature might be expected to cover all journals published. However, it has been demonstrated that a relatively small number of journals publish the majority of significant scholarly results. This principle is often referred to as Bradford's Law. A sidebar on the right lists "JOURNAL SELECTION PROCESS TRANSLATIONS" in various languages: Deutsch, English, Español, Français, Italiano, Polska, Português, Pycckè, Romàni, Türkçe, and 日本語.

Source: Marin Dacos:  
Digital Humanities in  
Open Access.  
<https://openaccess.mpg.de/1528691/Session-3-Dacos.pdf>

# Diversity not only in terms of languages but also:

- Publication cultures and research cultures

(what do we value when choosing a publication venue and how do we align them with our own research processes and data?)

- Disciplines

(different needs, traditions and practices of community)

- Content types and formats

(books! critical editions! Dynamic formats, living scholarship!How to go beyond legacy formats?)

- Diversity of players

(fragmented publication landscape, many small players grounded in language and disciplinary communities)





# Jussieu Call for Open science and bibliodiversity

<https://jussieucall.org/jussieu-call/>

# Self-archiving in the humanities: myths, fears, changing practices...and of course, infrastructure!



Mason Youngblood  
@MasYoungblood

A preprint of my new manuscript on conformity bias in the cultural transmission of music sampling traditions is up on [@arxiv](#) 🌐 Would love feedback/ideas from the community! [arxiv.org/abs/1906.11928](https://arxiv.org/abs/1906.11928)  
[#culturalevolution](#) [#musicology](#) [#machinelearning](#)



Lisa DeBruine 🇺🇸  
@LisaDeBruine

I used to tell people as an editor to submit data and materials as SI (as opposed to personal websites) to make sure they stayed available and linked to the paper. Now I tell them to share through [@figshare](#) or [@OSFramework](#) [twitter.com/vscooper/status...](https://twitter.com/vscooper/status/988888888888888888)

Vaughn Cooper @vscooper

Why do journals take so little care of Supplementary Files? Was just alerted that links to SI in our recent paper are now broken w/ website update. Not the first time for me either. Concerned that most important elements of pubs are being lost at an alarming rate. Need ombudsman!



Lisa Matthias  
@l\_matthia

Does anyone have any "good" (comprehensive, detailed) examples of pre-reg/pre-analysis plans for qualitative studies? The [@OSFramework](#) has some templates, but it would be great to see completed ones.  
[#openscience](#) [#phdchat](#)

Humanities Commons @humcommons · Apr 19, 2018

By creating a profile on [@humcommons](#), you strengthen your online presence and share your interests alongside your accomplishments, work, and education. Today's featured profile belongs to Zanne Domoney-Lyttle. Check it out!: [ow.ly/iHGn30jvSb3](https://ow.ly/iHGn30jvSb3)



R.D. Ingthorsson @RIngthorss · Aug 29

I just uploaded a **preprint** of 'Mario Bunge and the Current Revival of Causal Realism' to [@academia](#)! Published in a festschrift celebrating Bunge's 100 birthday:



Mario Bunge and the Current Revival of Causal Realism  
Mario Bunge's Causality and Modern Science is arguably one of the best treatments of the causal ...  
[academia.edu](https://academia.edu)



Jeroen Sondervan  
@jeroenson

Preprint publishing is fairly new in the humanities, but this is changing rapidly. Want to know more about what it is? And how it can speed up scholarly communication? Read the FAQs (where we listed a few further readings about the topic): [mediarxiv.com/faqs/](https://mediarxiv.com/faqs/)



Jack Grieve @JWGrieve · Aug 28

Our paper on attributing the Bixby Letter to John Hay over Abraham Lincoln has now officially been published:

[doi.org/10.1093/llc/fq...](https://doi.org/10.1093/llc/fq...)

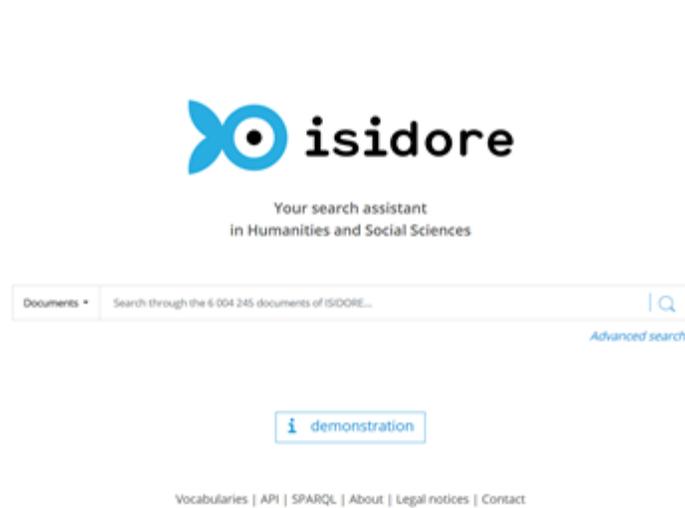
It's behind a paywall but the **preprint** is available on my **website** if you're interested.



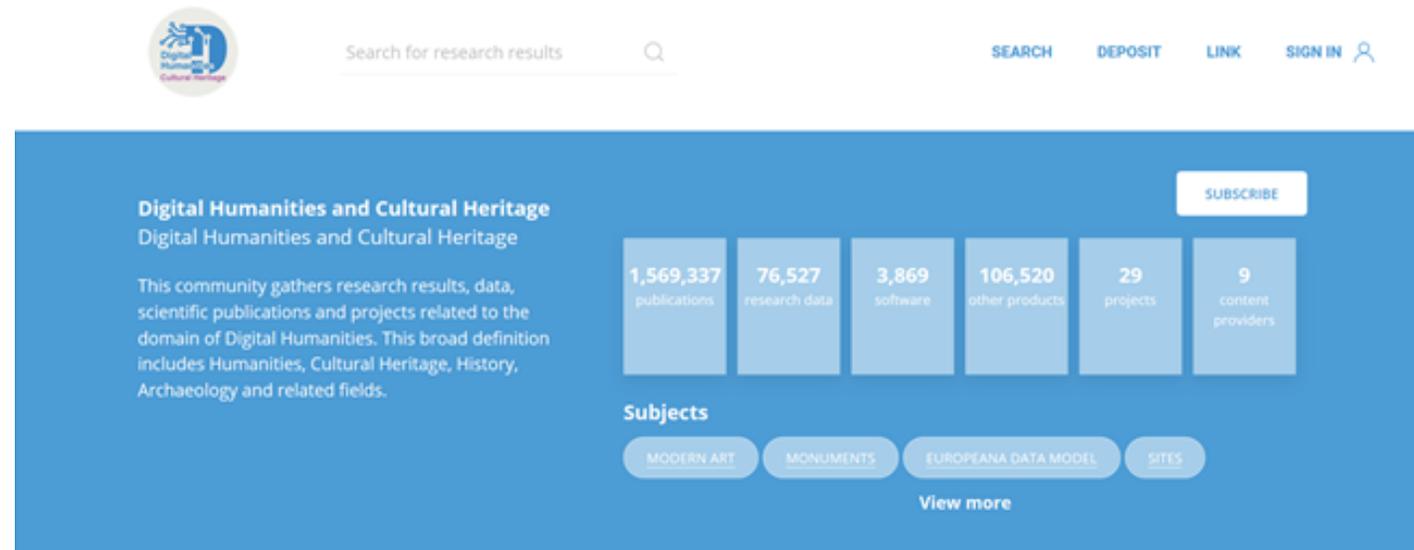
Attributing the Bixby Letter using n-gram tracing  
Abstract. There is a long-standing debate about the authorship of the Bixby Letter, one of the most famo...  
[academic.oup.com](https://academic.oup.com)

1 11 40

# Innovation in discovery platforms and practices



<https://isidore.science/>

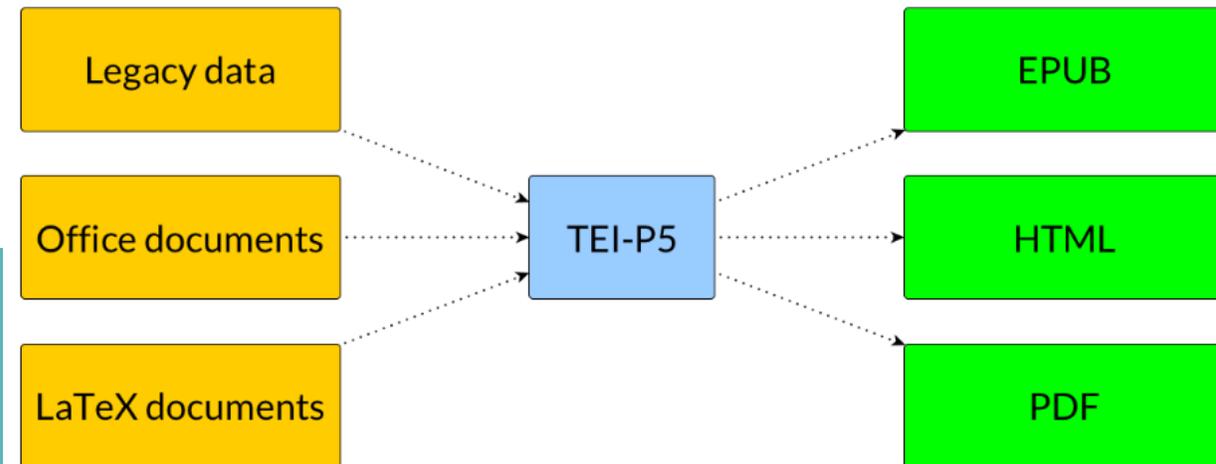


<https://beta.dh-ch.openaire.eu/>

# How about publications *as* data?

...because we do not only read publications but also enrich them, extracting information from them, analyze them etc.

Image [source](#)



An exemplary model of a sustainable workflow for scholarly communication. ([Thoden 2019](#))

# DH tools in the service of scholarly communication: the implementation of the NERD service in two different environments:

## entity-fishing Service Implementations on Publishing Platforms



### Named Entity Recognition and Disambiguation

Entity extraction and disambiguation is the task of determining the identity of entities mentioned in a text against a knowledge base. The identification and resolution of named entities like person-name, location, etc. provides many practical applications, e.g. possibility to extract lists of people, to map different texts, to generate timelines and to provide an enhanced search. This is of great importance not only for research but also for the publishing process.

### INRIA Entity-Fishing and its Advantages

- Entity-fishing, the NERD implementation developed by INRIA, is a service available within the DARIAH-EU infrastructure and used by the HIRMEOS partners to enrich Open Access digital monographs published on five digital platforms.
- No requirement for expertise in knowledge engineering
  - Multilingual support (currently it supports 5 languages: en, es, it, fr, de)
  - Standard REST API

Author	...
Title	...
Year	...
Language	...
Subject	...
Keywords	...
DOI	...
URL	...
Access	...
License	...
Version	...
Created	...
Updated	...
Published	...
Retrieved	...

### Typical Use Cases on Open Access Digital Platforms

- HIRMEOS partners have tested the integration of the NERD API in the workflow of publishing platforms to enhance discoverability and usage of enriched digital monographs.
- Enhancing the discoverability of monographs through facets linked with the concepts extracted
  - Aggregation of extracted entities, e.g., word clouds helping to understand the trending topics, author topic evolution, co-authorship or collaboration evolution, etc.
  - Classification or clustering to automatically generate collections
  - Highlighting annotated entities with different colours and showing pop-up definitions

### What can I do with Entity-Fishing?

- Entity extraction and resolution against Wikidata and Wikipedia in a row text or partially-annotated text segment
- Entity extraction and resolution in a PDF as an overlay on the original text
- Disambiguation of a search query
- Access to the Wikidata metamodel containing more than 35m entities and 154m statements
- Interactive disambiguation in text

### More Information

Do you want to know more about NERD services and the HIRMEOS implementations of the INRIA entity-fishing? Visit [www.hirmeos.eu](http://www.hirmeos.eu) and subscribe to our newsletter.



## On the HIRMEOS publishing platforms and on OpenMethods

### Know your implementation: Subgraphs in literary networks

Introduction by someone: The article is about subgraphs, etc...

We can visually distinguish three subgraphs: the main graph revolving around Faust and Mephisto, which basically comprises the entire plot of the play, except for two detached single scenes: Vorspiel auf dem Theater (Prelude in the Theater) and Walpurgisnachtstraum (Walpurgis Night's Dream). The two latter scenes do not feature any character from the main graph, which is

**NERD PLUGIN**

ALGORITHM: AVERAGE PATH LENGTH

**AVERAGE PATH LENGTH**

NORMALIZED: AVERAGE PATH LENGTH  
DOMAINS: COMPUTER\_SCIENCE, TELECOMMUNICATION  
CONF: 0

AVERAGE PATH LENGTH IS A CONCEPT IN NETWORK TOPOLOGY THAT IS DEFINED AS THE AVERAGE NUMBER OF STEPS ALONG THE SHORTEST PATHS FOR ALL POSSIBLE PAIRS OF NETWORK NODES. IT IS A MEASURE OF THE EFFICIENCY OF INFORMATION OR MASS TRANSPORT ON A NETWORK.

REFERENCES: WIKIPEDIA WIKIDATA

<https://dhmethods.hypotheses.org/182>

<https://www.hirmeos.eu/wp-content/uploads/2017/11/HI33-DARIAH-NERD-service-and-API-final.pdf>

# Thank you!



**Let's stay in touch!**

- Twitter: @etothczifra
- E-mail: [erzsebet.toth-czifra@dariah.eu](mailto:erzsebet.toth-czifra@dariah.eu)