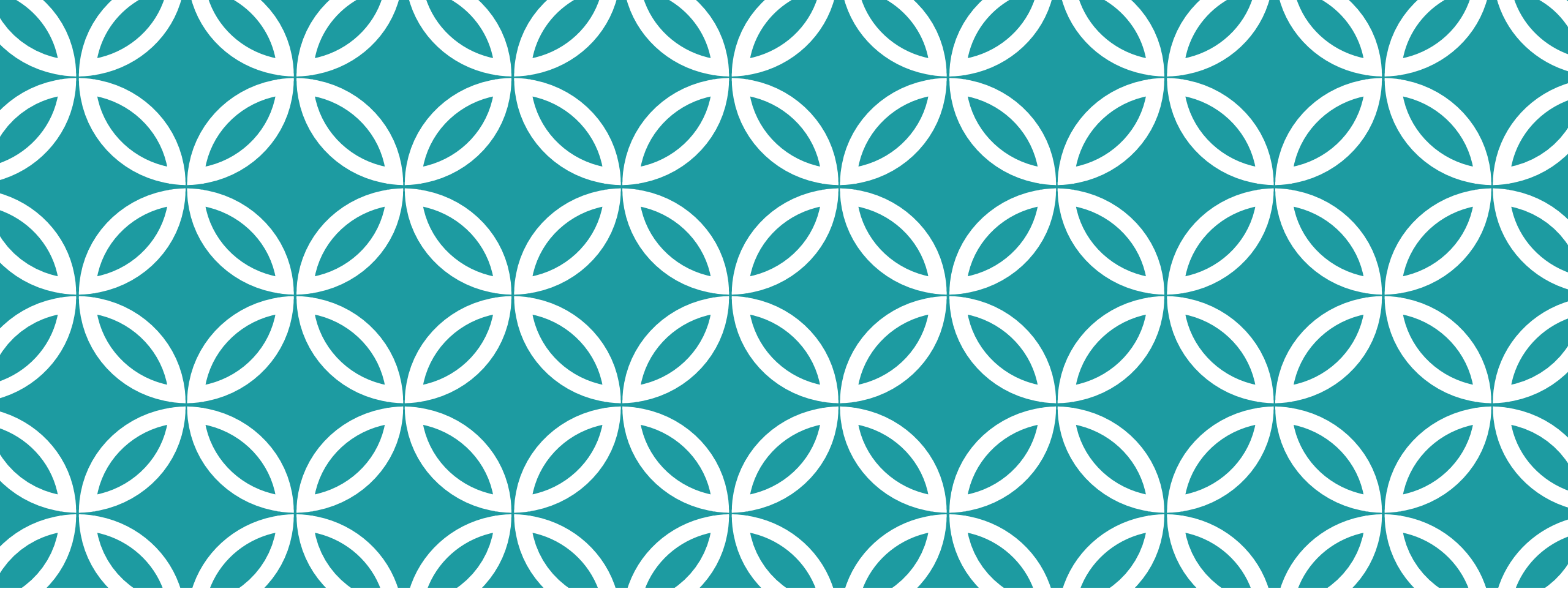# TURNING SCHOLIA INTO A PLATFORM FOR LIVING SCIENTOMETRIC STUDIES

Houcemeddine Turki

Faculty of Medicine of Sfax,
University of Sfax, Sfax, Tunisie

WikiCite 2020
26-28 October 2020, Online

INTRODUCTION

{} wikicite

SCIENTOMETRIC STUDY
DEFINITION

A research study that tries to analyze the quantitative, qualitative or behavioral evolution of scholarly research efforts:

- Research publications of an individual
- Research publications of the recipients of a scientific distinction
- Research publications on a topic of interest
- Research publications of a scientific institution
- Publications as part of a research project
- Research publications of a sovereign state
- Research publications during a specific period
- Coupling of two entities (e.g. Topic of interest in a sovereign state)

SCIENTOMETRIC STUDY
SECTIONS

Introduction

Methods

Results

Discussion

Conclusion

SCIENTOMETRIC STUDY
SECTIONS

Introduction :

- The analyzed entity in context

- List of scientometric studies on the entity

Methods

- Nature of the data used: Bibliographic metadata, Altmetric statistics, Scientific citations,

- Type of publications analyzed: Scholarly articles, Literature reviews, Invention patents, Legal texts ...

- Analysis restriction criteria: Data sources, Time span of analysis, Country, Keywords ...

- Data processing: Automatic natural language processing, Statistical analysis, Network analysis ...

SCIENTOMETRIC STUDY
SECTIONS

Results

- Raw scientometric indices and metrics (Number of citations, h-index)
- Standardized indicators with respect to a given criterion (Domain, Source)
- Relative indicators (Percentile rank according to a metric)
- Scientometric networks and derivation of centrality indicators (Citation, Co-citation, Collaboration, Bibliographic coupling)
- Statistical distribution and correlation

Discussion

- Comparison with the previous results of the same entity (Example: Tunisia between 2010 and 2015 vs. Tunisia between 2016 and 2020)
- Comparison with the results of other entities of the same class (Example: Tensorflow vs. PyTorch, SARS vs. COVID-19)
- Comparison with the results of a generalized entity (Example: Tunisia vs. Africa)

Conclusion

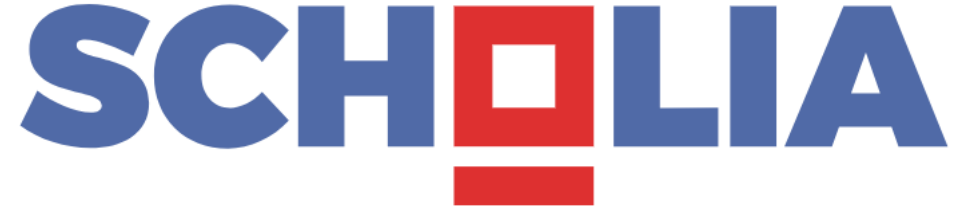- Identification of significant differences from the discussion
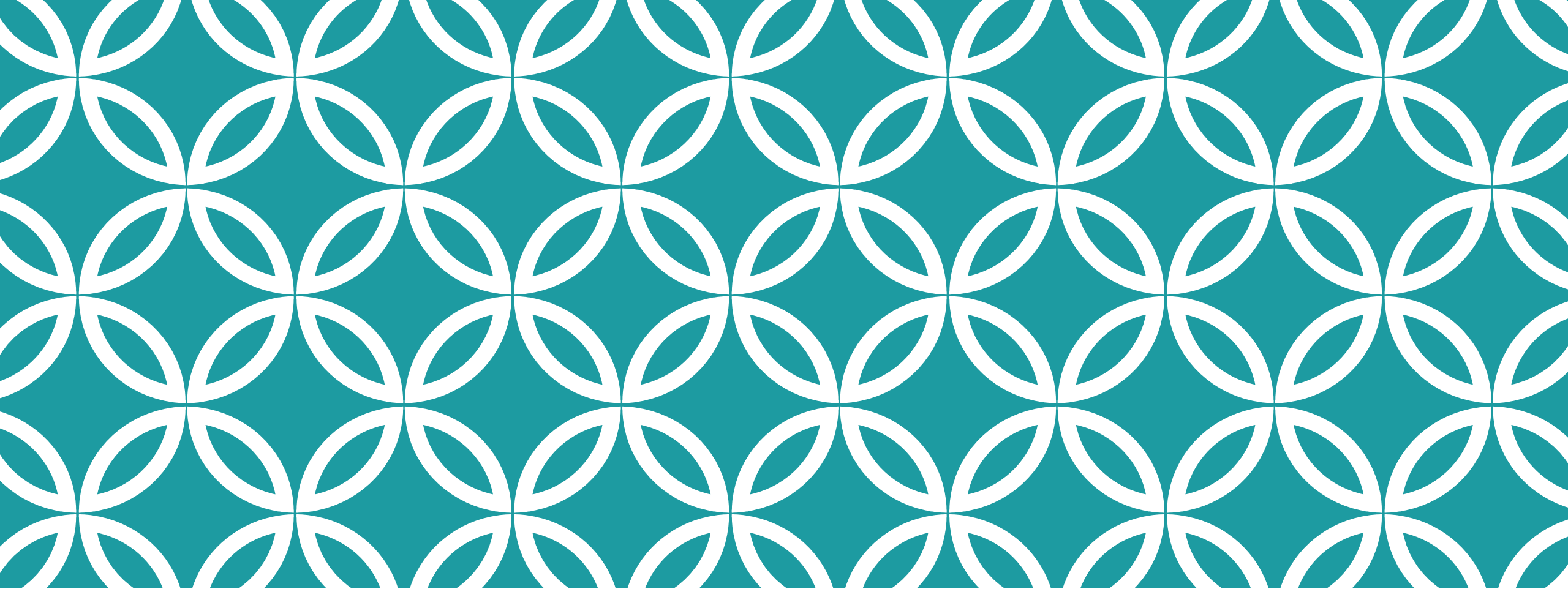
SCHOLIA : TOOL FOR SCHOLARLY PROFILING

{} wikicite

SCHOLIA : TOOL FOR SCHOLARLY PROFILING

HTTPS://SCHOLIA.TOOLFORGE.ORG

SPARQL queries can be used to query knowledge graphs and extract specific information in real time.

➔ Linked Bibliographic data in RDF format can be queried to generate knowledge that can be updated as long as the database is regularly enriched

A VARIETY OF VISUALISATIONS

Number of publications per year
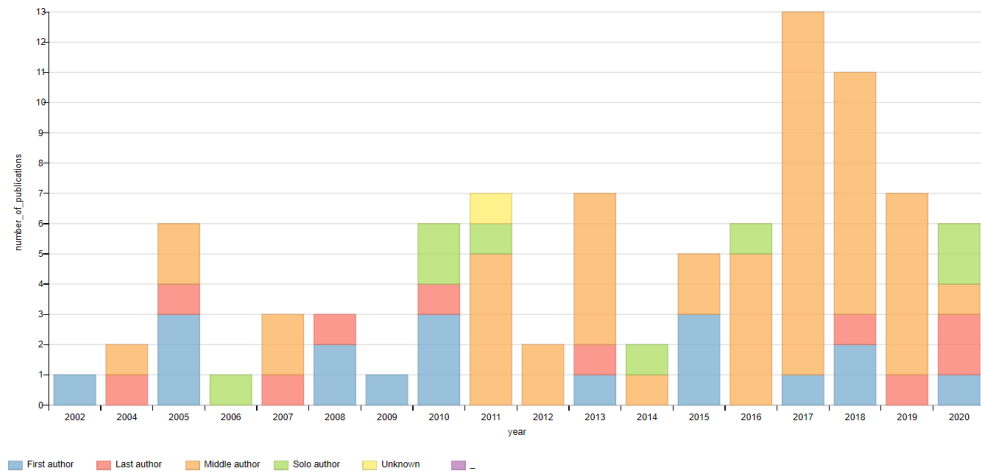
Co-author graph
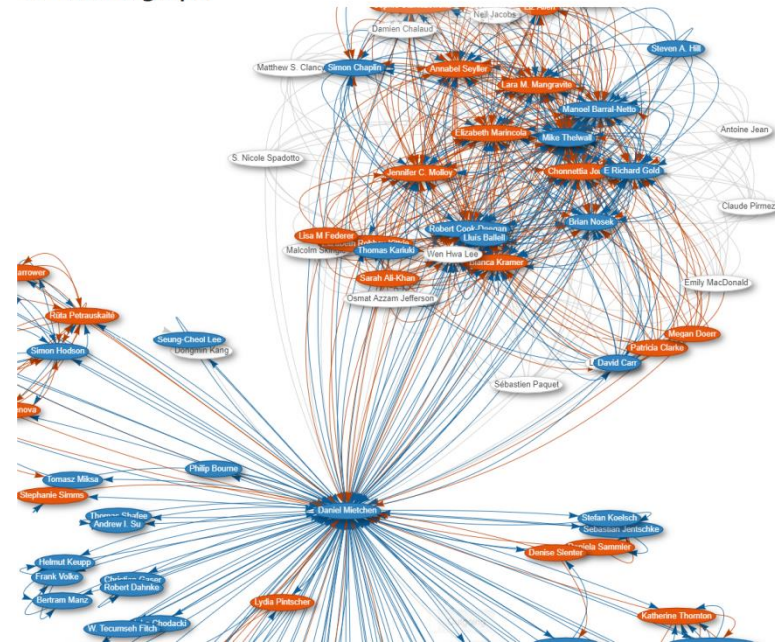
Timeline
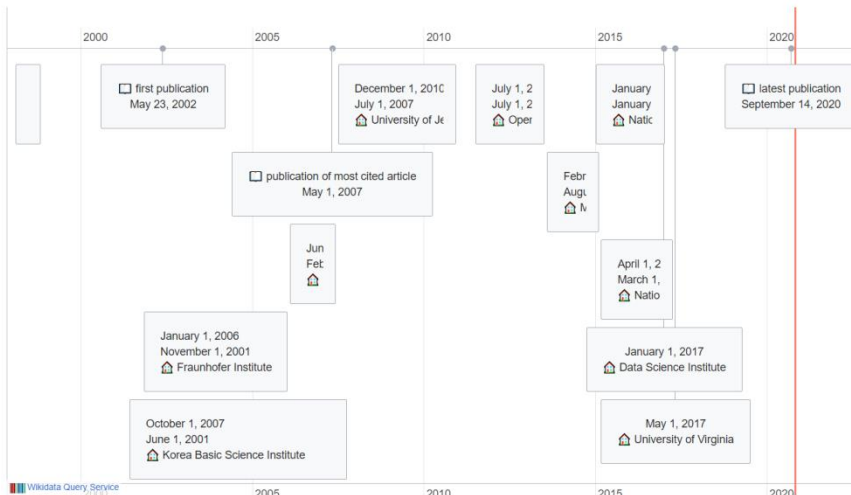
Academic tree

DEMO

Tunisia (Location, Country, Topic):
https://scholia.toolforge.org/topic/Q948

Daniel Mietchen (Author):
https://scholia.toolforge.org/author/Q20895785

University of Oxford (Publisher, Location, Sponsor, Organization, Topic):
https://scholia.toolforge.org/award/Q34433

Ruben Verborgh, James Hendler and Tim Berners-Lee (Authors):
https://scholia.toolforge.org/authors/Q80,Q6135847,Q30085536

LIMITATIONS

Raw visualization of data about the analyzed entity without contextualization or even discussion

Scientometric profiles are not personalized according to the user's needs (Example: Restricting the analysis to a period is not available)

For each entity, a limited series of visualizations is represented (Example: The network of individual collaborations is represented in the profile of a given country. However, the network of institutional collaborations is not represented in the same scientometric profile)
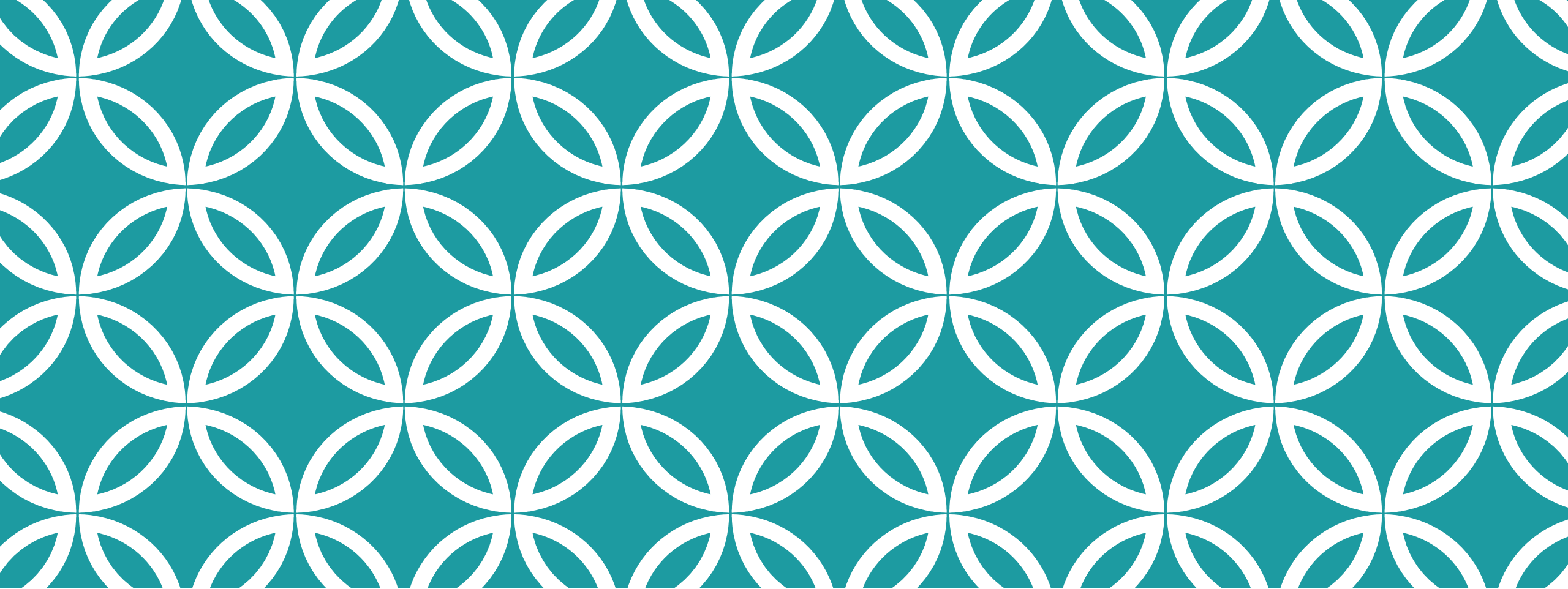
SCHOLIA AND LIVING SCIENTOMETRICS

WHAT SCHOLIA SHOULD BECOME

Since a SPARQL query can extract required data depending on the state of the knowledge graph, a Scholia profile can be converted into a scientometric study that can be updated in real time.

To achieve this interesting objective, we need to construct SPARQL queries based on Wikidata's linked bibliographic data to represent each constituent of a scientometric study.

TASKS

Task	Status
Introduction	
The analyzed entity in context	Only for research topics
List of scientometric work on the entity	No
Restriction to a given period	No
Coupled analysis of two entities	Yes
Methods	
Document explaining the Scholia profile generation method	No
Mulilingualism of the Scholia interface	No

TASKS

Task	Status
Results	
Support for Patents and Legal Texts	No, requires support of this information on Wikidata
Support for Publication Types	No
Support for raw, relative, or normalized scientometric metrics and indicators	No
Support for statistical distributions	Partially (publications for each year, publications for each area, etc.). No evaluation of scientometric distribution laws (Bradford's law …)

TASKS

Task	Status
Results	
Statistical correlation study	No
Scientometric networks	Partially (Only a few scientometric networks are represented for each type of entity. More comprehensive representation is required).
Support of scholarly citations	No (requires the mass upload of OpenCitations datasets or the use of federated SPARQL queries)
Centrality indicators of scientific entities in a scientometric network	No
Altmetric indicators (Wikipedia Pageviews, Wikipédia Citations)	No

TASKS

Task	Status
Discussion	
Comparison with previous results of the same entity	No, requires evaluation of statistical significance
Comparison with the results of other entities of the same class	No, requires evaluation of statistical significance
Comparison with the results of a generalized entity	No, requires evaluation of statistical significance
Conclusion	
Identification of significant differences from the discussion	No, requires evaluation of statistical significance

CONCLUSION

The creation of linked bibliographic data in RDF format on Wikidata is of great use in scientometric analyses

- Enables real-time quantitative and qualitative analysis of the evolution of scientific research using SPARQL.
- Easy to enrich, validate and develop from other resources.

Scholia is an interface that uses the SPARQL query language to generate scientometric profiles for several types of scientific entities.

- Profiles are updated as Wikidata is updated.
- The profiles can be enriched by adding SPARQL queries allowing the interface to become a platform for living scientometric studies.

REFERENCES

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European journal of operational research, 246*(1), 1-19.

Turki, H., Hadj Taieb, M. A., Ben Aouicha, M., & Abraham, A. (2020). Nature or Science: what Google Trends says. *Scientometrics, 124*(2), 1367-1385.

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM, 57*(10), 78-85.

Rasberry, L., Willighagen, E., Nielsen, F., & Mietchen, D. (2019). Robustifying Scholia: paving the way for knowledge discovery and research assessment through Wikidata. *Research Ideas and Outcomes, 5*, e35820.

Nielsen, F. Å., Mietchen, D., & Willighagen, E. (2017, May). Scholia, scientometrics and wikidata. In *European Semantic Web Conference* (pp. 237-259). Springer, Cham. doi:10.1007/978-3-319-70407-4_36.

Turki, H., Hadj Taieb, M. A., & Ben Aouicha, M. (2020). *Robustifying Scholia, An overview*. University of Sfax.
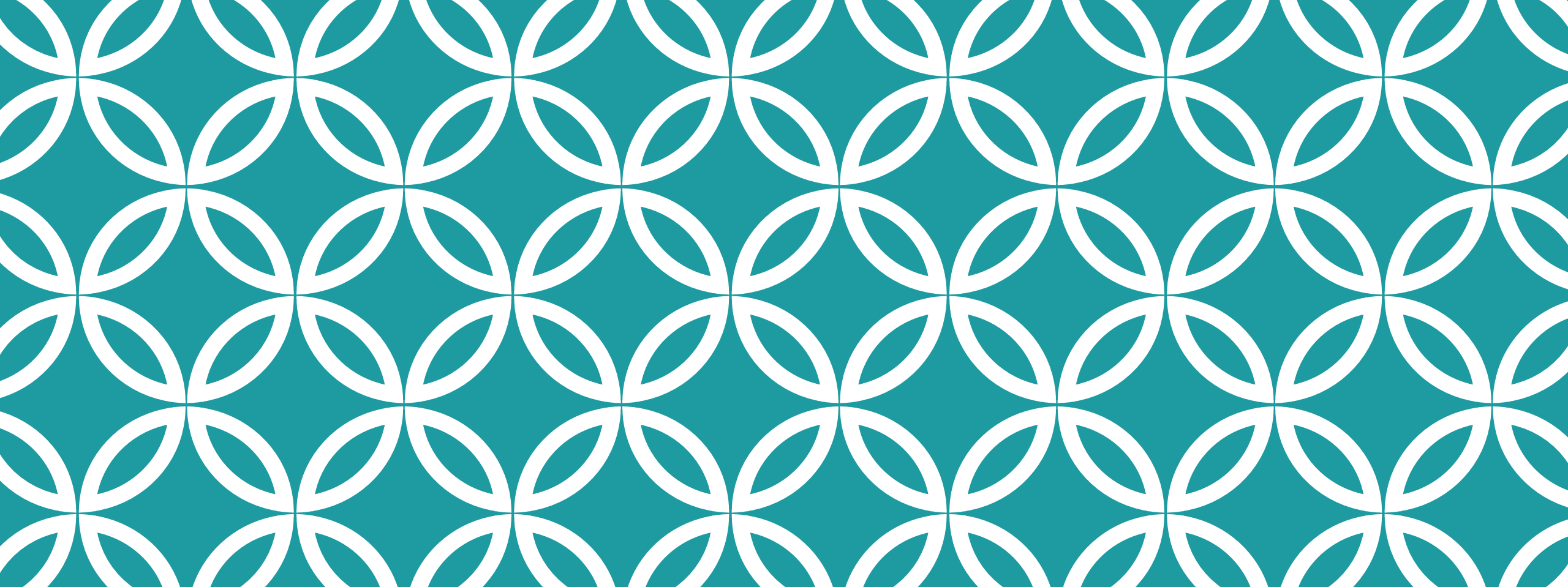
Houcemeddine Turki
houcemeddine.turki@medecinesfax.org

THANK YOU
Questions

{} wikicite