



Date: November 26, 2023

Prof. dr. P.T. Groth
Informatics Institute (IVI)

Mailing address:

P.O. Box 94323
1090 GH Amsterdam
Netherlands

Visiting Address:

UvA - LAB42
Science Park 900
1098 XH Amsterdam
Netherlands

+ 31 (0) 20 525 7460 (secr)
p.t.groth@uva.nl
pgroth.com

Dear Review Committee,

This is my review of the habilitation thesis of Dr. Jakub Klímek entitled “Improving data accessibility and interoperability using linked data”.

The thesis presents a comprehensive review of Dr. Klímek’s work on systems and approaches for both publishing data in the form of web standards compliant data (termed linked data) and as well as enabling this consumption. This habilitation is clearly the result of Dr. Klímek’s own work with his collaborators based on my analysis of the Turnitin results provided to me. Any overlapping content detected stems from his prior publications.

In terms of contribution, the habilitation shows a clear contribution to the domain of linked data publication. Exposing data in the form of web standards provides a number of technical and methodological challenges. Dr. Klímek’s work focuses on how to do this more efficiently and in a developer friendly manner. This is bundled together under the heading of LinkedDataPipes. This system provides a comprehensive extract transform load approach specifically tailored to the challenges of web standards data. An important point here is to look at how to do so in an efficient manner, here, the thesis introduces the notion of chunking to the literature and how to specifically tailor this method to the kinds of input data seen in linked data systems. Another contribution is the focus on being able to reuse components of these pipelines across multiple data publication tasks.

Another key aspect with respect to web standards oriented data is to look at how to publish data for consumption. Here, the thesis describes new methods for the exposure of underlying data using the web standard for data catalog metadata (DCAT). Importantly, this work builds on the LinkedDataPipes work discussed above to overcome the challenges in moving to a more distributed web oriented view of data catalogs. Another important contribution, in terms of consumption, is the analysis of tools that can natively consume linked data. The comprehensive survey of linked data consumption tools that is part of the thesis showed a path forward in making tools that can more effectively allow users to take advantage of the unique characteristics on linked data.

An often overlooked part of research contributions is their testing and application in real world scenarios. This habilitation shows a very strong embedding in such real world engagement, which is commendable. Specifically, the research has resulted in available open source systems which have been used by the

Czech Republic's open data catalog, the Czech Social Security Administration, a municipality (Brno), Charles University, and a startup to name a few. It is important to note that such systems research in computer science often takes longer but as shown here can translate to a larger impact.

In terms of publications, the habilitation consists of solid peer review publications. There are key publications in the major outlets within the Semantic Web community (e.g. the Journal of Web Semantics and the Semantic Web Journal) as well as smaller events affiliated with the community (e.g. poster presentations and workshops). For these later smaller publications, it would have been nice to see these publications to be expanded into larger contributions. Furthermore, while their methodological contribution of the results are evident to an expert in the field, a more overarching paper with the key methodological insights and generalization would have been a beneficial addition. That being said the candidates ability to publish is well documented in his CV.

Overall, the habilitation thesis is a comprehensive work in the area of the Semantic Web and provides an interesting approach to the overall problem of how to make and facilitate the consumption of web standards-based data. Hence, from my perspective *it meets the standard requirements for a habilitation thesis*.

Based on the habilitation, I think Dr. Klímek has the potential for shaping the research direction around data on the web especially with the deep connection he has both to creating systems and having them used in practice. The Semantic Web as an area of interest also has a strong systems tradition and this is evident in the habilitation. He is also active in the community in terms of reviewing and involvement in the creation of web standards. Such contributions are emphasized with this area of research. In my own institution, we look at a comprehensive view of a candidate to become associate professor, this includes their research output but also contributions to innovation (i.e. use of research) as well as education. Given these and the candidates CV, I would *recommend the further progression in the habilitation procedure*.

If you have any additional questions or would like me to expand anymore on this review, feel free to contact me.

Sincerely,

Paul Groth, Ph.D.
Professor of Algorithmic Data Science
Lead – Intelligent Data Engineering Lab
Scientific Director – Data Science Centre
University of Amsterdam