

May 29, 2023

Prof. Dr. Jan Hajič  
Chair of the Habilitation Committee  
Charles University  
Prague, Czechia

Dear Committee Members:

It is my pleasure to have reviewed the Habilitation Thesis of Dr. Jakub Klímek.

The thesis focuses on addressing the barriers-to-entry that many organisations face when deploying Linked Data in order to manage and publish their data. While the Linked Data principles provide a clear added value over publishing data in forms such as PDFs, CSVs, etc. – as best expressed by the “5-Star Scheme” – this added value can also come at a considerable cost in terms of reusing existing vocabularies, minting IRIs, setting up correct deferenceability and content negotiation, etc. Beyond this, the Linked Data may also not be easy to understand or consume; for example, the resulting Linked Data may not be compatible with the existing data management and processing platforms of the particular organisation.

In this context, the work of Dr. Klímek, as described by this Habilitation Thesis, has focused on the design, implementation and evaluation of tools to assist in the production and consumption of Linked Data, thus lowering the aforementioned barriers-to-entry for organisations. Much of the described work centres around LinkedPipesETL: a visual system that allows users to define a data flow – through the composition and interlinking of data processing components – supporting Linked Data preparation and consumption. Later work looks at various extensions to this tool, including enhancements to its scalability for ingesting very large datasets through data chunking strategies, visualisation of DCAT-AP metadata in the context of data catalogues, validation with respect to SKOS and RDF Data Cube vocabularies, storage of configurations in Solid PODs, and tools to support mapping Linked Data to CSV. A key theme in the work of Dr. Klímek has been the exploration of use-cases and the real-world deployment of the tools and techniques developed; such deployment includes applications involving the Czech National Open Data Catalog, the Czech Social Security Administration, the H2020 project OpenBudgets.eu, etc.

Overall, I have found the thesis to be well-written, well-motivated and with a clear vision. I wish to particularly commend the significant contributions in terms of the deployment of the proposed tools and techniques in real-world settings, which not only bolsters the impact of the work, but has also helped to evaluate the work in the target setting, and (I am sure) to inspire many of the improvements described. Such work on studying deployment is sometimes all too

rare within the Linked Data community. I must also commend the work on the *Survey of Tools for Linked Data Consumption*, which I find not only to be highly impressive in terms of the breadth of its scope and the depth of its analysis, but also the conceptual framework used to discuss relevant literature. This is certainly a very valuable contribution for the community.

In more general terms, and beyond the particular contributions described in the Habilitation Thesis itself, the DBLP profile for Dr. Klímek lists 60 publications coauthored by him, including 11 journal papers, 48 conference/workshop publications, and 1 book chapter. Here I can draw attention to publications in the Semantic Web Journal, the Journal of Web Semantics, the Knowledge and Information Systems journal, and the Extended Semantic Web Conference, which are top venues in the Linked Data/Semantic Web area. The acceptance of publications at these venues demonstrates the acceptance of Dr. Klímek's by his peers in the community. Dr. Klímek has further published two single-author publications, including in the highly-reputable Journal of Web Semantics, which – along with the variety of co-authors with which he has worked on other publications – helps to evidence his independence as a researcher. His Google Scholar profile indicates that his work has received 824 citations, with a *h*-index of 15, which illustrate the considerable academic impact of his work.

Upon reviewing the originality report produced by the Turnitin service, I can see no question of irregularity. The few instances that are highlighted either relate to Dr. Klímek citing his own previously-published work, expanding well-known acronyms, or discussing established concepts (such as the Linked Data principles) with adequate citation. I see no reason to doubt the originality of the novel text presented in the Habilitation Thesis.

Overall, I believe that Dr. Jakub Klímek meets the established requirements for this habilitation, and highly recommend that the Habilitation Thesis is accepted by the committee so that Dr. Klímek can continue his promising career.

Yours sincerely,

Prof. Dr. Aidan Hogan  
Associate Professor  
DCC, FCFM, Universidad de Chile  
ahogan@dcc.uchile.cl