Multilingual taxonomy initiative - TaDiRAH as community of practice

Borek, Luise

luise.borek@tu-darmstadt.de
Technical University of Darmstadt, Germany

Hastik, Canan

canan@hastik.de
Technical University of Darmstadt, Germany

Dombrowski, Quinn

qad@stanford.edu
Stanford University

Broeder, Daan

d.g.broeder@uu.nl
KNAW Humanities Cluster, The Netherlands

Rockenberger, Annika

annika.rockenberger@ub.uio.no
University of Oslo Library

Nagasaki, Kiyonori

nagasaki@dhii.jp
International Institute for Digital Humanities

Mochizuki, Ryo

rmochizu@uni-muenster.de
The University of Tokyo

Katakura, Shumpei

katakura-shumpei642@g.ecc.u-tokyo.ac.jp
Tokyo National Museum

Cupar, Drahomira

dcupar@unizd.hr
University of Zadar

Ohmukai, Ikki

i2k@l.u-tokyo.ac.jp
The University of Tokyo

The Taxonomy of Digital Research Activities in the Humanities (TaDiRAH), initially a transatlantic initiative for indexing Digital Humanities (DH) research practices, dates back before 2014 (TaDiRAH 2016). From the start, TaDiRAH has benefited greatly from input and contributions from the community which actively helped shape the taxonomy by identifying and defining DH research concepts and producing translations. [1] TaDiRAH supports a pragmatic classification of DH research content and increases transparency, findability, and reusability of resources. Due to its multilingual nature and its availability in seven languages, this can be particularly beneficial for content in otherwise marginalised languages. The current TaDiRAH SKOS core vocabulary application (TaDiRAH 2020a) emerged from the concept hierarchy of this grassroots initiative and promotes cross-disciplinary collaborative work.

Collaboration, interdisciplinarity and multilingualism remain the cornerstones of TaDiRAH (TaDiRAH 2022). The call for Digital Humanities 2023 enables building the TaDiRAH community with a common strategy that binds a group of users by the goal to describe, classify and link DH research activities. The half-day workshop provides the opportunity to engage new stakeholders and supporters from the global DH community to further shape the taxonomy collaboratively. The current Version 2.0, formalized in SKOS, provides a revised structure and semantics with a focus on research activities and techniques. Following the FAIR principles, the taxonomy is published as linked open data. It includes translations in German, French, Spanish, Portuguese, Serbian, Italian, and Japanese and is enriched with Wikidata definitions based on a workflow for further maintenance, development, and quality management. In addition, it is considered to use the taxonomy also to establish transdisciplinary connections in teaching and education, based on research activities, if these prove to be generic. Here, there is close collaboration with the NFDI Section Edutrain.

The key questions to be discussed in this workshop are whether TaDiRAH meets the needs of the DH community, how further experts can be inspired applying TaDiRAH, and what possible pain points are. From projects such as EOSC and SSH, it has become clear that community-specific vocabularies and taxonomies are extremely valuable for searching and retrieving materials, even if they are still incomplete and evolving. A special attention will be laid on multilingualism as the community-driven taxonomy makes resources visible that are designed in languages other than English. Here, the taxonomy can contribute to a more inclusive DH.

Aim of this DH2023 pre-conference half-day workshop is to discuss the current state and potential uses of TaDiRAH with various experts from different application areas. The workshop is structured in three sections. Section 1 introduces the taxonomy, its genesis, the conceptual model, its representation, and formalization. In Section 2, the invited pre-workshop leaders address the role of shared vocabularies, their curation, and the integration of TaDiRAH into catalogs, marketplaces, and specialized terminology with regard to multilingualism. In Section 3 the workflow for assigning term translation and correction is outlined, working groups are formed, and an intended Croatian translation will serve as a use case to introduce new languages into the taxonomy and update existing ones.

Section 1 — Introduction and State of TaDiRAH

30 minutes
Luise Borek, Technical University of Darmstadt (luise.borek@tu-darmstadt.de) / Canan Hastik, Technical University of Darmstadt (canan@hastik.de)

The transition of the taxonomy to a semantic vocabulary has been accompanied by some necessary changes. This introduction provides information about the latest version, its multilingualism, technical and infrastructural representation with a view to the col-

laboration and participation opportunities and tasks (TaDiRAH 2020b). This concerns in particular the implementation of the relations, which required a specification of scope notes. As a consequence, the multiple translations currently vary from the latest English version of the taxonomy. Aim is to ensure that the taxonomy continues to function pragmatically and is community-driven.

Section 2 — Community of Practice

90 Minutes

In this section, the community of practice reports with impulse talks — each in a length of 20 minutes — on its work on and experience with TaDiRAH discussing new multilingual collaborative project ideas. These cover implemented applications presenting the nature and function of the taxonomy by providing translations. With these use cases and applications, the the taxonomy will be presented, discussed, and collaboratively thought further.

Quinn Dombrowski, Academic Technology Specialist, Stanford University, Division of Literatures, Cultures, and Languages (DLCL) (qad@stanford.edu)

One dream of a shared, widely adopted taxonomy is that it will lead to an ever-growing network of resources connected through their use of a common vocabulary. But some of the most obvious candidates for such a vocabulary are difficult to maintain over the long-term due to the amount of labor involved. The DiRT Directory (DiRT 2015), one of the projects behind TaDiRAH's creation, has been defunct since 2016. This case study will speak to the role of TaDiRAH and other evolving vocabularies in relation to projects that die, and the data that lives on from them.

Daan Broeder, Project Manager & consultant at CLARIN ERIC, a partner in Social Sciences & Humanities Open Cloud (SSHOC) (d.g.broeder@uu.nl)

Currently there is an abundance of resource catalogs and recommender systems available in the Social Science Humanities (SSH) domain (EU 2022). By using metadata harvesting such as Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) it becomes easy to create yet another catalog. But without carefully selecting and curating that metadata not much added value will be provided for the user. Such an effort is costly and requires long-term collaboration within and between communities. Including searching and filtering options using ontologies like TaDiRAH in recommenders and catalogs will significantly add value for end-users.

Annika Rockenberger, Senior Academic Librarian, University of Oslo Library, Humanities and Social Sciences library, Research services (annika.rockenberger@ub.uio.no)

In 2020, the University of Oslo Library, Library for Humanities and Social Sciences organized a workshop introducing TaDiRAH as a concept model and a tool for mapping out the various digital research activities. Though the taxonomy and its potential applications were positively evaluated regarding research support and teaching, the lack of a Norwegian translation of research activities was one major obstacle in fully understanding, adopting, and incorporating its concepts and definitions. Our team collaborates with the Center for Computing in Science Education's task group on terminology. We have started translating the current TaDiRAH version into Norwegian (bokmål/nynorsk) which will be integrated into the official database of Norwegian specialized terminology (https://www.termportalen.no/) under the supervision of the Norwegian Council for Language (Sprakradet 2017).

Kiyonori Nagasaki, International Institute for Digital Humanities (nagasaki@dhii.jp) / Ikki Ohmukai, The University of To-kyo, (i2k@l.u-tokyo.ac.jp) / Ryo Mochizuki, The University of Tokyo (rmochizu@uni-muenster.de) / Shumpei Katakura, Tokyo National Museum (katakura-shumpei642@g.ecc.u-tokyo.ac.jp)

At the beginning of 2021, 17 mainly young researchers, gathered for the online translation process for the skos:prefLabel, skos:definition, and skos:notes (SKOS 2009) in TaDiRAH. Since each of the researchers worked in a different field in the humanities and computer science, it was necessary to discuss whether or not to unify the translations of technical words, and the process was managed by three editors. After the editors approved the translations, several rounds of corrections were made. One of the difficulties in translation was handling the slightly different ways of classifying concepts in Japanese scholarship and the fact that one English word is sometimes translated into other words in various fields in the Japanese-speaking world.

Section 3 — Collaboration: Work on Language Versions

60 minutes

With the aim of collaboration, cross-national and -disciplinary integration, this section aims to encourage additional language versions of TaDiRAH. Together, bottom-up, an application-oriented workflow for future joint translation work will be defined and evaluated, encouraging further translation efforts in the region. Depending on the interests of the participants, this section is also open to the formation of further working groups on other tasks, and supports the provision of new language versions.

Drahomira Cupar, University of Zadar (dcupar@unizd.hr)

Aim of this section is to show preparatory steps which are crucial to ensure adaptation and translation of TaDiRAH vocabulary in Croatian language. The interesting starting point is that the Croatian version of TaDiRAH can build on the existing Serbian version, due to similarities and semantics the two languages have in common. Harmonization and alignment of those two languages and parallel examples from English can serve as a model for the future translations. Participants will have the opportunity to work with actual linguistic data and learn how to prepare and initiate the translation project. The goal is to showcase Croatian language translation in order to empower more language communities to work on translations in other languages.

The addressed target audience includes researchers from DH, related fields, cross-disciplinary scientific communities from NFDI e.V. particularly involved content providers and curators of educational material, and interested research data stewards. Given previous participation in the taxonomy and the assumption that the contributors are engaged, we expect participation of a minimum of 25 people.

Notes

1. In 2021, a study on the use and impact of TaDiRAH was conducted as part of the DARIAH final report, which also documented the benefits for the community (Borek / Hastik 2021).

Bibliography

Borek, Luise / **Hastik, Canan** (2021): "TaDiRAH – Taxonomy in the Digital Research Activities. A DAR-

IAH Impact Case Study", in: Edmond, Jennifer / de Loof, Chris (eds): DARIAH ERIC. *A Research infrastructure to enhance and support digitally-enabled research and teaching across the Arts and Humanities*. Annual Report 2021. https://www.dariah.eu/activities/impact-case-studies/tadirah-the-taxonomy-of-digital-research-activities-in-the-humanities/ [24.04.2023].

Borek, Luise / **Hastik, Canan** / **Khramova, Vera** / **Geiger, Jonathan** (2020): *TaDiRAH – Taxonomy of Digital Research Activities in the Humanities*. Version 2.0.1. https://vocabs.dariah.eu/tadirah/en/ [24.04.2023].

DiRT (2015): "Digital Research Tools (DiRT) Directory". https://digitalhumanities.berkeley.edu/resources/digital-research-tools-dirt-directory [24.04.2023].

EU (2022): "SSH Integration. Integrating Social Sciences and Humanities into each of the priorities of Horizon 2020". https://research-and-innovation.ec.europa.eu/research-area/social-sciences-and-humanities/ssh-integration_en [24.04.2023].

SKOS (2009): "SKOS Simple Knowledge Organization System Namespace Document". https://www.w3.org/2009/08/skos-reference/skos.html [24.04.2023].

Sprakradet (2017): "Rettleiing i terminologiarbeid". https://www.sprakradet.no/Sprakarbeid/Terminologi/slik-lagar-du-ei-termliste/Rettleiing-i-terminologiarbeid [24.04.2023].

TaDiRAH – Taxonomy of Digital Research Activities in the Humanities (2022): https://tadirah.info/ [24.04.2023].

TaDiRAH – Taxonomy of Digital Research Activities in the Humanities (2020): Future Tasks. https://tadirah.info/pages/Tasks.html [24.04.2023].

TaDiRAH – Taxonomy of Digital Research Activities in the Humanities (2016): https://www.vocabularyserver.com/tadirah/en/sobre.php [24.04.2023].