Beitrag ID: **3** Typ: **nicht angegeben**

Dark Data in Mathematics

Mittwoch, 26. Oktober 2022 11:25 (25 Minuten)

Dark data is data that is poorly managed [1, 2]. It is diametrically opposed to FAIR data because its epistemic status is unclear, and it is neither findable, accessible, interoperable, nor reusable. For example, research data may be uncurated, unavailable, unannotated, biased, or incomplete. Examples of dark data in scientific computing include the vast amounts of data that are held unavailable, unsearchable, and unannotated on the parallel file systems or tape archive storages of high-performance computing centers [2]. But what is dark data in mathematics, i.e. dark data concerning mathematical research data such as models, formulas and other abstract artefacts [3]? Can these also be unavailable, unsearchable or unannotated, depending on their specifics? The talk will explore the extent to which even these more abstract types of assets and research data in mathematics can become dark in the absence of good research data management practice.
Dark data as a negative category serves as an analytical tool to work out how the data management processes can be improved and aligned with the FAIR principles. It is a requirement for research data infrastructures that they enable good practices in research data management [4]. For this purpose, we propose a strategy based on metadata standardization by ontologies expressed simultaneously and consistently in OWL description logic (to permit SPARQL queries, etc.) and first order logic (e.g., to facilitate answer set programming) [5]. Our primary aim is to support the documentation of epistemic metadata [6], i.e., information about the knowledge status of data, which we propose to standardize by a mid-level ontology [7]. Below this, at the domain level, documenting the epistemic metadata in a way that is adequate for each academic community will require an evaluation of disciplinary scientific practices and conventions in detail [6]. It is here that NFDI consortia can provide dedicated support by encouraging academic communities to reflect upon their own practices and engage in community-driven metadata standardization. The key objective is for all research data to attain epistemic FAIRness: A status where it is accessible and intelligible to all stakeholders in what way the data has been given an interpretation as knowledge, making that knowledge and its reuse machine-actionable and thereby avoiding that it falls into darkness.

[1] Heidorn, P. B. “Shedding light on the dark data in the long tail of science.” Library Trends 57.2 (2008): 280-299.
[2] Schembera, B., and Durán, J. M. “Dark data as the new challenge for big data science and the introduction of the scientific data officer.” Philosophy & Technology 33.1 (2020): 93-115.
[3] Koprucki, T., Tabelow, K., and Kleinod, I. “Mathematical research data.” PAMM 16.1 (2016): 959-960.
[4] Horsch, M. T., et al. “Interoperability and architecture requirements analysis and metadata standardization for a research data infrastructure in catalysis.”In Pozanenko, A., et al. (eds.), Proceedings of DAMDID/RCDL 2021, CCIS no. 1620, Springer, 2022.
[5] Horsch, M. T. “Mereosemiotics: Parts and signs.”In Sanfilippo, E. M., et al. (eds.), Proceedings of JOWO 2021, CEUR no. 2969, CEUR-WS, 2021.
[6] Schembera, B., and Horsch, M. T. “Dark data and epistemic metadata in molecular modelling.”Submitted, 2022.
[7] Horsch, M. T., and Schembera, B. “Documentation of epistemic metadata by a mid-level ontology of cognitive processes.”Submitted (preprint doi:10.5281/zenodo.6638457), 2022.

Hauptautoren: Dr. SCHEMBERA, Björn (IANS / University of Stuttgart); Dr. HORSCH, Martin Thomas (Department of Data Science, Norwegian University of Life Sciences)

Vortragende(r): Dr. SCHEMBERA, Björn (IANS / University of Stuttgart)

Sitzung Einordnung: Contributed Talks