FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Data Science Day 2024

8 May 2024

Friedrich Schiller University Jena

Book of Abstracts

Olia Blacher (Ed.)
olia.blacher@uni-jena.de

Contents

Construction of Machine Learning descriptors for modelling glass properties

Authors: Felix Arendt; Marek Artur Sierka

Friedrich Schiller University Jena

Corresponding Authors: felix.arendt@uni-jena.de, marek.sierka@uni-jena.de

There are three sources of information for modelling glasses: composition, structure, and processing parameters. Glass composition is the most readily available information, so it is desirable to extract as much performance from it as possible. An approach to this is ab initio descriptors. Here ab initio derived descriptors refer to mean values of the ab initio calculated properties of the glass components, weighted according to their respective mole fraction. Machine Learning (ML) models trained with ab initio descriptors can reproduce the performance of their compositional counterparts for a large, comprehensive data set, and in some cases even improve it for a smaller data set covering a variety of oxides, chalcogenides, and metallic glasses.

However, comprehensive and consistent datasets are rarely available. High-throughput molecular dynamics simulations offer a good approach to generate a consistent dataset for the training of ML models. Here, simulations were performed for the NABS system to calculate the glass density, elastic moduli and glass transition temperature. Predictions of the trained ML models are in good agreement with data from the SciGlass database and measurements from the Bundesanstalt für Materialforschung und -prüfung (BAM). In addition, the predictive performance is further improved by including BAM measurement series in the training set. This combined approach of ML, molecular dynamics simulations, measurements and novel ab initio derived descriptors provides promising results and could be a valuable toolbox to accelerate time efficient and resource efficient glass design.

Digital forest inventory based on drone imagery

Author: Steffen Dietenberger[1]

Co-authors: Marlin M. Mueller[1]; Markus Adam[1]; Felix Bachmann[2]; Boris Stöcker[3]; Sören Hese[2]; Clémence Dubois[1]; Christian Thiel[1]

[1] *DLR-Institut für Datenwissenschaften*

[2] *FSU-Lehrstuhl für Fernerkundung*

[3] *Universität Münster-Institut für Geoinformatik*

Corresponding Author: steffen.dietenberger@dlr.de

Accurate and up-to-date data is essential for sustainable forest management and effective monitoring. In the German forestry sector, several factors underline the importance of this data: a) significant impacts of climate change-related events such as drought and heat on forest structure and health, b) staff shortages in parts of the forestry sector, c) trend towards conversion of monoculture stands into more diverse and complex forests, and d) increasing implementation of digitalization strategies. To address these challenges, UAV-based structure from motion (SfM) data products offer a cost-effective approach. We use the spectral and geometric datasets to develop methods for a digital forest inventory and automatically derive parameters such as breast height diameter (BHD), tree stem positions, delineation of individual tree crowns and deadwood. Various flight configurations, including nadir and oblique camera angles, and different acquisition times were combined to generate a comprehensive digital representation (3D point clouds, orthomosaics and elevation models) of the forest structure, containing ground, stem and canopy information.

In a study area within the Hainich National Park, we analyzed the combination of data from leaf-on and leaf-off conditions to enhance the derivation of the tree stem positions and the delineation of individual tree crowns. Stem coordinates were obtained from the leaf-off point cloud using a cluster algorithm with an accuracy of 0.87 and were used as markers for a crown delineation in the leaf-on dataset. The combination of the datasets slightly improved the crown delineation. Additionally, the leaf-off datasets enable the creation of orthomosaics of the forest floor. These orthomosaics were employed to train a deep learning model (U-Net), which successfully segmented lying deadwood trunks, detecting 96.89% of the deadwood objects.

Figure 1: A profile cross-section of the normalized point clouds from the study area in Hainich National Park was created using UAV RGB images and an structure from motion (SfM) approach. The top section displays the point cloud without leaves from March 29, 2022, the middle section shows the point cloud with leaves from August 4, 2021, and the bottom section presents a merged dataset combining both.

Neuromorphic Information Processing in Nonlinear Optical Fibers

Author: Sobhi Saeed

Co-author: Mario Chemnitz

Leibniz Institute of Photonic Technology, Albert-Einstein Str. 9, 07745 Jena, Germany

Corresponding Authors: saeed.sobhi@uni-jena.de, mario.chemnitz@leibniz-ipht.de

The performance limitations of traditional digital computer architectures have led to a rise in the development of neuromorphic hardware, with optical solutions gaining popularity due to their energy efficiency, high speed, and scalability. In recent theoretical and experimental studies, nonlinear interactions between optical spatial modes have been used to emulate basic neural network functionalities. The use of such transient nonlinear dynamics may enable electro-optical limitations to be bypassed in signal conversion, and may also facilitate the development of new approaches towards highly efficient, multi-layer (deep) processing.

The main focus of this research work is the exploration and functionalization of highly complex and nonlinear wave dynamics, such as broadband light generation known as supercontinuum generation. Such spectral broadening inside fibers performs an input-output nonlinear mapping similar to the hidden layers of neural networks. Our research addresses the key challenge of identifying benchmarks to reveal the computational power of such analog wave computers.

We build an optical computing system, that enables encoding a dataset into an optical signal via spectral domain phase modulation and processing this signal in a nonlinear fiber. The system is read out via a spectrometer, following which a weight matrix is trained on this system output to map the high-dimensional space of the system to interpretable prediction results, a training method known from reservoir computing. We share our benchmark-tests results to identify performance scalability and the system's capability to solve nonlinear problems. The research thus extends our understanding of analog brain-inspired hardware for information processing in the optical domain to help address global challenges such as green computing, Big Data communications, and intelligent medical diagnostics.

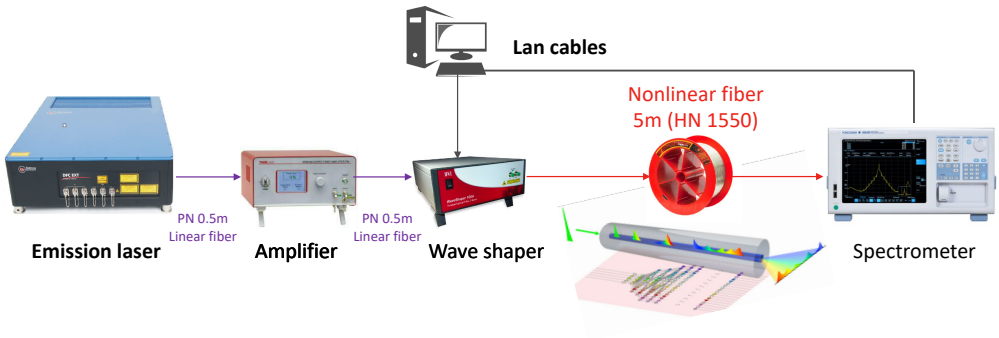

Figure 1: The figure shows the propagation of a light signal, originating from a femtosecond laser, undergoes amplification by an EDFA, spectral phase modification (Encoding in our NN) by a Wave Shaper, spectral broadening (non-linear mapping in our NN) in a non-linear fiber, and subsequent recording and storage.

Exploring Machine Learning Approaches for Language Identification in a Code-switching Dataset

Author: Olha Kanishcheva

Co-author: Maria Shvedova

Friedrich Schiller University Jena

Corresponding Authors: kanichshevaolga@gmail.com, masha.shvedova@gmail.com

In multilingual societies, code-switching or code-mixing, where individuals alternate between languages within discourse, poses a challenge for natural language processing (NLP). Code-switching, a common linguistic phenomenon, is observed globally, such as when bilingual individuals seamlessly transition between English and Spanish or German and Turkish in a conversation. In multilingual communities like India, speakers may effortlessly blend Hindi and English within the same discourse [1-2]. In the framework of our research, we created the code-switching Ukrainian-Russian dataset, labeled it, and applied different models of language identification on the token level. There are different code-switching datasets, in our work, we investigate the Ukrainian-Russian dataset, which adds complexity to identification since these languages have almost the same alphabet. We focused specifically on intra-sentential and intra-word categories of code-switching in our analysis, omitting inter-sentential instances.

Our dataset consists of separate sentences selected from the corpus of Ukrainian parliamentary transcripts of the Verkhovna Rada [3]. All tokens of these sentences were labeled by 6 categories: Ukrainian (UK), Russian (RU), Ukrainian-Russian hybridized words (MIX), numbers (NUM), dialects, other languages, etc. (OTH), and punctuation (PUNCT). The statistical information about these categories is presented in the table below.

To build the code-switching language identification model, we explored the Conditional Random Fields (CRF), Long Short Term Memory networks (LSTM), and BERT models for language identification on the token level. Our experiments indicate that LSTM and BERT showed good results for this task, F1-measure of more than 89% for two categories ‘Ukrainian’ and ‘Russian’.

However, these models fail in hybrid word classification when trained on this dataset, this is due to the lack of data for this category. Currently, there are already code-switching corpora that have morphological and syntactic markup. Theoretically, adding these properties will allow for more accurate identification of the language, namely tokens from the ‘MIX’ category.

Labels	Description	Count	%
UK	Ukrainian words	91 592	41.85
RU	Russian words	82 375	37.65
MIX	Ukrainian-Russian hybridized words (Surzhik)	652	0.29
NUM	Numbers	2 123	0.97
OTH	Dialects, other languages, etc.	234	0.1
PUNCT	Punctuation	41 832	19.11

Table 1: Corpus statistics for the language pair Ukr-Rus.

References

[1] Tanmay Chavan, Omkar Gokhale, Aditya Kane, Shantanu Patankar, and Raviraj Joshi. 2023. My Boli: Code-mixed Marathi-English Corpora, Pretrained Language Models and Evaluation Benchmarks. In Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings), pages 242–249, Nusa Dua, Bali. Association for Computational Linguistics.
[2] A. F. Hidayatullah, A. Qazi, D. T. C. Lai and R. A. Apong, “A Systematic Review on Language Identification of Code-Mixed Text: Techniques, Data Availability, Challenges, and Framework Development”, in IEEE Access, vol. 10, pp. 122812-122831, 2022, DOI: 10.1109/ACCESS.2022.3223703.
[3] Olha Kanishcheva, Tetiana Kovalova, Maria Shvedova, and Ruprecht von Waldenfels. 2023. The Parliamentary Code-Switching Corpus: Bilingualism in the Ukrainian Parliament in the 1990s-2020s. In Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP), pages 79–90, Dubrovnik, Croatia. Association for Computational Linguistics.

Supporting sustainable agricultural decision-making with geodata. The FieldMApp - a versatile data collection tool

Authors: Eric Schonert[1]; Felix Bachmann[2]; Martina Wenzel[3]; Ursula Gessner[3]; Maximilian Enderling[1]; Friederike Klan[1]; Sina Truckenbrodt[1]

[1] *Deutsches Zentrum für Luft und Raumfahrt e.V. (DLR), Institut für Datenwissenschaften*

[2] *FSU-Lehrstuhl für Fernerkundung*

[3] *Deutsches Zentrum für Luft und Raumfahrt e.V. (DLR), Deutsches Fernerkundungsdatenzentrum*

Corresponding Authors: sina.truckenbrodt@dlr.de, maximilian.enderling@uni-jena.de, ursula.gessner@dlr.de, eric.krueger@dlr.de, friederike.klan@dlr.de, felix.bachmann@uni-jena.de

Agricultural companies face the challenge to work economically and ecologically sustainable to ensure their long-term existence, while they are confronted with an increasing shortage of skilled labour. Digital solutions can partly help to address these three challenge dimensions. In crop production, geo- and Earth observation (EO) data emerge as an integral building block for digitising decision-making processes for a sustainable management of arable land. Therefore, EO data must be interpretable with regard to the variables relevant for decisions in crop production. This requires that the knowledge of domain experts, such as farmers and agricultural consultants, as well as complementary in-situ data can be linked to the EO data in form of machine-readable (geo-)data. The FieldMApp, a modularly structured and flexible mobile application supports this by enabling the digital acquisition and appropriate storage of required data.

The FieldMApp enables: users to collect data via forms, use case specific interfaces and serves as a geographic information system (GIS) to acquire and visualize geo-data in vector and raster format. Thus, openly available data sources for example national weather data, satellite images and products derived from satellite data like the Vegetation Deficit Layer (VDL) are obtainable via the FieldMApp GIS. The VDL is based on Sentinel-2 sensor data and describes current crop vitality. It serves as a meaningful example data set, which can be combined with local knowledge of farmers to provide an effective decision-making tool for more sustainable management of low yield areas. Together with farmers and agricultural consultants their requirements regarding the FieldMApp are identified and implemented.

The poster spotlights a use case for more sustainable management of low yield areas in detail and provides insights into the FieldMApp's ecosystem. Further, an architectural overview of the FieldMApp's software infrastructure will be shown. Open source technologies and open standards are utilized to construct the ecosystem. In compliance with OGC API Features and SpatioTemporal Assets Catalog (STAC) maximum integrability is ensured (e.g., QGIS). FieldMApp GIS capabilities are available online and offline. By integrating an external two frequency global navigation satellite system (GNSS) sensor, the FieldMApp enables precise mapping and digitization tasks, thus ensuring accurate agricultural operations in remote locations. The contribution will discuss the FieldMApp functionalities, future directions, and the role of digital tools in supporting agricultural decision-making.

Architectural Threat Elicitation Through Hybrid Information Flow Analysis

Author: Bernd Gruner

Co-author: Clemens-Alexander Brust

DLR-Institut für Datenwissenschaften

Corresponding Authors: bernd.gruner@dlr.de, clemens-alexander.brust@dlr.de

Software processes a vast amount of sensitive data, such as passwords, certificates, or configurations. However, tracing information flows within complex programs could help to identify and mitigate threats but poses challenges. One potential threat is the accidental disclosure of sensitive information to unauthorized recipients. Existing methods are insufficient in addressing the problem of information flow tracing. Fuzzers only detect crashes and hangs, taint analysis encounters difficulties due to overapproximation, and symbolic verification is overly restrictive for practical application.

We propose an approach for reconstructing and refining information flow graphs to uncover potential threats. We use static analysis and machine learning techniques to automatically reconstruct the information flow graph. Subsequently, we refine identified information flows using information flow fuzzing. Finally, we associate the threats using a rule-based system. Our approach provides a validated information flow graph of the software along with a list of elicited threats.

A differentiable rasteriser for point cloud data

Author: Wolfhart Sebastian Feldmeier

UKJ

Corresponding Author: wolfhart.feldmeier@uni-jena.de

We present a software package that performs rasterisation of point cloud data into regular grids in a way that is differentiable with respect to all relevant parameters.

This enables gradient-based methods for various tasks in which volumes are represented by point clouds, such as the 3D reconstruction of volumes from multiple projection images, or the estimation of camera poses.

As an application, we show how this package is used to estimate the volumetric shape of an object from tomography data.

The source code is freely available at https://github.com/microscopic-image-analysis/DiffPointRasterisation.jl

Utilizing Transformer Deep Neural Networks for Forest Height Estimation From Satellite Time Series

Authors: Markus Joachim Zehner[1]; Valentin Kasburg[1]; Clémence Dubois[1, 2]; Christian Thiel[2]; Alexander Brenning[1]; Jussi Baade[1]; Nina Kukowski[1]; Christiane Schmullius[1]

[1] *Friedrich Schiller University Jena*

[2] *DLR*

Corresponding Author: markus.zehner@uni-jena.de

Satellite remote sensing delivers cost-efficient information for an area-wide and up-to-date forest height estimation, one of the priority biodiversity metrics. Recent studies utilize spectral signatures and spatial patterns within Random Forest and U-Nets; the temporal domain in the data has yet to be fully explored. To tackle forest height estimation as a time series problem, we employ and modify the lightweight temporal attention encoder to leverage the full available time series and the temporal dependence within the data. We test our method in two target regions of different forest types: Hainich (Germany) and Remningstorp (Sweden). Furthermore, we study the influence of added information on forest height prediction accuracy, such as weather conditions and viewing geometry. We increase the accuracy achieved with Copernicus Sentinel 1 SAR (S-1) by supplying temperature conditions and the viewing geometry along the backscatter measurements. With additional fusing of information from S-1 and Copernicus Sentinel 2 Multispectral Instrument in a per-pixel regression, we achieve good agreements compared to LiDAR-derived forest height as a reference with an RMSE of 3.8 m and MAE of 2.4 m on spatially distinct holdout data within the same forest of the target regions.

An Einsum-inspired Tensor Contraction Engine

Authors: Max Engel; Alexander Breuer

Friedrich Schiller University Jena

Corresponding Author: max.engel@uni-jena.de

The Einstein summation convention (einsum) allows to define complex linear algebra expressions using a concise yet expressive notation. Tensor Networks and especially Quantum Circuits are important problems that can be formulated through Einstein summation.
We present an engine for the efficient evaluation of einsum expressions. Our engine identifies two core primitives for fast tensor contractions within an evaluation: Small GEMMs and small packed GEMMs. We achieve performance portability by using a just-in-time code generation approach for the appearing primitives. Further, our memory manager minimizes the overall memory footprint of our engine and targets high cache reuse.
We demonstrate the performance portability of the presented approach by evaluating our engine on Arm Neoverse, AMD Ryzen, and Intel Xeon CPUs. We show that our backend achieves peak utilization of over 40% for a set of demanding einsum expressions. In summary, our results show that the engine can outperform PyTorch's einsum implementation by over 2x.

SciGlass Next: bringing the largest open-access database of glass properties online

Authors: Ya-Fan Chen; Marek Artur Sierka

Friedrich Schiller University Jena

Corresponding Authors: marek.sierka@uni-jena.de, ya-fan.chen@uni-jena.de

The currently available glass databases are SciGlass [1] (which in 2019 becomes open source, with about 420,000 glasses) and INTERGLAD [2] (a commercial database with about 380,000 glasses). Since the plain data in SciGlass has become open source under the ODC Open Database License (ODbL), there are no longer any distributors in the world and no one to update the database. Given the importance of SciGlass to the glass community and the numerous scientific publications on predicting glass properties based on machine learning, in this work we have started to maintain the SciGlass database, fixed many erroneous data and developed an open-access web-based version of SciGlass, called SciGlass Next [3]. As an open source project, we particularly look forward to collaboration and input from the community to progressively integrate (general) predictive models and analysis tools for glass properties to further enrich SciGlass Next and turn data into knowledge and accelerate the development of novel glass materials.

[1] SciGlass database. https://github.com/epam/SciGlass
[2] INTERGLAD Ver. 8. https://www.newglass.jp/interglad_n/gaiyo/info_e.html
[3] SciGlass Next. https://sciglass.uni-jena.de

Encoding and Decoding the Microverse

Authors: Aristeidis Litos[1]; Daniel Rios Garza[2]; Bas E. Dutilh[1, 3]

[1] *Institute of Biodiversity, Faculty of Biological Sciences, Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena*

[2] *Université Paris-Saclay, INRAE, PROSE, 92160 Antony, France*

[3] *Theoretical Biology and Bioinformatics, Science4Life, Utrecht University, Padualaan 8, 3584 CH, Utrecht, the Netherlands*

Corresponding Authors: aristeidis.litos@uni-jena.de, danielriosgarza@gmail.com, b.e.dutilh@uni-jena.de

Complex dynamics and co-occurrence patterns transpire throughout the Microverse and are reflected in the composition of microbial communities. Such patterns and dynamics largely shape microbial community structure and function. Environmental parameters, such as available nutrients, play an important role in the microbiome composition. Understanding those patterns and their causing factors can benefit predictive microbiome modeling.

Context-specific exploration of patterns and dynamics in microbiomes has led to important milestones, such as antibiotics discovery, and substantial quantities of data. As a result, a holistic approach becomes feasible but challenging due to limitations that include the inherent sparsity of the dataset and the specificity of upstream analysis tools.

With this study, we aim to explore meaningful dimensionality reduction for microbiome data, establish techniques for analyses of large volumes of microbial communities, and highlight potential underlying biases.

We developed a neural network with an autoencoder architecture that encodes and decodes microbial community compositions. Our model highlights the biological information of microbiomes by projecting each sample on a universal taxonomic tree and applying the Generalized Unifrac distance, as a loss function, to calculate the difference between original communities and those reconstructed after embedding.

With this approach, we provide a biologically informed embedding of microbiomes in low dimensions. Furthermore, we examine potential biases on the data via our models.

A probabilistic model for biomolecular latent space trajectories

Author: Andreas Kröpelin

UKJ, AG Mikroskopische Bildanalyse

Corresponding Author: andreas.kroepelin@uni-jena.de

Cryogenic electron microscopy has proved to be a very poweful method for discovering the structure of large biomolecules. Many of such complexes can only fulfil their critical role in cellular processes because they are able to change their conformation, i.e. the spatial position of their atoms relative to eachother. It is therefore of high interest to also study such conformational changes, based on the static structures reconstructed using cryo-EM.

In this poster, the focus is on one subproblem: Given an unordered set of observed molecular structures, how can we find a continuous trajectory of structures that fits the observations?

To this end, an approach using lower dimensional latent spaces is used. A latent space is assumed to capture the complexity of the conformational changes but in much less dimensions than the conformational space.

Based on embedded molecular structures, the poster then explores a probabilistic model for piecewise linear trajectories in the latent space. It allows for a principled way of formalising reasonable requirements on such trajectories by the means of a posterior based score function. Finally, applications to biomolecular data and performing maximum aposteriori estimations of latent trajectories explaining their conformational changes are explored.

Convexity Certificates for Symbolic Tensor Expressions

Authors: Joachim Giesen; Paul Gerhardt Rump

Co-authors: Julien Klaus; Maurice Wenig; Niklas Merk

Friedrich Schiller University Jena

Corresponding Authors: joachim.giesen@uni-jena.de, paul.gerhardt.rump@uni-jena.de, julien.klaus@uni-jena.de

Knowing that a function is convex ensures that a global optimum can be computed efficiently. Here, we implement an approach to certify the convexity of functions by certifying that their second-order derivative is positive semidefinite. Both the computation of the second-order derivative and the certification of positive semidefiniteness are done symbolically. Previous implementations of this approach assume that the function to be optimized takes symbolic scalar or vector inputs, meaning that the second-order derivative is at most a matrix that can be expressed as a symbolic function of the input. However, the input of many machine learning problems is naturally given in the form of matrices or higher order tensors, in which case the second-order derivative becomes a tensor of at least fourth order. The familiar linear algebra notations and known rules for determining whether a matrix is positive semidefinite are not sufficient to deal with these higher order expressions. Here, we present a notation for tensor expressions that allows us to generalize semidefiniteness to higher order tensors. We show that for differentiable functions, the implementation of this approach is more powerful than existing solutions.

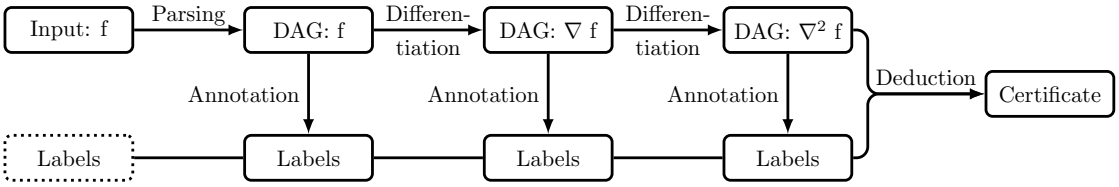

Figure 1: Overview of the symbolic Hessian approach, which is used to certify convexity. The user provides a formal expression for the input function f. The certifying algorithm parses the input f into an expression DAG and computes its gradient and Hessian by automatic symbolic differentiation. Some nodes of the expression DAGs of the input and its derivatives can be directly annotated with labels such as *diagonal*, *symmetric*, or *positive semidefinte (PSD)*. Finally, a set of deduction rules is used to propagate these labels to the root of the expression DAG of the Hessian. If the root is labeled PSD, then the deductions constitute a convexity certificate for f.

The Thuringian joint effort for research data management

Authors: Cora Assmann[1]; Jessica Rex[2]; Kevin Lang[3]; Kevin Lindt[2]; Nadine Neute[4]; Roman Gerlach[1]; Sarah Boelter[5]; Stefan Kirsch[5]

[1] *Friedrich-Schiller-Universität Jena*

[2] *Technische Universität Ilmenau*

[3] *Bauhaus Universität Weimar*

[4] *Universität Erfurt*

[5] *Ernst-Abbe-Hochschule Jena*

Corresponding Authors: roman.gerlach@uni-jena.de, cora.assmann@uni-jena.de

Handling data properly is essential for Data Scientists in order to gain informed insights and to achieve high-quality research outcomes. The field of research data management provides practices, processes and tools that ensure research data is effectively managed throughout Data Science projects.

The "Thüringer Kompetenznetzwerk Forschungsdatenmanagement" (TKFDM) is a collaborative effort of research data management helpdesks and initiatives in Thuringia to provide support to researchers of all Thuringian universities. The TKFDM conducts information and training events on all aspects of research data management, including workshops, coffee lectures and other activities addressing researchers at all levels and ranging from introductory level to specialized topics.

The poster presents TKFDM services, events and materials. This includes materials like handouts, the Research Data Scary Tales card game and the TKFDM Coffee Lecture Series. It furthermore presents the Data Steward Pilot Project and the soon to be launched Repository for Research Data in Thuringia (REFODAT). The cooperation between TKFDM and the research data management initiative of Thuringian's universities of applied sciences, the FDM-HAW Competence Cluster Jena-Erfurt-Nordhausen-Schmalkalden (FDM-HAWK) is also highlighted.

FAIR Assessment Tools: An evaluation of assessment tools of data sets according to the FAIR principles

Authors: Cora Assmann[1]; Jessica Rex[2]; Kevin Lang[3]; Nadine Neute[4]; Roman Gerlach[1]

[1] *Friedrich-Schiller-Universität Jena*

[2] *Technische Universität Ilmenau*

[3] *Bauhaus Universität Weimar*

[4] *Universität Erfurt*

Corresponding Authors: cora.assmann@uni-jena.de, roman.gerlach@uni-jena.de

Since the publication of the FAIR principles in 2016, they have become increasingly important and various tools have been developed to help assess published data with regard to compliance with the FAIR principles. There is a wide range of fair assessment tools currently available, from simple printable PDF checklists to fully automated tools that only require a DOI or URL to perform the assessment. Researchers hoping for feedback on how to optimize their own dataset with regard to the FAIR principles have different requirements than data stewards who need a quick overview of the quality of the datasets published in the repository. In order to get an orientation as to which tools are suitable for which user group and which question, we evaluated the FAIR assessment tools available in the period from July to August 2022. In our evaluation, we considered the following aspects, among others: the duration of processing, the target group of the tool, whether prior knowledge (in the field of IT and RDM) is necessary for using the tool and for understanding the results.
The poster summarizes the evaluation of the FAIR assessment tools by assigning them to four categories: Fully Configurable Tools, Automatic Tools, Improved Survey Tools, and Regular List Tools. The categorization gives users an overview and thus supports them in selecting the right tool for their needs.