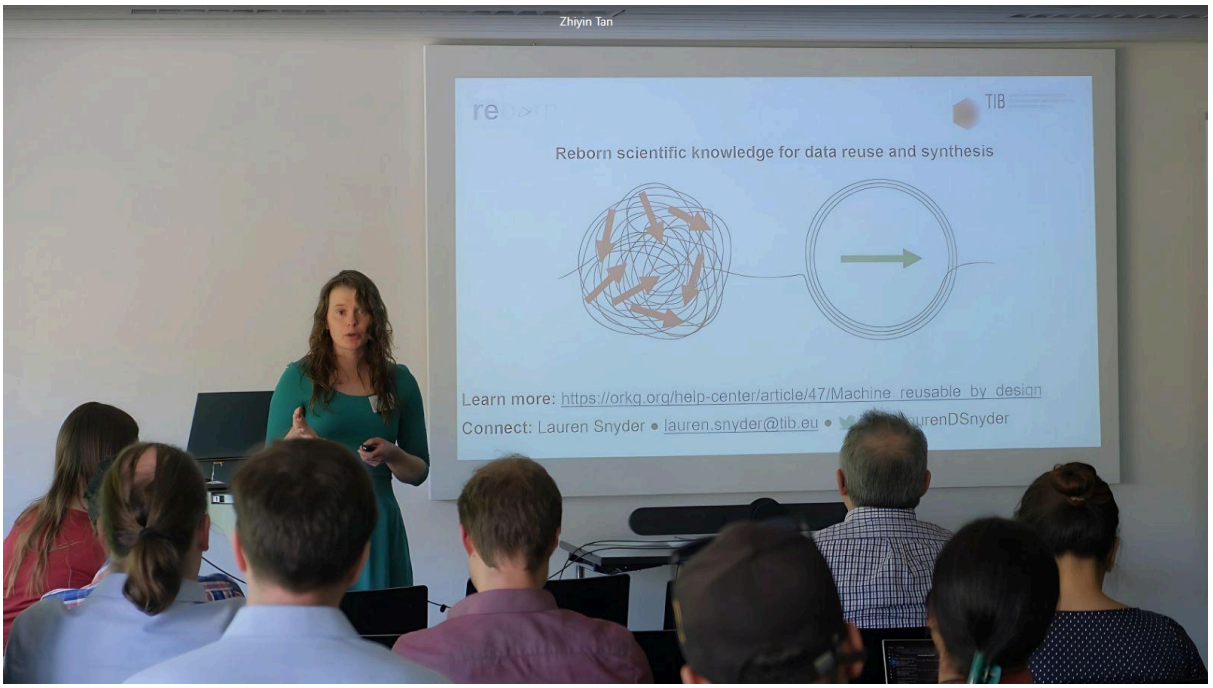
6 May 2025

PRESS RELEASE

Researchers introduce simple approach that could transform the way we produce and communicate science



Co-author Lauren Snyder presents the reborn article approach at the 2024 Research Knowledge Graph Symposium in Hannover, Germany. Photo Credit: Zhiyin Tan

Using existing data analysis tools, researchers can now make their scientific results readable by humans and machines, making them easier to reproduce and reuse

Despite significant advances in digital technologies, modern scientific results are still communicated using antiquated methods. In nearly four hundred years, scientific literature has progressed from physically printed articles to PDFs, but these electronic documents are still text-based and therefore not machine-readable. This means your computer cannot interpret the information they contain without human assistance. With millions of scientific articles published annually, the need for machine-assisted information retrieval and processing is rapidly growing. Most efforts to address this need have attempted to train machines to interpret text-based information using artificial intelligence (AI) approaches, usually with limited success.

Recently, a research team from the TIB – Leibniz Information Centre for Science and Technology proposed tackling the problem with a different mindset. Rather than trying to teach machines our language, why not produce science in a language they already understand? In an article published in Scientific Data, the team introduces reborn articles, an open-source approach that allows researchers to produce scientific findings in a machine-readable format.

Dr. Markus Stocker, first author and head of the Lab Knowledge Infrastructures at the TIB, explained, “Many scientists already use data analysis tools that produce results machines can read. But the standard way of publishing these results is to organize them in a PDF document that is not readable by machines. This means that if anyone wants to reuse these results, which is the entire point of publishing them, they first have to extract and restructure them. Wouldn’t it be more efficient if we could publish results in a way that preserves their original structure? That’s what reborn articles enables.”

The reborn articles approach works with common data analysis tools like R and Python, and allows researchers to produce results that can be easily read by both humans and machines. This means other researchers can reproduce the analyses themselves and even download reborn article data as Excel or CSV files, which are also machine readable. This may seem trivial, but the main alternatives for reusing published data are to either copy and paste individual values from PDF articles by hand, which is time-consuming and error-prone, or use AI-based tools, which are inaccurate.

Overcoming the current fixation on AI-based information extraction has been a challenge when explaining how the approach works. As co-author and TIB postdoctoral researcher Dr. Lauren Snyder noted, “AI-based extraction tools are a hot topic. It seems every field of science is looking for ways to use large language models and other extraction-related approaches. While they are powerful tools in certain situations, I wonder if fixating on them is not doing us an overall disservice. Imagine renovating your home and trying to tackle every job with drilling tools. That just doesn’t make sense. I worry this fixation on information extraction will lead us to miss opportunities to develop tools that can tackle certain tasks more efficiently. I hope our work inspires others to start thinking beyond mainstream approaches.”

Dr. Stocker added, “People have been pointing out the inefficiencies of how we produce scientific knowledge for at least a quarter century. In that time, AI-based extraction has not solved the problem and if we continue with the mindset that extraction is all we can do, mid-century we might still be struggling with the same problems. If instead we had begun using long-existing technologies to ensure scientific knowledge is produced and published machine readable, today we would have vast databases of organized knowledge. While we may be a little late to the game, any time is a good time to begin with disruptive approaches.”

More about the study:
The study “Rethinking the production and publication of machine-readable expressions of research findings” was published as an open access article in *Scientific Data*. The full article is available at: https://doi.org/10.1038/s41597-025-04905-0

The authors are Markus Stocker, Lauren Snyder, Matthew Anfuso, Oliver Ludwig, Freya Thießen, Kheir Eddine Farfar, Muhammad Haris, Allard Oelen, and Mohamad Yaser Jaradeh.

Explore reborn article data on the emerging digital library ORKG reborn: https://reborn.orkg.org/

Co-author Lauren Snyder recently produced a video describing reborn articles for the special AI/quantum category of the 2025 “Dance your PhD” competition run by AAAS and *Science*. The video describes the limitations of text-based scientific articles from a machine’s perspective and demonstrates how machine-readable scientific knowledge can make research more impactful. Watch the video and dance along here: https://youtu.be/lM0OolYmD9c?si=XM7iyzT6UF_rL-FM

For more information, please contact:
Markus Stocker, Head of the Lab Knowledge Infrastructures at the TIB, markus.stocker@tib.eu (German & English speaker)

Lauren Snyder, Postdoctoral Researcher in the Lab Knowledge Infrastructures at the TIB, lauren.snyder@tib.eu (English speaker)

The contact persons are available for live and pre-recorded television, radio, or podcast interviews.

About FAIR2Adapt
The HorizonEurope project FAIR2Adapt is funded by the European Union’s Horizon Europe programme under grant agreement No 101188256. The project started in January 2025 and has a duration of three years. The project is coordinated by Simula, with the following, participating institutions and companies: TIB – Leibniz Information Centre for Science and Technology University Library, Universidad Politécnica de Madrid (UPM), Institut Francais De Recherche Pour L’Exploration De La Mer (IFREMER), Poznan Supercomputing and Networking Center (PSNC), ALPHA Consult, Expert System IBERIA SLU (EXPERT.AI), University of Hamburg (UHAM), FCiências.ID – Associação para a Investigação e Desenvolvimento de Ciências (FC.ID), SEI Oxford Office Ltd., Plan4All ZS (P4A) and its third party Lesprojekt Sluzby SRO National and Kapodistrian University of Athens (NKUA), Nansen Environmental and Remote Sensing Center (NERSC), GO FAIR Foundation (GFF), Fresh Thoughts Consulting GmbH (FTC), adelphi research gemeinnützige GmbH, and the associated partner DeSci Labs.

For more information, please visit:

Website	LinkedIn company page	YouTube channel
Bluesky account	CORDIS	Zenodo account

Press Contacts

Alexandra Garatzogianni	**Michael Fribus**

Communication & Dissemination Lead	Communication & Dissemination Deputy
Alexandra.Garatzogianni@tib.eu TIB – Leibniz Information Centre for Science & Technology	Michael.Fribus@tib.eu TIB – Leibniz Information Centre for Science & Technology

This project has received funding from the European Union's HorizonEurope programme under the grant agreement number 101188256. Any dissemination of results here presented reflects only the consortium view. The Commission is not responsible for any use that may be made of the information it contains.