

Pressemitteilung | 05. Januar 2023

## Preprints als Informationsquelle besser nutzbar machen

Ulrike Ostrzinski  
Pressesprecherin ZB MED

Fon: +49 (0) 178 5300050  
Fax: +49 (0) 221 478 7124  
pressestelle@zbmed.de

TH Köln und ZB MED entwickeln im DFG-Projekt eine Information Extraction Pipeline für Preprints

**Im Projekt *PIXLS – Preprint Information eXtraction for Life Sciences* werden die TH Köln und ZB MED – Informationszentrum Lebenswissenschaften in den kommenden drei Jahren eine Anwendung entwickeln, die Preprint Server automatisiert erschließt. Damit kann die Forschungscommunity die aktuellen Informationen, die auf Preprint Servern veröffentlicht wurden – und daher in klassischen Nachweis- und Suchsystemen kaum auftauchen – besser nutzen. Die Deutsche Forschungsgemeinschaft (DFG) fördert das Projekt im Rahmenprogramm *e-Research Technologien*.**

Preprints stellen eine relativ neue Art dar, wissenschaftliche Ergebnisse noch vor dem Peer Review und der Veröffentlichung in einem Fachjournal für die Forschung bereitzustellen. So können aktuelle Erkenntnisse schnellstmöglich verbreitet und nachgenutzt werden. Mittlerweile gibt es verschiedene Server, die Preprints für unterschiedliche Forschungscommunitys bereitstellen. Sie unterscheiden sich technisch oder inhaltlich. *Preprints.org* ist beispielsweise multidisziplinär ausgerichtet, während *bioRxiv* ausschließlich Preprints aus den Lebenswissenschaften enthält.

Die Arbeitsgruppen aus der TH Köln – unter Leitung von Prof. Dr. Philipp Schaer – und von ZB MED – unter Leitung von Prof. Dr. Konrad Förstner – werden in *PIXLS* ein System entwickeln, das die bisher noch vernachlässigten Informationsquellen auf den Preprint Servern systematisch erschließt und durch Mehrwertdienste besser zugänglich macht. So können sowohl die Volltexte als auch die Metadaten einfacher nachgenutzt werden. Das *PIXLS*-Team setzt dabei auf eine sogenannte Information Extraction Pipeline, eine Anwendung, die aus den unstrukturierten Daten der Preprints – also Fließtexte oder Zahlen – strukturierte Informationen extrahiert. Solche strukturierten Informationen können von Datenbanken besser verarbeitet werden. Sie können dadurch auf vielfältige Weise nachgenutzt werden und sind darüber hinaus einfacher auffindbar.

Die extrahierten Daten werden im ZB MED Knowledge Environment – eine von ZB MED bereits entwickelte Datenbank – zusammengeführt und vereinheitlicht. Diese Datenbasis macht es dann möglich, Mehrwertdienste zu entwickeln und für die Forschung bereit zu stellen. Das können zum Beispiel Linked-Open-Data-Schnittstellen oder innovative Reputations- und Trendindikatoren sein. Die Daten werden auch über *LIVIVO*, den Discovery Service von ZB MED, verfügbar sein. Im

Sinne von Open Science werden sowohl die Daten als auch die Technologie der Bibliotheks- und Wissenschaftscommunity zur Nachnutzung zur Verfügung gestellt.

Ulrike Ostrzinski  
Pressesprecherin ZB MED

Fon: +49 (0) 221 478 5687  
Fax: +49 (0) 221 478 7124  
pressestelle@zbmed.de

### **Hintergrund-Informationen:**

TH Köln

Die **TH Köln** zählt zu den innovativsten Hochschulen für Angewandte Wissenschaften. Sie bietet Studierenden sowie Wissenschaftlerinnen und Wissenschaftlern aus dem In- und Ausland ein inspirierendes Lern-, Arbeits- und Forschungsumfeld in den Sozial-, Kultur-, Gesellschafts-, Ingenieur- und Naturwissenschaften. Zurzeit sind rund 25.000 Studierende in etwa 100 Bachelor- und Masterstudiengängen eingeschrieben. Die TH Köln gestaltet Soziale Innovation – mit diesem Anspruch begegnen wir den Herausforderungen der Gesellschaft. Unser interdisziplinäres Denken und Handeln, unsere regionalen, nationalen und internationalen Aktivitäten machen uns in vielen Bereichen zur geschätzten Kooperationspartnerin und Wegbereiterin.

Weitere Informationen unter [www.th-koeln.de](http://www.th-koeln.de)

ZB MED – Informationszentrum Lebenswissenschaften

**ZB MED** ist das nationale Infrastruktur- und Forschungszentrum für lebenswissenschaftliche Daten und Informationen. Unsere Dienstleistungen unterstützen die Forschung für Mensch und Umwelt: von Medizin über Biodiversität bis hin zu Umweltschutz.

Information Hub for Life Sciences

Wir stellen die Infrastrukturen für die Literatur- und Informationsversorgung, das Datenmanagement und die Handhabung großer Datenmengen bereit.

Data Sciences for Life Sciences

Wir forschen angewandt und transdisziplinär. Wir ermöglichen Datenanalysen und generieren Datenpotenziale. Als Partner der Forschung schaffen wir neues Wissen, das sich an den Bedürfnissen der Community orientiert.

Weitere Informationen unter [www.zbmed.de](http://www.zbmed.de)