

# Die SciKnow Vision:

wissenschaftliche Entdeckungen erleichtern und beschleunigen

Im Rahmen der Exzellenzinitiative des Bundes und der Länder ist an der LUH die „SciKnow – Scientific Knowledge Collider“ Clusterinitiative entstanden, die das Ziel hat, Methoden der Künstlichen Intelligenz für die Organisation und Nutzung von Daten, Informationen und Wissen in der Forschung anzuwenden. In diesem Artikel geben wir einen kurzen Überblick über die SciKnow Vision und stellen mit Wissensgraph-basierter personalisierter Medizin einen zentralen Anwendungsfall der Initiative vor.

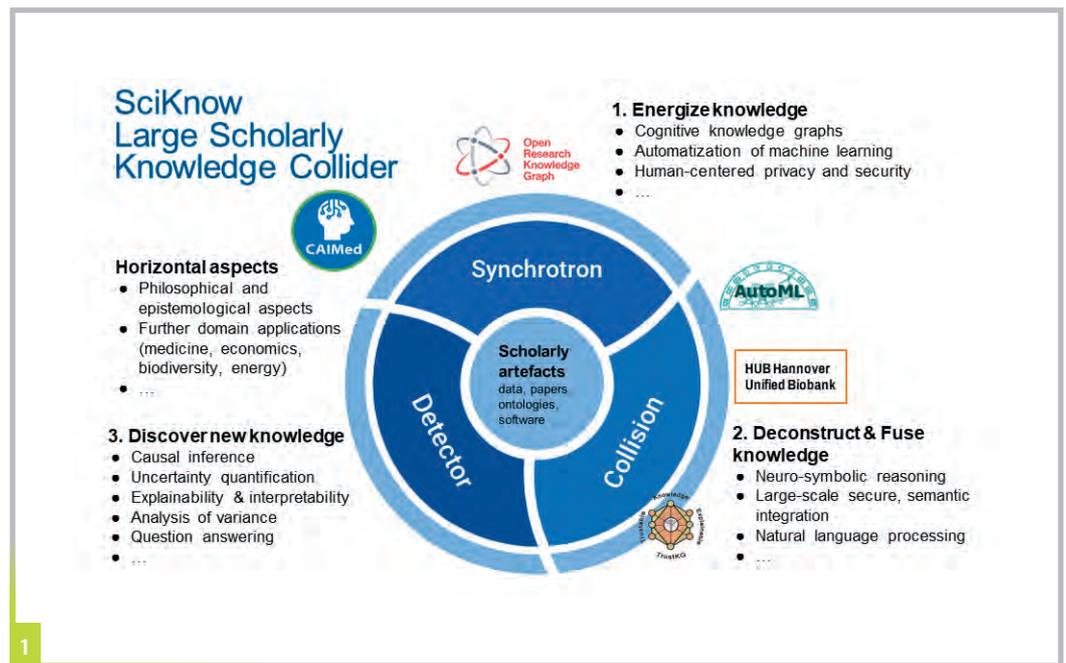


Abbildung 1  
Überblick über den Scientific Knowledge Collider mit den drei Kernkomponenten Synchrotron (Synchronisierung und Energetisierung von Wissen), Collision (Dekonstruktion und Fusion von Wissen) und Detector (Entdeckung neuen Wissens).  
Quelle: L3Seigene Darstellung

Große gesellschaftliche Herausforderungen wie Klimawandel, Kreislaufwirtschaft und Energieeffizienz, Infektionskrankheiten und personalisierte Medizin oder sichere und verantwortungsvolle digitale Dienste können nur durch die Vernetzung und Integration zahlreicher wissenschaftlicher Einzelergebnisse mit den zugrunde liegenden Forschungsdaten gelöst werden. Die derzeitigen Methoden – Veröffentlichung statischer Artikel und bestenfalls oberflächliche Speicherung relevanter Rohdaten in Repositorien – können diese Herausforderung nicht angemessen bewältigen. Schlimmer noch, sie haben zu einer Reihe

von Problemen geführt, wie zum Beispiel der Krise der Reproduzierbarkeit, der stetigen Verschlechterung des Peer-Review-Verfahrens, repetitiven und ineffizienten Forschungsprozessen und der mangelnden Interoperabilität von Forschungsdaten. Wir brauchen daher einen neuen Ansatz für den Austausch und die Nutzung wissenschaftlicher Daten und Erkenntnisse, der es uns beispielsweise ermöglicht, in akuten Krisen wie der Corona-Pandemie schneller und global vernetzt Lösungen zu entwickeln.

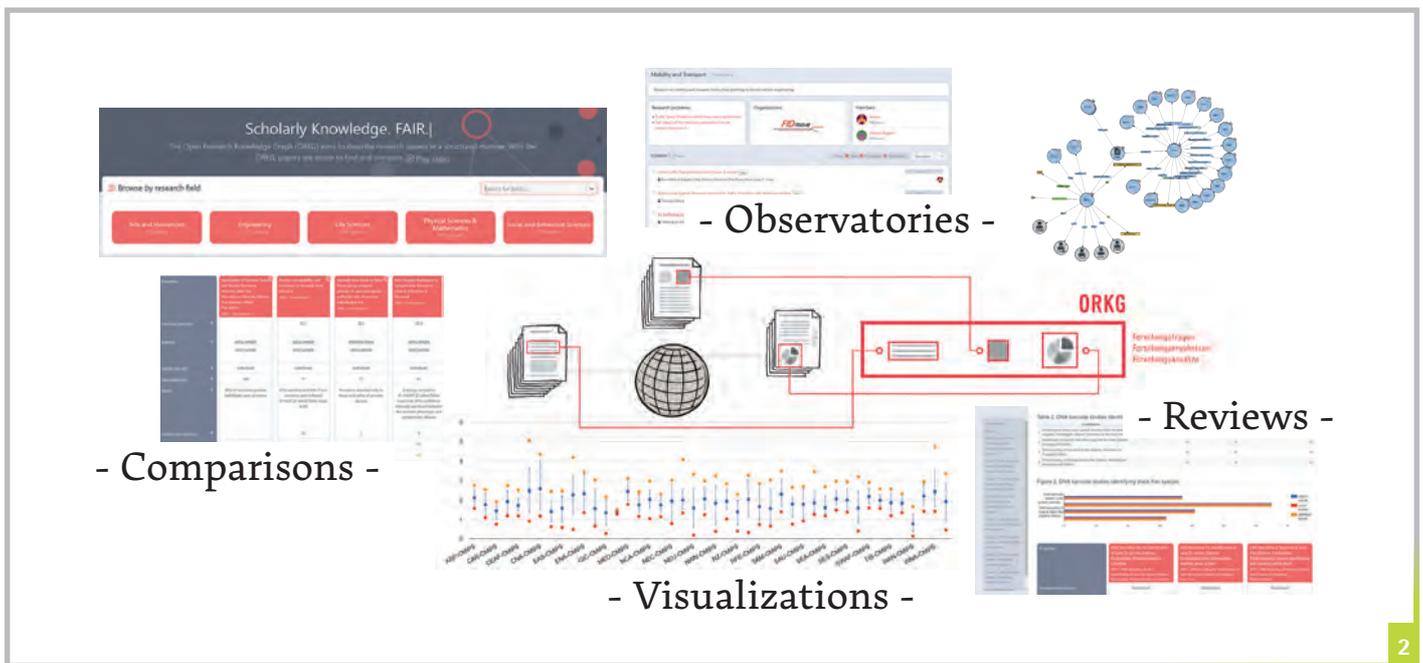
Das Ziel des Scientific Knowledge Collider ist (in Anlehnung an einen Teilchenbe-

schleuniger) die Beschleunigung und Kollision (Zerlegung, Verschmelzung, kritische Reflexion und Integration) von Forschungsergebnissen und Daten, um wissenschaftliche Entdeckungen zu erleichtern und zu beschleunigen. Der Scientific Knowledge Collider wird realisiert durch die Erforschung und Umsetzung neuer Methoden für: eine einheitliche Darstellung wissenschaftlicher Daten und Erkenntnisse (wie Ideen, Theorien, Ansätze, Bilder und Behauptungen usw.) in einer neuen Form von kognitiven Wissensgraphen – Synchronisierung und Energetisierung von Wissen

- ein neues Kooperationsmodell, das große Mengen von Forschungsdaten in kognitiven Wissensgraphen mit Hilfe von maschinellem Lernen verbindet, integriert und analysiert und synergetisch mit menschlicher und gemeinschaftlicher Intelligenz kombiniert – *Dekonstruktion und Fusion von Wissen*
- Entwicklung und Anwendung eines systematischen und ethischer Hinsicht untersucht, um Standards für die zukünftige Forschung zu setzen. Verlässlichkeit und Vertrauenswürdigkeit sind Voraussetzungen, um die Vision des Scientific Knowledge Collider als Katalysator für gesellschaftlich relevante, global vernetzte Spitzenforschung zu verwirklichen. Dies erfordert die Bündelung von Forschungsaktivitäten aus den unterschiedlichsten Diszipli-

sein. Als zentrale Anwendungsdomäne beschäftigen wir uns gemeinsam mit Partnern der Medizinischen Hochschule Hannover (MHH) mit der personalisierten Medizin sowie mit spezifischeren Anwendungen in den Bereichen Geoinformation, Ökologie, Kognitionswissenschaften und IT-Sicherheit.

Eine zentrale Herausforderung in diesem Bereich ist die



2

konzeptionellen Rahmens für die Nutzung kognitiver Wissensgraphen zur Analyse von Hypothesen auf der Grundlage von Korrelationen in Daten zusammen mit kontextuellem Wissen, um Beweise für die Feststellung von Kausalität zu liefern – *Entdeckung neuen Wissens*

Die Gewährleistung von Datenschutz und Sicherheit sensibler Forschungsdaten und Softwaresysteme einerseits und größtmöglicher Transparenz bei der Datennutzung andererseits wird als eine der großen Herausforderungen bei allen Entwicklungen sowohl in technischer als auch in philosophischer, kognitiver

nen: Künstliche Intelligenz, Datenmanagement, Knowledge Engineering, Scholarly Communication, Informationswissenschaft, Softwaresysteme, Philosophie, Wissenschaftsforschung und Softwaresicherheit.

Ein wissenschaftlich-methodischer Kern des Scientific Knowledge Colliders wird die Kombination von symbolischen (zum Beispiel semantische Repräsentationen und Wissensgraphen) und subsymbolischen Verfahren (maschinelles Lernen) – sogenannte HybridAI – und deren Integration mit menschlicher Intelligenz (zum Beispiel durch Kuratierung und Crowdsourcing)

einheitliche Darstellung von Forschungsergebnissen aus verschiedenen Formen multimodaler Repräsentationsformen und unzähligen Datenquellen (wie zum Beispiel generiert durch die NFDI German National Data Infrastructures oder EOSC European Open Science Cloud). So können beispielsweise experimentelle Ergebnisse verschiedener Publikationen über Tabellen, in Textform, unter Verwendung von Diagrammen oder Bildern kommuniziert werden. Die Extraktion der (im Text beschriebenen) experimentellen Bedingungen, die zu den Ergebnissen geführt haben, und die Zusammenführung der verschiedenen

Abbildung 2  
 Überblick über die wichtigsten ORKG-Dienste: Tabellarische Vergleiche wissenschaftlicher Informationen, Visualisierungen quantitativer Vergleichsdaten, thematische Übersichten mit ORKG-Inhalten und expertenbasierte wissenschaftliche Wissensorganisation in Observatorien.  
 Quelle: L3S/eigene Darstellung

Formen der wissenschaftlichen Kommunikation in einer gemeinsamen Darstellung ist eine Voraussetzung, um Forschungsergebnisse mit anderen Formen von Forschungsdaten (zum Beispiel klinischen oder experimentellen) zu kombinieren. Diese integrierte Sicht auf das Wissen ist von entscheidender Bedeutung, um die Implikationen kritisch zu reflektieren und neue Arbeitshypothesen für

Zweck werden die Artikel dem ORKG hinzugefügt, indem die wichtigsten Metadaten des Artikels über DOI von CrossRef abgerufen (oder manuell hinzugefügt) werden und dann der Inhalt des Forschungsartikels mit Hilfe spezieller Eingabefelder beschrieben wird. Solche strukturierten Inhalte wissenschaftlicher Beiträge sollten das behandelte Forschungsproblem, die verwendeten Materialien und

der ORKG den automatisierten Vergleich der Beiträge in einer tabellarischen Übersicht. Vergleiche unterstützen den Nutzer dabei, sich einen Überblick über den Stand der Technik zu verschaffen. Ein klassisches Beispiel aus der Informatik ist der Vergleich der Best/Worst-Case-Performance von Sortieralgorithmen oder der Precision und Recall von Algorithmen zur Objekterkennung in Bildern. Für Forscher in der Virologie und Epidemiologie ist es interessant, die Reproduktionszahlen verschiedener Viren vergleichen zu können. Solche Vergleiche bieten einen Überblick über die wichtigsten Informationen zu einem Forschungsproblem in Dutzenden oder Hunderten von Veröffentlichungen und sind somit ein wertvolles Instrument, um sich einen Überblick über den Stand der Technik in einem Bereich zu verschaffen.

Mit der engen Verzahnung von menschlicher Intelligenz (durch Kuratierung und Crowd-Sourcing), der Repräsentation des Forschungswissens in einem Wissensgraph und der Nutzung maschineller Lernverfahren für die Extraktion von Wissen aus PDF-Artikeln realisiert der ORKG bereits die zentrale Idee des Scientific Knowledge Collider. Wir sind allerdings in vieler Hinsicht noch sehr weit von der vollständigen Realisierung des Scientific Knowledge Collider entfernt: Die semantische Vernetzung und Integration vieler Artefakte des Forschungszyklus (neben Publikationen auch Forschungsdaten, Software usw.) steckt noch in den Kinderschuhen und die automatisierten Verfahren zur Extraktion von Wissen aus Publikationen erreichen noch nicht die notwendige Genauigkeit. Eine große Herausforderung insbesondere auch für medizinische Anwendungen ist der Schutz sensibler persönlicher

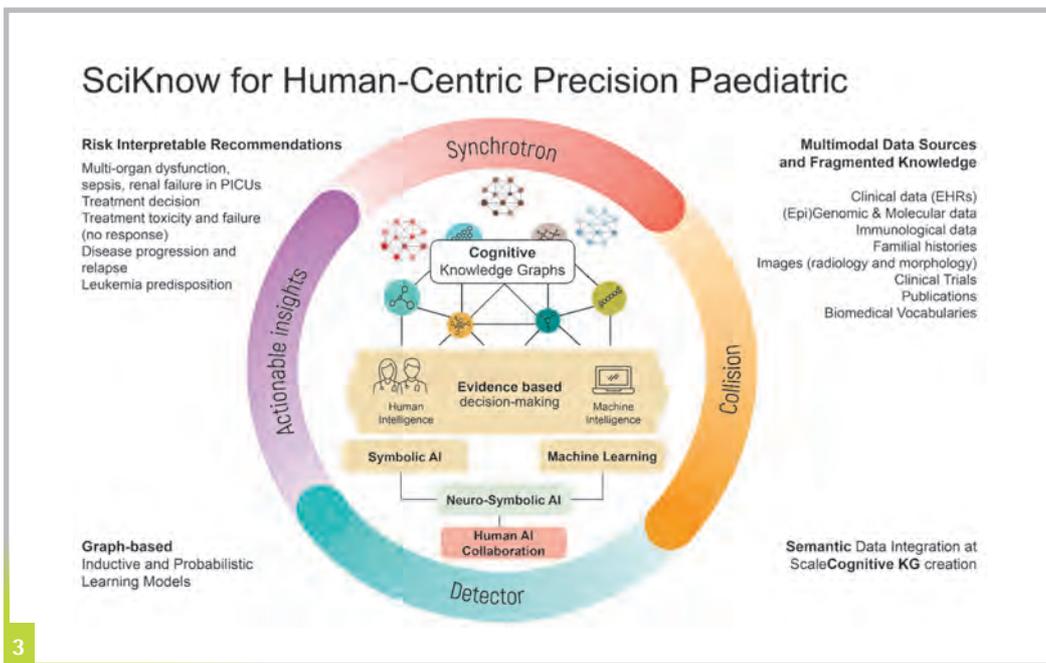


Abbildung 3  
Anwendung des Scientific Knowledge Collider Konzepts für Knowledge-Graph-basierte personalisierte Onkologie (In BioRender<sup>2</sup> gezeichnet).  
Quelle: L3S/eigene Darstellung

spezifische Anwendungen zu generieren. Dies wiederum kann nur funktionieren, wenn das gewonnene Wissen den Menschen vermittelt werden kann (Erklärbarkeit) und umgekehrt, wenn es möglich ist, menschliche Expertise sinnvoll zu integrieren.

Eine zentrale Komponente des Scientific Knowledge Collider existiert bereits: der Open Research Knowledge Graph (ORKG). Der ORKG ermöglicht eine strukturierte und semantische Beschreibung von Forschungsbeiträgen, die traditionell in unstrukturierten wissenschaftlichen Artikeln (meist publiziert als PDF) beschrieben werden. Zu diesem

Methoden sowie die erzielten Ergebnisse so beschreiben, dass der Beitrag mit anderen Artikeln, die das gleiche Forschungsproblem behandeln, vergleichbar wird. Die semantische Beschreibung folgt dem RDF-Subjekt-Prädikat-Objekt-Paradigma und kann vom Nutzer jederzeit flexibel um eigene zusätzliche Prädikate (Eigenschaften oder Attribute) erweitert werden. Eine Vorschlagsfunktion erleichtert das Auffinden und die Wiederverwendung bestehender Prädikate und Entitäten.

Wenn mehrere wissenschaftliche Beiträge zur Lösung einer Forschungsfrage im ORKG eingetragen sind ermöglicht

Daten und die Transparenz und Nachvollziehbarkeit von automatisiert generierten Empfehlungen und Analysen.

Eine zentrale Anwendung des Scientific Knowledge Colliders, die wir gemeinsam mit Kollegen der Medizinischen Hochschule Hannover bearbeiten ist die personalisierte Medizin insbesondere Onkologie. Die Onkologie stellt eine besondere Herausforderung dar, da Krebserkrankungen durch komplexe genetische und mikroökologische Wechselwirkungen zwischen Tumoren verursacht und beein-

flusst werden. Diese Besonderheiten können die Art und Weise beeinflussen, wie ein Tumor wächst und metastasiert. Um diese individuellen Krankheitsmuster zu erkennen und ein genetisches Profil einer Person zu erstellen, sind präzise Analysen personalisierter Daten (zum Beispiel OMICS, Arzneimittelreaktionen und -resistenzen) erforderlich. Aggregierte Daten aus klinischen Aufzeichnungen, Tier- und Zellmodellen, klinischen Versuchen und wissenschaftlicher Literatur bieten reichhaltige Wissensquellen zur Aufdeckung von Mustern,

um Arzneimittelreaktionen und -resistenzen besser zu verstehen und die aufgedeckten Muster zu validieren, zu erklären und zu interpretieren.

<sup>1</sup> <https://orkg.org>

<sup>2</sup> <https://app.biorender.com/>

**Sören Auer**  
**Maria Esther Vidal**  
**Bodo Rosenhahn**  
**Wolfgang Nejd**

→ Weitere Informationen  
siehe Autor\*innenseiten.

**starting**  
**BUSINESS**  
GRÜNDUNGSSERVICE DER  
LEIBNIZ UNIVERSITÄT HANNOVER



11  
102  
1004  
Leibniz  
Universität  
Hannover

[WWW.STARTING-BUSINESS.DE](http://WWW.STARTING-BUSINESS.DE)

**TRÄUMEN ODER  
MACHEN?**

JETZT EIGENES **STARTUP** GRÜNDEN  
UND FÖRDERUNG SICHERN!