

DER WISSENSCHAFTSRAT BERÄT DIE BUNDESREGIERUNG
UND DIE REGIERUNGEN DER LÄNDER IN FRAGEN
DER INHALTLICHEN UND STRUKTURELLEN ENTWICKLUNG DER
HOCHSCHULEN, DER WISSENSCHAFT UND DER FORSCHUNG.

HINTERGRUNDINFORMATION

Köln 26.10.2020

Zum Wandel in den Wissenschaften durch datenintensive Forschung

Beispiele aus verschiedenen Wissenschaftsgebieten

Neben einem Nachwort, das den Wandel in den Wissenschaften durch datenintensive Forschung anhand des Umgangs mit der COVID-19-Pandemie illustriert, diskutiert das Papier weitere Beispiele, an denen sich zeigt, wie vielschichtig dieser Wandel ist:

- _ In der **maschinellen Sprachübersetzung** wurden durch die Entwicklung neuer, lernender Methoden in Kombination mit der Nutzung großer Datenbestände und gesteigerter Rechenleistungen in den letzten Jahren deutliche Fortschritte erzielt. Dieses Beispiel steht zugleich stellvertretend für Forschungsbereiche, in denen private Unternehmen durch den Besitz großer Datenbestände und die Verfügbarkeit größerer Rechenkapazitäten den öffentlich finanzierten Forschungseinrichtungen überlegen sind.
- _ Das **Global Trade Analysis Project** (GTAP) ist ein Beispiel für eine kooperativ aufgebaute und über viele Jahre nachhaltig fortentwickelte Datensammlung. Ausgehend von Anfängen im Bereich der Agrarökonomie wird die Datensammlung inzwischen von verschiedenen Fachgebieten nachgefragt und hat sich zu einer Anlaufstelle auch für die Entwicklung von Methoden und Software entwickelt. Durch entsprechende Zugangsregelungen wird dafür gesorgt, dass Nutzerinnen und Nutzer zugleich einen Anreiz haben, zur Erweiterung und Verbesserung der Datensammlung beizutragen.
- _ Anhand eines Projekts, in dem heterogene Daten für **neuartige Städtemodelle** zusammengeführt werden, wird gezeigt, dass große Datenbestände, deren Generierung teils sehr kostenintensiv ist, inzwischen durch relativ kleine Forscherteams organisiert werden können. Diese Aggregation von Daten aus unterschiedlichsten Quellen ermöglicht zusätzliche Nutzungen innerhalb und außerhalb der Wissenschaft, die jedoch auch ethische Herausforderungen darstellen können.
- _ Das **Global Carbon Project** (GCP) illustriert, wie durch das Assessment von Daten aus verschiedenen Datenbanken eine hohe Informationsverdichtung mit weitreichenden politischen Implikationen erzielt wird. Das Fallbeispiel zeigt zugleich, welche neuen

Herausforderungen durch datenintensive Forschung für die Publikation von Daten und Software sowie deren langfristige Verfügbarkeit entstehen.

- _ Die praktischen Herausforderungen von Forschung an der Schnittstelle von Medizin und Bioinformatik werden am Beispiel **systemmedizinischer Ansätze zur Untersuchung von Lungenkrebs** demonstriert. Notwendig sind hier die Entwicklung einer disziplinenübergreifend einheitlichen Sprache und institutionenübergreifender Workflows sowie eine professionelle Beratung zu umfangreichen und sich dynamisch entwickelnden rechtlichen Vorschriften.
- _ Im Kontext der **Digitalisierung industrieller Wertschöpfungsprozesse** (Industrie 4.0) nimmt die Diskussion um den Einsatz datenbasierter Lösungen und *Big Data Analytics* einen zunehmend breiten Raum ein. In der Zusammenarbeit von Ingenieurwissenschaften und Industrie werden mit Hilfe von Methoden des maschinellen Lernens deutlich effizientere Produktentwicklungen sowie schnellere Ergebniserreichung ermöglicht.
- _ Wie neu entstandene digitale Textsammlungen datenintensive Forschung an der Schnittstelle von Computerphilologie und Computerlinguistik ermöglichen, zeigt das Beispiel der **Analyse eines großen Korpus von Hefromanen**. Methoden aus der Informatik werden dort für Analysen und Vergleiche großer Textmengen wenig beforschter Gattungen nutzbar gemacht, um ältere Forschungsergebnisse zu überprüfen und neuen Forschungsfragen nachzugehen.