

**TIB** LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK

**Verbundprojekt**

# ***TextTransfer II (Hauptprojekt)***

Methode zur korpusgestützten Prognose von  
Impactmustern in wissenschaftlichen Texten

**Abschlussbericht TIB**

nach Nr. 3.2. BNBest-BMBF 98



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

## 1. Eckdaten des Forschungsprojektes

|   |  |
|---|--|
| Vorhabenbezeichnung                         | <i>TextTransferII (Hauptprojekt)</i> – Methode zur korpusgestützten Prognose von Impactmustern in wissenschaftlichen Texten  |
| Projektart                                  | Verbundprojekt   |
| Verbundpartner                              | <ul style="list-style-type: none"> <li>▪ Leibniz-Institut für Deutsche Sprache (IDS), Mannheim<br/><i>Teilprojekt IDS: Implementierung und Praxis im Anwendungsfall Linguistik</i></li> <li>▪ Technische Informationsbibliothek (TIB) - Leibniz-Informationszentrum Technik und Naturwissenschaft, Universitätsbibliothek, Hannover<br/><i>Teilprojekt TIB: Implementierung und Praxis im Anwendungsfall Bibliothekswesen</i></li> </ul> |
| Zuwendungsempfänger und ausführende Stellen | <ul style="list-style-type: none"> <li>▪ Leibniz-Institut für Deutsche Sprache (IDS), R5, 6-13, 68161 Mannheim</li> <li>▪ Technische Informationsbibliothek (TIB) - Leibniz-Informationszentrum Technik und Naturwissenschaft, Universitätsbibliothek, Welfengarten 1B, 30167 Hannover</li> </ul>  |
| Gesamtprojektleitung                        | Prof. Dr. Andreas Witt (IDS) ( <a href="mailto:witt@ids-mannheim.de">witt@ids-mannheim.de</a> )  |
| Förderkennzeichen                           | <ul style="list-style-type: none"> <li>▪ 01IO2002A (IDS)</li> <li>▪ 01IO2002B (TIB)</li> </ul>   |
| Förderer                                    | Bundesministerium für Bildung und Forschung (BMBF)   |
| Projektträger                               | <p>01.06.2020-30.11.2023: Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR); Bereich Gesellschaft, Innovation, Technologie<br/>Heinrich-Konen-Straße 1, 53227 Bonn</p> <p>01.12.2023-31.05.2024: Projektträger Jülich; Nachhaltige Entwicklung und Innovation; Hochschulen, Innovationsstrukturen, Gesundheit I Hochschulen (HIG 5); Forschungszentrum Jülich GmbH, 52425 Jülich</p>  |

|                        |   |
|------------------------|---|
| Laufzeit des Vorhabens | 01.06.2020 – 31.05.2023 (TIB)<br>01.06.2020 – 31.05.2024 (IDS, inkl. Kostenneutraler Laufzeitverlängerung)  |
| Berichtszeitraum       | 01.06.2020 – 31.05.2024   |
| Berichtstyp            | Abschlussbericht Projektpartner TIB   |
| Stichwörter            | Deutsche Forschungsberichte, Text Mining, Maschinelles Lernen, Korpusanalyse, Computerlinguistik, Korpuslinguistik, Impact, Impact Assessment, Wissenstransfer, Forschungsimpact, Forschungsbewertung, Impact-Indikatoren, Transfer-Potenzial, IDS, TIB |
| Version                | 1.0   |
| Verfasser              | Lambert Heller (TIB)  |
| Datum                  | 27.06.2024  |

**Inhalt:**

|  |           |
|--|-----------|
| <b>1. ECKDATEN DES FORSCHUNGSPROJEKTES</b>   | <b>2</b>  |
| <b>2. KURZE DARSTELLUNG</b>  | <b>7</b>  |
| 2.1. <b>Aufgabenstellung</b>   | 7         |
| 2.2. <b>Voraussetzungen, unter denen das Vorhaben durchgeführt wurde</b>   | 8         |
| 2.3. <b>Planung und Ablauf des Vorhabens</b>   | <b>10</b> |
| 2.3.1. Planung   | 10        |
| 2.3.2. Abstimmung innerhalb von <i>TextTransfer II</i>   | 11        |
| 2.3.3. Danksagung  | 12        |
| 2.3.4. Rechte  | 13        |
| 2.4. <b>Wissenschaftlicher und technischer Stand, an den angeknüpft wurde, insbesondere</b>                                  | <b>13</b> |
| 2.4.1. - Angabe bekannter Konstruktionen, Verfahren und Schutzrechte, die für die Durchführung des Vorhabens benutzt wurden. | 13        |
| 2.4.2. - Angabe der verwendeten Fachliteratur sowie der benutzten Informations- und Dokumentationsdienste                    | 13        |
| 2.4.2.1. Fachliteratur (bezogen auf das Gesamtprojekt)   | 13        |
| 2.4.2.2. Informations- und Dokumentationsdienste   | 18        |
| 2.5. <b>Zusammenarbeit mit anderen Stellen</b>   | <b>19</b> |
| <b>3. EINGEHENDE DARSTELLUNG</b>   | <b>20</b> |
| 3.1. <b>Verwendung der Zuwendung und der erzielten Ergebnisse im Einzelnen</b>   | <b>20</b> |
| 3.1.1. AP1: Bezugsrahmen   | 20        |
| 3.1.2. AP2: Anwendungsfälle  | 22        |
| 3.1.3. AP3: Stichprobe   | 29        |
| 3.1.4. AP5: Technische und rechtliche Rahmenbedingungen  | 30        |
|  | 4         |

|             |  |           |
|-------------|--|-----------|
| 3.1.5.      | AP6: Implementierungskonzept   | 32        |
| 3.1.6.      | AP7: Kommunikationskonzept   | 34        |
| 3.1.7.      | AP8: Projektmanagement   | 37        |
| <b>3.2.</b> | <b>Die wichtigsten Positionen des zahlenmäßigen Nachweises</b>   | <b>37</b> |
| <b>3.3.</b> | <b>Notwendigkeit und Angemessenheit der geleisteten Arbeit</b>   | <b>38</b> |
| <b>3.4.</b> | <b>Voraussichtlicher Nutzen, insbesondere die Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans</b>           | <b>39</b> |
| <b>3.5.</b> | <b>Zum Zeitpunkt der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen</b> | <b>39</b> |
| <b>3.6.</b> | <b>Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 6 (BNNest-BMBF 98)</b>   | <b>40</b> |
| 3.6.1.      | Vorträge   | 40        |
| 3.6.2.      | Veranstaltung, Workshops, Kurse  | 40        |
| 3.6.3.      | Publikationen/Poster   | 40        |

## **Abbildungsverzeichnis:**

|        |   |    |
|--------|---|----|
| Abb. 1 | Systematisierung möglicher Implementierungskonzepte                       | 33 |
| Abb. 2 | Die Webpräsenz des Projektes unter „texttransfer.org“                     | 34 |
| Abb. 3 | Unter „texttransfer.org“ wird der Ansatz der Methode TextTransfer erklärt | 36 |

## **2. Kurze Darstellung**

### **2.1. Aufgabenstellung**

Forschungsergebnisse liegen zumeist als wissenschaftliche Texterzeugnisse vor, deren unüberschaubare Vielzahl in ihrem Gehalt und ihren Anwendungsmöglichkeiten oft nur zeitverzögert oder gar nicht wahrgenommen wird. Gleichzeitig wird – trotz der Investitionen öffentlicher Mittel in die Langzeitarchivierung – das in diesen Quellen vermutete Potenzial aus der Anwendung von Forschungsergebnissen bislang nicht hinreichend genutzt. Die Ausschöpfung des gesellschaftlichen oder wirtschaftlichen Potenzials aus der öffentlichen Rezeption und Anwendung von Forschungsergebnissen kann jedoch von wissenschaftlichen Publikationen mangels gemeinsamer Sprache und fehlender Kompatibilität in den zugrundeliegenden Formaten kaum erreicht werden. Der Forschung selbst fehlen trotz vorhandener Strukturen der Portfolio- und Marktanalyse bisweilen eigene Fähigkeiten und Kapazitäten, ihre meist schriftlich fixierten Forschungsergebnisse zeitökonomisch und zuverlässig auf erweiterte Anwendungspotenziale außerhalb der Wissenschaft zu untersuchen.

Entsprechend komplex gestaltet es sich, mit vertretbarem Aufwand und ohne allzu umfassende Fachkenntnis einzuschätzen, inwieweit unübersehbare Datenmengen wissenschaftlicher Texte hinsichtlich sprachlicher Hinweise auf durch bestimmte Formate oder Prozesse der Forschung (Wissens- und Technologietransfer bis hin zu wirtschaftlicher Verwertung) angestoßene Effekte einer existenziellen, gesellschaftlichen, wirtschaftlichen oder ökologischen Veränderung untersucht werden können, um das Potenzial gesellschaftlicher Investitionen in problemorientierte Forschung noch besser ausschöpfen zu können als bisher.

Klassische Auswertungsverfahren sind hierfür überfordert. Einfache Wortsuchen in mutmaßlich relevanten Texten führen zumeist durch einen zu heterogenen Sprachduktus schon innerhalb derselben Disziplin kaum zu brauchbaren Ergebnissen. Darüber hinaus eröffnen sie nur selten bzw. nur mit erheblichem Aufwand einen Sinnzusammenhang zwischen Projektgegenstand, angewandter Forschungsmethode und ggf. transferrelevantem Ergebnis. In diesem Spannungsfeld hat das Projekt

in einer ersten Förderphase (FKZ: 01IO1634 (IDS) / FKZ: 01IO1635 (TIB), im Folgenden *TextTransfer I (Pilot)*<sup>1</sup> genannt, erfolgreich den Beweis erbracht, mittels eines technischen Analyseverfahrens automatisiert in großen Berichtsdatenmengen regelhafte sprachliche Merkmale zu identifizieren, die Projekten mit hohem Transfer- und Impactpotenzial (Wahrscheinlichkeiten aufgrund textstruktureller oder transferbezogener indikatorgestützter Elemente) eigen sind. Das Hauptprojekt *TextTransfer II* griff die Ergebnisse von *TextTransfer I* auf und führte sie konsequent weiter.

Die zentrale Aufgabenstellung des Verbundprojektes TextTransfer II (Hauptprojekt) sah dabei in seiner Zielplanung die Stabilisierung der in der Pilotphase entwickelten Methode mittels Erweiterung der Datenbasis um neue Quelltypen und Domänen vor, um die Reichweite des entwickelten Transferinstruments je nach Nutzungsart skalierbar zu gestalten – mit dem Ziel, die Belastbarkeit der Methode, die im Pilotprojekt entwickelt wurde, zu testen und einer Verstetigung entgegenzuführen. Darüber hinaus wurden weiterführende Referenzindikatoren und Methoden zur automatischen Rekonstruktion von Wirkweisen rezipierten Wissens geprüft. *TextTransfer II* definierte dabei **Impact** folgendermaßen: „**Eine Auswirkung auf, eine Veränderung oder ein Nutzen für Wirtschaft, Gesellschaft, Kultur, Politik oder Recht, Technologie oder Umwelt, über den akademischen Bereich hinaus**“.

## 2.2. Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Voraussetzung für das Projekt waren die Kompetenzen, die die Projektpartner jeweils in das Projekt mit einbrachten. Für die Bearbeitung deutschsprachiger Textdaten hinsichtlich der im Projekt gewählten Fragestellung wurde erneut auf die im Pilotprojekt von TextTransfer etablierten Strukturen zurückgegriffen – mit den vorhandenen Kompetenzen im Bereich Korpusaufbau, Text Mining und maschinellem Impact Assessment.

---

<sup>1</sup> Der Abschlussbericht von TextTransfer I findet sich in der TIB Bibliothek öffentlich zugänglich unter <https://www.tib.eu/de/suchen/id/TIBKAT:1747197327/TextTransfer-Pilot-korpusgestützte-Erkennung-von?cHash=f56c7df1117392f268358ce611858ca4>

Die Aufgaben der Verbundpartner stellten sich dabei wie folgt dar:

Der Projektkoordinator **IDS**<sup>2</sup> konzentrierte sich im Rahmen von *TextTransfer II* auf die Auswahl der Stichprobe verschiedener Domänen und Quelltypen, die Vorbereitung der Daten für das maschinelle Lernverfahren (die Annotation von Textquellen), die Erarbeitung der Rahmenbedingungen, die für die Nutzung der Methode nötig sind, die Implementierung des Verfahrens zur Nutzung des Verfahrens am IDS selbst, aber auch im Hinblick auf eine (Nach-)Nutzung der Methode im Rahmen eines Kommunikationskonzeptes für weitere Anwender nach Projektende zu ermöglichen. Dem IDS oblagen überdies sämtliche projektadministrative Aufgaben.

Die Aufgabe des Projektpartners **TIB**<sup>3</sup> war es, ein bedarfsgerecht zugeschnittenes Korpus von Forschungsberichten als Stichprobe im PDF-Format zur Verfügung zu stellen. Die TIB gewährleistete außerdem die automatische Konvertierung der Stichprobe in das Zielformat txt, das zur Vorbereitung der Daten für das Maschinelle Lernen notwendig war. Neben der Entwicklung zur Anwendung der Methode im eigenen Haus arbeitete die TIB außerdem gemeinsam mit dem Projektpartner IDS an Empfehlungen zur Methodenimplementierung hinsichtlich der technischen und rechtlichen Rahmenbedingungen. Die TIB war entsprechend an den Arbeitspaketen 1, 2, 3, 5 und 6 beteiligt.

Unterstützt wurden die Verbundpartner von den Unterauftragsnehmern **Görgen & Köller GmbH** (G&K)<sup>4</sup> und **Prof. Dr. Jana Diesner** von der **School of Information Sciences / The iSchool der Universität von Illinois at Urbana-Champaign (UIUC)**<sup>5</sup> und ihrer Arbeitsgruppe. G&K unterstützte das IDS bei der Erbringung von TransfERNachweisen für die die Stichprobe bildenden Projekte bzw. sowohl beim Implementierungs- als auch Kommunikationskonzept. Die UIUC passte die in *TextTransfer I (Pilot)* zur Textanalyse notwendige Software entsprechend der neu hinzugekommenen Domänen und Quelltypen an.

---

<sup>2</sup> Vgl. u.a. <http://www1.ids-mannheim.de>

<sup>3</sup> Vgl. u.a. <https://www.tib.eu/de/>

<sup>4</sup> Vgl. u. a. <https://gk-mb.com>

<sup>5</sup> Vgl. u.a. <https://ischool.illinois.edu/people/jana-diesner>

Die Aufgabenverteilung im Verbundprojekt entsprach den Kernkompetenzen der jeweiligen Partner.

## 2.3. Planung und Ablauf des Vorhabens

### 2.3.1. Planung

Mit seinen auf Teilprojektebene angesiedelten Einzelkomponenten konnte die Vorstudie *TextTransfer (Pilot)* eine erste Basis dafür liefern, die Bedarfslücke für die maschinengestützte semantische Analyse deutschsprachiger Textdaten langfristig zu schließen. Der vergleichsweise Rückstand des deutschen Marktes in diesem Bereich ist nicht zuletzt auf personelle Engpässe qualifizierter Computerlinguistinnen/Computerlinguisten und Informatikerinnen/Informatiker im wissenschaftlichen Umfeld als auch einer Zurückhaltung in der Produktentwicklung im Bereich der Künstlichen Intelligenz auf Unternehmensseite im deutschsprachigen Raum zurückzuführen. Das Projekt wird entsprechende Fähigkeiten der semantischen Textanalyse auf Basis maschineller Lernverfahren aufbauen, sollte eine Förderphase des Hauptprojektes *TextTransfer*, das sich zum Zeitpunkt der Berichtserstellung in der Antragsphase befindet, zustande kommen.

Unter diesen Voraussetzungen gestaltete sich die Projektdurchführung des in Deutschland bisher unerprobten, prototypischen Verfahrens vornehmlich auf experimenteller Basis. Der endgültige Projektansatz zur Methodenentwicklung aus dediziertem Stichprobendesign, Informationsgewinnung, Systematik der Impact-Klassifikation sowie der technischen Umsetzung im maschinellen Lernverfahren wurden erst im Zuge des Projektfortschritts finalisiert. Entsprechende Anpassungen von der ursprünglichen antragsgemäßen Umsetzungsplanung werden im Folgenden erläutert (vgl. hierzu Kap. 3, insbesondere 3.1.3 *Inventar*).

Um diese Herausforderungen abzufedern, war bereits in der Antragsphase für *TextTransfer (Pilot)* auf IDS-Seite vorgesehen, die Stelle einer wissenschaftlichen Mitarbeiterin bzw. eines wissenschaftlichen Mitarbeiters mit Schwerpunkt Computerlinguistik (Pos. F082) personengebunden zu besetzen. Aufgrund unvorhergesehener personeller Änderungen konnte die Stelle nicht an die eingeplante Mitarbeiterin des IDS vergeben werden. In Abstimmung mit dem Projektträger konnte sie stattdessen zunächst nur übergangsweise und erst zum Oktober 2017 endgültig besetzt werden. Diese und weitere

administrative Verzögerungen auf Seiten des Fördermittelgebers bei der Freigabe von Projektmitteln für außereuropäische Unteraufträge bzw. infolgedessen entstandene Umschichtungen im Kostenplan als auch eine damit im Zusammenhang stehende erste kostenneutrale Laufzeitverlängerung des Projektes um vier Monate (Dezember 2018 bis März 2019) konnten vom IDS in Zusammenarbeit mit dem Projektträger gelöst werden. Um die Rahmenbedingungen für eine unmittelbare Anschlussfinanzierung für eine weitere Projektphase zu schaffen, im Zuge derer die Pilotstudie zu stabilisieren und auszuweiten war, wurde auf Initiative des Projektpartners IDS eine zweite, aus Eigenmitteln der Projektpartner IDS und TIB gestemmte Phase der kostenneutralen Verlängerung von neun Monaten (April bis Dezember 2019) in die Wege geleitet.

Auch beim Projektpartner TIB konnte nach einigen Verzögerungen bzw. Zwischenlösungen die zu besetzende Stelle eines Informatikers (m/w/d), die u.a. die Konvertierung der Daten in das IDS Textformat vorsah, erst Anfang 2018 final besetzt werden.

Die Zielerreichung des Verbundes im Berichtszeitraum war davon unberührt.

### **2.3.2. Abstimmung innerhalb von *TextTransfer II***

Die einzelnen Arbeits- und Projektschritte zwischen den Partnern IDS und TIB wurden im IDS koordiniert. Entsprechend wurden auch den Unterauftragnehmern einzelne Arbeits- und Projektschritte zugewiesen.

Statustreffen (virtuell oder in Präsenz) des Projektteams fanden zu folgenden Terminen statt:

- #1 21.04.2021 / virtuell,
- #2 16.06.2021 / virtuell,
- #3 19.08.2021 / virtuelles Meeting geplant, musste aber kurzfristig auf Grund der Erkrankung der Projektleitung abgesagt werden,
- #4 09.11.2021 / Präsenz-Treffen in den Räumlichkeiten des Projektpartners TIB / Hannover,
- #5 06.12.2021 / virtuell,
- #6 22.02.2022 / virtuell,
- #7 07.04.2022 / virtuell,

- #8 30.06.2022 / virtuell,
- #9 03.11.2022 / virtuell,
- #11 18.07.2023 / virtuell,
- #12 09.10.2023 / virtuell,
- #13 14.12.2023 / virtuell.

Präsenstreffen des Projektteams unter Teilnahme des Beirats fanden zu folgenden Terminen statt:

- #10 22.03.2023 / Präsenstreffen in den Räumlichkeiten der TIB, Hannover,
- #15 28.05.2024 / Projektabschlussstreffen in Präsenz in den Räumlichkeiten des IDS, Mannheim.

Zwecks Absprachen und Zusammenarbeit fanden bei Bedarf außerdem Treffen – virtuell oder in Präsenz – in kleineren Gruppen statt.

Darüber hinaus wurde der Projektstatus Quo bei für das Gesamtprojekt relevanten Entwicklungen zeitnah per Cloudlösung, im Rahmen eines gemeinsamen Rolling Documents oder per Email an alle Projektbeteiligte kommuniziert.

Zum Austausch von Daten größeren Umfangs wurde sowohl auf Gigamove als auch auf institutsinterne Cloud-Lösungen der Projektpartner TIB und IDS zurückgegriffen.

### **2.3.3. Danksagung**

Das Projekt "*TextTransfer II (Hauptprojekt) – Methode zur korpusgestützten Prognose von Impactmustern in wissenschaftlichen Texten*" wurde vom Bundesministerium für Bildung und Forschung (BMBF) unter den Förderkennzahlen 01IO2002A (IDS) und 01IO2002B (TIB) gefördert und vom Deutschen Zentrum für Luft- und Raumfahrt e.V. (DLR) vom 01.06.2020 bis zum 30.11.2023 und vom Projektträger Jülich – Nachhaltige Entwicklung und Innovation; Hochschulen, Innovationsstrukturen, Gesundheit vom 01.12.2023 bis 31.05.2024 betreut.

Unser Dank gilt neben den Förderern allen am Projekt Beteiligten, dabei insbesondere den Unterauftragnehmern G&K und UIUC, die dank ihrer Unterstützung auch nach dem offiziellen

Projektende und ihrer weiterhin sehr engagierten Mitarbeit im Rahmen der kostenneutralen Verlängerung zu den Ergebnissen des Projektes maßgeblich mit beigetragen haben.

#### **2.3.4. Rechte**

Die alleinige Verantwortung für den Inhalt dieser Publikation liegt bei den Autorinnen bzw. Autoren.

### **2.4. Wissenschaftlicher und technischer Stand, an den angeknüpft wurde, insbesondere**

#### **2.4.1. - Angabe bekannter Konstruktionen, Verfahren und Schutzrechte, die für die Durchführung des Vorhabens benutzt wurden.**

Für die Durchführung des Vorhabens wurden keine bekannten Konstruktionen, Verfahren oder Schutzrechte benutzt.

#### **2.4.2. - Angabe der verwendeten Fachliteratur sowie der benutzen Informations- und Dokumentationsdienste**

##### **2.4.2.1. Fachliteratur (bezogen auf das Gesamtprojekt)**

Aksnes, D. W., Langfeldt, L., and Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, 9(1):2158244019829575.

Barrett, D. and Leddy, S. (2008). Assessing creative media's social impact. *The Fledgling Fund*.

Becker, D. R., Harris, C. C., McLaughlin, W. J., and Nielsen, E. A. (2003). A participatory approach to social impact assessment: the interactive community forum. *Environmental Impact Assessment Review*, 23(3):367– 382.

Becker, H. A. (2001). Social impact assessment. *European Journal of Operational Research*, 128(2):311– 321.

Berendt, B. (2019). Ai for the common good?! pitfalls, challenges, and ethics pen-testing. *Paladyn, Journal of Behavioral Robotics*, 10(1):44–65.

- Blakley, J., Huang, G., Nahm, S., and Shin, H. (2016). Changing appetites & changing minds: Measuring the impact of "food, inc.". The USC Annenberg Norman Lear Center.
- Bornmann, L. and Daniel, H.-D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3):391–392.
- Bornmann, L. (2012). Measuring the societal impact of research: research is less and less assessed on scientific impact alone—we should aim to quantify the increasingly important contributions of science to society. *EMBO reports*, 13(8):673–676.
- Bornmann, L. (2013). What is societal impact of research and how can it be assessed? a literature survey. *Journal of the American Society for Information Science and Technology*, 64(2):217–233.
- Bornmann, L. (2015). Usefulness of altmetrics for measuring the broader impact of research: A case study using data from plos and f1000prime. *Aslib Journal of Information Management*, 67(3):305–319.
- Bornmann, L. (2017). Measuring impact in research evaluations: a thorough discussion of methods for, effects of and problems with impact measurements. *Higher Education*, 73(5):775–787.
- Chattoo, C. B. and Das, A. (2014). Assessing the social impact of issues-focused documentaries: Research methods & future considerations. Center for Media and Social Impact, School of Communication at American University.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.
- Clark, J. and Abrash, B. (2011). Social justice documentary: Designing for impact. Center for Social Media.
- Coates Ulrichsen, T., & Athanassopoulou, N. (2024). Commercialising Social Science Research: Insights from the University of Cambridge on key barriers, enablers and pathways to success. IfM Engage and the Policy Evidence Unit for University Commercialisation and Innovation, University of Cambridge. <https://doi.org/10.17863/CAM.107657>

- Diesner, J., Rezapour, R., and Jiang, M. (2016). Assessing public awareness of social justice documentary films based on news coverage versus social media. IConference 2016 Proceedings.
- Gabrys-Deutscher, E., and Lütjen, A. (2022). Deutsche Forschungsberichte – eine Sonderform Grauer Literatur in der Technischen Informationsbibliothek (TIB): Rückblick auf über ein halbes Jahrhundert Bestandsaufbau und Ausblick in die Zukunft. O-Bib. Das Offene Bibliotheksjournal Herausgeber VDB, 9(1), 1-13. <https://doi.org/10.5282/o-bib/5768>
- Gomes, D. and Stavropoulou, C. (2019). The impact generated by publicly and charity-funded research in the united kingdom: a systematic literature review. *Health research policy and systems*, 17(1):22.
- Gori, M. and Pucci, A. (2006). Research paper recommender systems: A random-walk based approach. In 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06), pages 778–781. IEEE.
- Greenhalgh, T., Raftery, J., Hanney, S., and Glover, M. (2016). Research impact: a narrative review. *BMC medicine*, 14(1):78.
- Grimmer, J. and Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3):267–297.
- Hennon, D., Dewaele, A., De Smet, E., Buysse, A. (2019): Guide to Impact Planning, Ghent 2019.
- Heyeres, M., Tsey, K., Yang, Y., Yan, L., and Jiang, H. (2019). The characteristics and reporting quality of research impact case studies: A systematic review. *Evaluation and program planning*, 73:10–23.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, 102(46):16569–16572.
- Holden, G., Rosenberg, G., and Barker, K. (2005). Bibliometrics: A potential decision making aid in hiring, reappointment, tenure and promotion decisions. *Social Work in Health Care*, 41(3-4):67–92.
- Latané, B. (1981). The psychology of social impact. *American psychologist*, 36(4):343.

- Loria, S., Keen, P., Honnibal, M., Yankovsky, R., Karesh, D., Dempsey, E., et al. (2013). Textblob: simplified text processing; 2018.
- Ma, Y. and Uzzi, B. (2018). Scientific prize network predicts who pushes the boundaries of science. *Proceedings of the National Academy of Sciences*, 115(50):12608– 12615.
- Ma, Y., Oliveira, D. F., Woodruff, T. K., and Uzzi, B. (2019). Women who win prizes get less money and prestige.
- Mishra, S., Fegley, B. D., Diesner, J., and Torvik, V. I. (2018). Self-citation is the hallmark of productive authors, of any gender. *PLoS one*, 13(9):e0195773.
- Otto, M., Scherer, A. (2015): *Technologietransfer in eigenständigen Organisationsformen: ein Leitfadens für die außeruniversitäre Forschung*, Potsdam: Deutsches GeoForschungsZentrum GFZ, 57 S. (DOI: <http://doi.org/10.2312/GFZ.WTT.2015.001>)
- Parker, J. and Van Teijlingen, E. (2012). The research excellence framework (ref): Assessing the impact of social work research on society. *Practice*, 24(1):41–52.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.
- Piwowar, H. (2013). Altmetrics: Value all research products. *Nature*, 493(7431):159.
- Pulido, C. M., Redondo-Sama, G., Sordé-Martí, T., and Flecha, R. (2018). Social impact in social media: A new method to evaluate the social impact of research. *PLoS one*, 13(8):e0203117.
- Rezapour, R. and Diesner, J. (2017). Classification and detection of micro-level impact of issue-focused documentary films based on reviews. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1419–1431. ACM.
- Schomaker, Rahel M.; Sitter, Alexander (2020): Die PESTEL-Analyse – Status quo und innovative Anpassungen. *Der Betriebswirt* 61(1): 9-27. Weblink:<https://www.researchgate.net/profile/Rahel->

Schomaker/publication/341876583\_Die\_PESTEL-Analyse\_-\_Status\_quo\_und\_innovative\_Anpassungen/links/5fd6285192851c13fe816690/Die-PESTEL-Analyse-Status-quo-und-innovative-Anpassungen.pdf

Shmueli, G. et al. (2010). To explain or to predict? *Statistical science*, 25(3):289–310.

Smalheiser, N. R. and Torvik, V. I. (2008). The place of literature-based discovery in contemporary scientific practice. In *Literature-based discovery*, pages 13–22. Springer.

Subramanyam, K. (1983). Bibliometric studies of research collaboration: A review. *Journal of information Science*, 6(1):33–38.

Swanson, D. R., Smalheiser, N. R., and Torvik, V. I. (2006). Ranking indirect connections in literature-based discovery: The role of medical subject headings. *Journal of the American Society for Information Science and Technology*, 57(11):1427–1439.

Taylor, M. (2013). Exploring the boundaries: How altmetrics can expand our vision of scholarly communication and social impact. *Information Standards Quarterly*, 25(2):27–32.

Tsey, K., Onnis, L.-a., Whiteside, M., McCalman, J., Williams, M., Heyeres, M., Lui, S. M. C., Klieve, H., Cadet-James, Y., Baird, L., et al. (2019). Assessing research impact: Australian research council criteria and the case of family wellbeing research. *Evaluation and program planning*, 73:176–186.

Tsey, K. (2019). Planning for and tracking research impact: Australian research council framework. In *Working on Wicked Problems*, pages 65–74. Springer.

Universität Koblenz-Landau, Zentrales Institut für Scientific Entrepreneurship & International Transfer, 15.08.2016: Wertschöpfender Wissens- und Technologietransfer außeruniversitärer Forschungseinrichtungen: Schlussbericht; FKZ 03 IO 1314 (<https://edocs.tib.eu/files/e01fb16/871694468.pdf>).

Van Raan, A. (1996). Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*, 36(3):397–420.

Van Raan, A. F. (2004). Measuring science. In Handbook of quantitative science and technology research, pages 19–50. Springer.

Vanclay, F. (2006). Principles for social impact assessment: A critical comparison between the international and us documents. *Environmental Impact Assessment Review*, 26(1):3–14.

Verbeek, A., Debackere, K., Luwel, M., and Zimmermann, E. (2002). Measuring progress and evolution in science and technology–i: The multiple uses of bibliometric indicators. *international Journal of management reviews*, 4(2):179–211.

Witt, A., Diesner, J., Steffen, D., Rezapour, R., Bopp, J., Fiedler, N., Köller, C., Raster, M., and Wockenfuß, J. (2018). Impact of scientific research beyond academia: an alternative classification schema. *Proceedings of the LREC 2018 Workshop on Computational Impact Detection from Text Data*, pages 34–39.

Wolf, B., Lindenthal, T., Szerencsits, M., Holbrook, J. B., and Heß, J. (2013). Evaluating research beyond scientific impact how to include criteria for productive interactions and impact on practice and society. *GAIA-Ecological Perspectives for Science and Society*, 22(2):104–114.

Wyndham, J., Vitullo, M., Kraska, K., Sianko, N., Carbajales, P., Nuñez-Eddy, C., and Platts, E. (2017). *Giving meaning to the right to science: A global and multidisciplinary approach*. Washington, DC: AAAS.

#### 2.4.2.2. Informations- und Dokumentationsdienste

- Förderkatalog des Bundes/Projektsuche: <https://foerderportal.bund.de/foekat/jsp/SucheAction.do?actionMode=searchmask>
- Gemeinsamer Bibliotheksverbund: <https://www.gbv.de>
- TIB Informations- und Dokumentationsdatenbank: <https://www.tib.eu/de/> und <https://www.tib.eu/de/researchieren-entdecken/sammelschwerpunkte/deutsche-forschungsberichte/>
- Gemeinsamer Bibliotheksverbund: <https://www.gbv.de>

## **2.5. Zusammenarbeit mit anderen Stellen**

Die Komplexität der Methodenentwicklung war nur durch die enge Zusammenarbeit und den jeweiligen Kernkompetenzen zwischen den Projektpartnern TIB und IDS bzw. dessen Unterauftragnehmern G&K und UIUC möglich.

### 3. Eingehende Darstellung

#### 3.1. Verwendung der Zuwendung und der erzielten Ergebnisse im Einzelnen

Bei Beendigung des Projektes waren die wissenschaftlich-technischen Ergebnisse wie bei Antragstellung geplant erreicht. Nachfolgend finden sich die Ergebnisse aus den Arbeitspaketen ausführlich dargestellt.

##### 3.1.1. AP1: Bezugsrahmen

AP1 diente der Schaffung von Voraussetzungen, um den Projektgegenstand zu vertiefen und weiterzuentwickeln mit dem Ziel, die Belastbarkeit der Methode zu testen und einer Verstetigung entgegenzuführen.

Die Forschungsfrage des Vorhabens musste basierend auf den Ergebnissen aus *TextTransfer I* unter Berücksichtigung der erweiterten Arbeitsfelder mit zum Teil – im Vergleich zum Pilotprojekt – neuen Inhalten spezifiziert, dokumentiert und Handlungsfelder abgeleitet werden. Zu diesem Zweck wurden Parameter zur Erweiterung bestehender Datensammlungen und zur Identifikation neuer relevanter Domänen aufgestellt.

Der Untersuchungsgegenstand „Forschungswissen“ wurde in der Pilotphase primär zum Funktionsnachweis anhand von Projektendberichten fixiert. Im Sinne einer umseitigen Stabilisierung der Methode, die mittels des automatisierten Abgleichs von Forschungstexten und klassifizierenden Metainformationen auf die Prognose von Impactwahrscheinlichkeiten zielt, waren im Hauptprojekt in den Arbeitspaketen 2 und 3 alternative Textformate auf ihre Nutzbarkeit für die Methode zu eruieren. In der Folge war dann ein breiteres Arsenal an Referenzdaten für das Maschinelle Lernen zu generieren, anhand dessen unterschiedliche Formen der Rezeption und des Impacts von Forschung abzulesen sein sollte. Hierfür sollten zusätzliche mögliche Quelltypen – neben dem der Projektberichte aus dem Pilotprojekt –, die sich für die Anwendung der Methode eignen, identifiziert bzw. evaluiert werden. Um die Methoden künftig auch auf größeren und mit Blick auf das Spektrum der Domänen heterogenen Datenmengen zu testen, sollte außerdem u.a. eine weitere Domäne für AP 2 Anwendungsfälle – neben den Domänen Mobilität, Künstlichen Intelligenz und Germanistischen

20

Linguistik – aus dem Bereich der Geisteswissenschaften, konkret die der Musikwissenschaften, herangezogen werden. Als dritter Baustein neben Domäne und Quelltyp, der die Erweiterung der Methode *TextTransfer* betrifft, kommt dem Anwender eine entscheidende Rolle zu. Im Verlauf des Projektes wurde daher eine Fokussierung auf potenzielle Anwender als am Transfer interessierte oder für den Transfer zuständige Personen der Methode und damit der Entwicklung eines Instruments des Wissens- und Technologietransfers vorgenommen: Neben den beteiligten Instituten selbst und den im Bereich des Transfers verankerten Anwenderpositionen (u.a. Direktion, Transferbeauftragten, Öffentlichkeitsarbeit) aus unterschiedlichsten wissenschaftlichen Disziplinen sollte die Methode perspektivisch für weitere Anwendersegmente auch außerhalb der Wissenschaft geöffnet werden. Die Methode sollte damit innerhalb der beteiligten Institutionen fest verankert werden, aber auch für Dritte, an Impact- und Transferauswertungen und -aktivitäten Interessierten nutzbar gemacht werden. Details zum Thema Auswahl der Domänen und Quelltypen vgl. im Folgenden.

Um zusätzliche Möglichkeiten der Erhebung weiterer Daten zu generieren und die Datenbasis, mit der die Methode aus *TextTransfer I* weiter optimiert werden konnte, zu erhöhen, wurden im Rahmen von AP 1 außerdem zwei Aufgabenbereiche identifiziert, die sich auf Grund aktueller Entwicklungen bei den Projektpartnern zu Projektbeginn als mögliche Hebel hinsichtlich Ressourcenumverteilung zu erweisen schienen, die aber zum Zeitpunkt der Antragsstellung noch nicht abzusehen waren.

Zum einen handelte es sich um den Bereich der **Referenzdatenerhebung**, wie sie in Fehler: Verweis nicht gefunden vorgesehen war. Referenzdaten dienen der Klassifizierung des Impacts in Quelltexten für das maschinelle Lernverfahren. In *TextTransfe I* wurden zur Erhebung der Referenzdatenbasis vornehmlich Interviews mit den Autoren des Quelltyps Projektbericht durchgeführt, um klassifizierende Aussagen zum Impact einzelner Projekte zu gewinnen. Der Aufwand dieses Ansatzes entpuppte sich jedoch im Laufe des Pilotprojekts als für die praktische Umsetzung und Nachnutzung bei potenziellen Anwendern als zu groß. Dieser Ansatz wurde daher durch die Adaption eines Online-Umfragetools im laufenden Berichtsjahr prozessoptimiert. Durch diesen Ansatz war es möglich, sowohl eine deutlich größere Anzahl von Aussagen zum Impactpotenzial zu erhalten, als auch zusätzliche Gruppen von Probanden, die für die Untersuchung kontaktiert werden sollten, zu erreichen.

Eine weitere Prozessanpassung war beim Thema Vor-/Aufbereitung der Daten für das Maschinelle Lernverfahren, ebenfalls Teil von AP 3 Stichprobe, im Bereich der **Konvertierung** vorgesehen. Anstelle der sehr aufwendigen Konvertierung der im PDF-Format vorliegenden Daten in die Markup-Sprache TEI XML (i5) war angedacht, auf die zwischenzeitlich existierende Möglichkeit des Projektpartners TIB einer Automatisierung der Umwandlung von PDFs in Textdaten zurückzugreifen.

AP1 war zu Projektende planmäßig abgeschlossen.

### 3.1.2. AP2: Anwendungsfälle

Der vorliegende Entwicklungsstand der Methode aus *TextTransfer I* gestattet linguistische Analyseverfahren einer auf den Quelltypus Projektbericht und die Domäne Mobilität zu Testzwecken eingegrenzten Datenbasis. Das Arbeitspaket zielte auf die Schaffung einer erweiterten Datengrundgesamtheit für das maschinelle Lernverfahren. Die Funktionalität der Methode sollte hierbei durch Hinzuziehung neuer Quelltypen – und somit neuer Formate – als auch neuer Domänen – und somit thematisch unterschiedlichen Texten – erprobt und erweitert werden.

#### Domänenerweiterung für den Quelltyp Projektberichte

Das Transparenzgebot für die Verwendung steuerlicher Mittel legt öffentlich geförderter Forschung die Verpflichtung auf, die Ergebnisse von wissenschaftlicher Forschung in Projektabschlussberichten darzustellen. Auch hinsichtlich ihrer öffentlichen Verfügbarkeit bei der als Datengeber kooperierenden TIB und aufgrund ihrer im Vergleich zu anderen Formaten wissenschaftlicher Publikation relativen strukturellen und sprachlichen Homogenität (Berichtsduktus, Standardaufbau, dedizierte Transferabschnitte) wurde diese Textgattung für ein empirisch gestütztes und maschinell getriebenes Auswertungsverfahren des Textmining als besonders geeigneter Ausgangspunkt für den Funktionsnachweis der Methode für das Pilotprojekt *TextTransfer I* ausgesucht, die auch im Hauptprojekt zwecks Vertiefung der Methode erneut betrachtet wurden.

Als erweiterter Bezugsrahmen wurden vom Verbundpartner IDS entsprechend Forschungsprojekte im Berichtszeitraum ausgewählt,

- die aus öffentlich geförderten Verbundprojekten stammen,
- deren Projektpartner sowohl aus öffentlicher Forschung als auch der freien Wirtschaft in Deutschland kommen,
- deren deutschsprachige Projektabschlussberichte in der TIB frei zugänglich in digitaler Form vorliegen, wobei ein Projekt einen Abschlussbericht stellvertretend für das Gesamtprojekt und/oder Einzelberichte pro Projektpartner auszuweisen hat.

### **„Alt“-Domäne Mobilität**

Projekte der Domäne Mobilität, die von unterschiedlichen Projektträgern gefördert und die alle in der TIB unter dem Recherchebegriff „Mobilität“ geführt werden – womit eine quantitative Erweiterung zum Pilotprojekt vorgenommen wurde –, wurden im Verlauf des Projektes erneut – zwecks Abschluss bzw. Vertiefung des Ansatzes aus *TextTransfer I* – genauer betrachtet. Die Domäne wurde im Vorprojekt aufgrund hoher gesellschaftlicher Relevanz und potenziell hohem Impactpotenzial beibehalten. Die Erfahrungen aus *TextTransfer I* hatten die Eignung der Domäne als Evaluierungsgegenstand bestätigt. Eine große Datenbasis, wie sie mit der Domäne Mobilität im Vergleich zu *TextTransfer I* vorliegt, diene außerdem der Erhöhung der Skalierbarkeit der Methode

### **Neu-Domänen**

Für den Quelltyp der Projektberichte sollten zwecks Aufbau einer neuen Datenbasis weitere Domänen aus dem Bestand des Projektpartners TIB betrachtet werden, von denen einerseits entlang des bisherigen Suchfilters eine hohe Impactwahrscheinlichkeit zu erwarten war, die aber auch vor allem eine Bedeutung für die Anwendung im Rahmen des Impact- und Transferaktivitäten der beteiligten Partner haben sollten. Zudem sollte im Sinne eines breit aufgestellten Transferverständnisses und eines ganzheitlichen Transferauftrages über alle Disziplinen hinweg eine weitere Domäne aus dem Bereich Geisteswissenschaften herangezogen werden. Entsprechend den Vorgaben wurde folgende weiteren Domänen hinzugezogen:

### **Domäne Künstliche Intelligenz**

Die Domäne Künstliche Intelligenz wurde als Thema von hoher gesellschaftlicher Relevanz ausgewählt. Das Thema war zudem Gegenstand des Wissenschaftsjahres 2019<sup>6</sup>. Im Hinblick auf ihr Impact- und Transferpotenzial war sie für die methodische Ausrichtung des Projektpartners IDS von zentraler Bedeutung: Das Projekt wurde aus der Abteilung Digitale Sprachwissenschaft des IDS geleitet und bediente sich bei der Entwicklung der Methode der Technik des maschinellen Lernverfahrens, einem Teilgebiet der Künstlichen Intelligenz.

### **Domäne Germanistische Linguistik**

Um die Methode auf eine Domäne im Kern-Forschungsinteresse des Projektpartners IDS als sprachwissenschaftliche Einrichtung zu erweitern und gleichzeitig das Thema Transfer- und Impactpotenzial am IDS selbst zu unterstützen, wurde als weitere neue Domäne die der Linguistik ausgewählt.

### **Domäne Musikwissenschaften**

Neben den bereits genannten Domänen sollte eine weitere Domäne aus dem Bereich der Geisteswissenschaften herangezogen werden. Die Wahl fiel aus zwei wesentlichen Gründen auf die Domäne Musikwissenschaften: Einerseits sollte ein zusätzliches thematisches Feld aus dem Bereichen Geistes- und Kulturwissenschaften erschlossen werden, um zu eruieren, welche Ergebnisse ein maschinelles Lernverfahren generiert, das sich bei der Stichprobenanalyse nicht auf klassische Indikatoren und Formate des Transfers stützen kann. Dabei erschien es interessant, um den Bestand an geeigneten Referenzdaten zu erweitern. Eruiert werden sollte die Hypothese, inwiefern kulturelles Forschungswissen in stärker öffentlichen und für die Erhebung besser zugänglichen Formaten wie beispielsweise Zeitungstexten rezipiert wird. Idealerweise sollte sich hier auch im Projektverlauf ein weiterer Quelltyp ergeben. Andererseits ist der Fachbereich im Rahmen der Nationalen Forschungsdateninfrastruktur (NFDI)-Initiative der Deutschen Forschungsgemeinschaft sehr aktiv in

---

<sup>6</sup> <https://www.wissenschaftsjahr.de/2019/>

der Etablierung umfassender Datenbestände. Zusammengenommen sollte dies dem Vorhaben gute Rahmenbedingungen zur Verifizierbarkeit der Methodenfunktionalitäten bei gleichzeitig guten Voraussetzungen für die Maschinenlesbarkeit ermöglichen. Als Partner sollte das Musikwissenschaftliche Seminar der Universität Detmold/Paderborn, das über Kooperationen im Rahmen der NFDI – namentlich der beiden Verbünde Text+ sowie NFDI4Culture –, mit dem Projektpartner IDS verbunden ist, gewonnen werden. Vertiefte Gespräche und Briefings zum Ansatz des Projektes *TextTransfer* mit dem Musikwissenschaftlichen Seminar der Universität Detmold/Paderborn als auch erste Recherchen zum Thema Anwendungsfälle – speziell der Recherche einer Grundgesamtheit der Domäne Musikwissenschaften als auch nach einem potenziellen neuen Quelltyp – durch das Musikwissenschaftliche Seminar der Universität Detmold/Paderborn ergaben, dass für die Domäne Musikwissenschaften weder im Bestand des Seminars selbst als auch erweitert bei spezifischen Partnereinrichtungen der Universität Detmold/Paderborn eine nennenswerte Grundgesamtheit des Quelltyps Projektberichte identifiziert werden konnte, die nicht bereits auch im Bestand des Projektpartners TIB nach Durchlauf einer ersten Recherche zum Thema Musikwissenschaften vorhanden gewesen wäre.

Dem potenziellen Kooperationspartner wurde daher die Identifizierung eines neuen Quelltyps (etwa monografische Schriften) und eines passenden Referenzdatentyps (bspw. Zeitungstexte) vorgeschlagen.

Im Zuge der engeren Abstimmung dieses Vorgehens war es jedoch aufgrund definitorischer und forschungskultureller Hürden nicht zu einem Kooperationsvertragsabschluss gekommen.

Umso mehr zeigte sich, wie wichtig es bei zunehmender Fokussierung auf Forschungstransfer ist, auf dem Weg der Entwicklung von Mechanismen zur Qualitätssicherung von Zielvereinbarungen auch bei

Einrichtungen jenseits des Paktes für Forschung und Innovation als Rückgrat der deutschen Forschungslandschaft<sup>7</sup> künftig auf eine flankierende Strategie zur Akzeptanzwahrung zu achten.

Um diese Domäne dennoch als Stichprobe dem Maschinellen Lernverfahren zuzuführen, wurde letztendlich erneut auf den Bestand der TIB zurückgegriffen.

Der Projektpartner IDS hatte für die Etablierung der Grundgesamtheiten einen kriteriengestützten Suchfilter pro Domäne und innerhalb einer Domäne nach Projekten mit ihren jeweiligen Berichten mit den folgenden Kriterien eingerichtet, die der Projektpartner TIB auf seinen Bestand anwandte: Verbundname Gesamtprojekt, Akronym, Verbundnummer, Name Teilprojekt, Name und Kontaktdaten Forschungseinrichtung/Zuwendungsempfänger, Projektleiter, Fördernummer, Projektträger, Projektlaufzeit, Öffentliche Verfügbarkeit des Berichts, Digitale Verfügbarkeit des Berichts im TIB-Bestand, Berichtstypus (Konsortial-, Gesamtbericht/Individualbericht), Seitenzahl, Link zu PDF in der TIB-Bibliothek. Die jeweilige Grundgesamtheit pro Domäne wurde in Form einer Exceltabelle an das IDS zwecks finaler Sichtung der Zusammenstellung mit Blick auf die finale Eignung eines Datensatzes unter Berücksichtigung der Kriterien des erweiterten Bezugsrahmens für die Stichproben aus AP 3 übergeben.

### Quelltyperweiterung

Neben dem Quelltyp Projektendberichte sollten weitere Quelltypen hinsichtlich des erweiterten Transferbegriffs als Lerndaten evaluiert werden. Hierzu wurden folgende Ansätze verfolgt:

Ein erster Ansatz, der im Austausch mit dem Projektträger DLR entstand, bezog sich auf die vom Projektträger sekundär erstellten, öffentlich gemachten Projektsteckbriefe, die übersichtsartige Angaben zu Anträgen von Gesamtprojekten und Verwertungsplänen enthalten. Hier sollte überprüft werden, inwieweit dieser Quelltyp dem in *TextTransfer II* gewählten methodischen Ansatz genüge. Die Recherche ergab jedoch, dass der Quelltyp Projektskizze nicht öffentlich zur Verfügung steht und die

---

<sup>7</sup> Vgl. Koalitionsvertrag 2021-2025, S. 22 ([https://www.fdp.de/sites/default/files/2021-11/Koalitionsvertrag%202021-2025\\_0.pdf](https://www.fdp.de/sites/default/files/2021-11/Koalitionsvertrag%202021-2025_0.pdf))

Archivierungspraxis einzelner Projektträger stark voneinander abweicht. Ein neuer Quelltyp ließ sich für die Gattung öffentliche Projektsteckbriefe abschließend nicht identifizieren.

Ein weiterer Ansatz, der im Berichtszeitraum verfolgt wurde, war der der Prüfung von Forschungsanträgen als möglicher zusätzlicher Quelltyp. Dabei sollten Projektanträge zur Stichprobe der Domäne Mobilität aus *TextTransfer I* herangezogen werden, da für diese bereits das Impactpotenzial bekannt ist. Als besondere wissenschaftliche Herausforderung stellten sich hierbei die evolutionäre Distanz von Wissenstand und Kapazität der Projektbeteiligten in der Antragsphase hin zum endgültigen Projektergebnis dar. Auf der Eignung von Antragstexten zur Impactprognose unter Einbeziehung eines möglichen analytischen Zwischenschritts der Ergebnisanalyse musste damit das Hauptaugenmerk dieses Ansatzes gerichtet werden. In der Praxis gestaltete sich jedoch der Aufwand der Einholung einer Nutzungsgenehmigung dieses nicht-öffentlichen Quelltyps sowie die Eruiierung von rechtsstabilen Ansprechpersonen als noch aufwendiger und sensibler als die Recherche von externen Referenzdaten nach dem deduktiven Ansatz aus dem Evaluierungsprojekt *TextTransfer I*. Eine Reproduktion dieses Vorgehens sowohl innerhalb des vorliegenden Projektplans von *TextTransfer* selbst als auch bei der angestrebten Nachnutzung der Methode im Forschungsalltag erschien daher kapazitär nicht vertretbar. Das Thema wurde somit nicht weiterverfolgt.

Wie bereits im Vorangegangenen zum Thema Neudomäne Musikwissenschaften beschrieben wurde ein weiterer Ansatz – die Recherche eines weiteren Quelltyps im Zusammenhang mit der Domäne Musikwissenschaften – aufgrund des Nichtzustandekommens einer Kooperationsvereinbarung mit dem Musikwissenschaftlichen Seminar der Universität Detmold/Paderborn nicht weiter vorangetrieben.

Neben den oben genannten Ansätzen wurde außerdem die Idee eruiert, einen neuen Quelltyp mit Hilfe von Informationen aus dem Internet zu generieren:

Eine ganze Kategorie möglicher Datenquellen sind dabei öffentlich zugängliche Online-Quellen – darunter z.B. verschiedenartige Websites, Presseerklärungen sowie Social-Media-Präsenzen von Forschungseinrichtungen, aber auch journalistische Angebote im Web. Allein die Erwähnung eines

Forschungsprojekts sowie deren Urheber, Häufigkeit, Kontexte und Zeiträume der jeweiligen Erwähnungen könnten wertvolle ergänzende Hinweise im Mosaik von Mustern der Verwertbarkeit offenbaren. Einzelne, aktuelle Quellen dieser Art lassen sich durch Linksammlungen, Websuchmaschinen und Website-interne Suchfunktionen finden und prüfen. Schwieriger würde es sein, eine große Anzahl ausgewählter Websites mit reproduzierbarer Umfassendheit und Präzision auf Stichwortkombinationen hin zu untersuchen, zumal wenn es gelten würde, umfassende Sammlungen heruntergeladener Web-Dokumente aus der Vergangenheit zu berücksichtigen. In diesem Zusammenhang hätte es daher gegolten, in Kooperation mit umfassenden Web-Archiven<sup>8</sup> Datenquellen zu erschließen. Web-Archive erlauben dabei keine Stichwortsuche. Mit der Technik des fokussierten Crawlens wäre es allerdings möglich, bedarfsgerecht zugeschnittene aktuelle oder historische Kollektionen von Web-Dokumenten erstellen zu lassen, die dann lokal auf Rechnern von Mitarbeiterinnen und Mitarbeitern des Projekts analysiert werden könnten. Hierbei wäre die Herausforderung die Bewältigung des Umgangs mit verhältnismäßig großen Datenmengen, Suchkriterien zu entwickeln, mit denen sich den Kollektionen der Web-Dokumente die erforderlichen Informationen entziehen lassen, sowie die Ergebnisdaten schließlich in einem Modell zu erfassen, das Überblick und Vergleichbarkeit zu gewinnen helfen würde.

Nach erster Prüfung wurde dieser Ansatz einer neuen Quelltyperschließung mittels Daten aus Web-Archiven als interessant und durchführbar befunden. Es wurde dem Projektträger DLR entsprechend vorgeschlagen, angesichts der zu erwartenden hohen Aufwände für diesen lohnenden Ansatz einen systematischen Suchdienst für alternative, Impact-relevante Daten aus dem Web-Archive in seine künftigen strategischen Überlegungen für weiterführende Forschungsprogramme aufzunehmen.

AP2 war zu Projektende planmäßig abgeschlossen.

---

<sup>8</sup> u.a. dem Unternehmen Internet Archive (<https://archive.org>)

### 3.1.3. AP3: Stichprobe

Der methodische Ansatz des Verbundes sah die Generierung von geeigneten Stichproben vor, die als Bezugsbasis für das maschinelle Lernen dienen sollten. Hierfür griffen für den Quelltyp Projektberichte die Prozesse der physischen Aufbereitung (Strukturierung und Konvertierung) durch den Projektpartner TIB sowie der informationellen (Auswahl, Annotation) Aufbereitung durch den Projektpartner IDS ineinander.

Mit den im 2. Projektjahr in AP 2 Anwendungsfälle am Vorgehen im Pilotprojekt orientierten Eingrenzungen konnten dem Projekt für die genannten Domänen folgende Stichproben aus dem TIB-Bestand an das IDS zwecks weiterer Bearbeitung übergeben werden:

- Domäne Mobilität (Mob): 239 Projekte mit 906 Einzelberichten<sup>9</sup>
- Domäne Künstliche Intelligenz (KI): 116 Projekte mit 497 Einzelberichten
- Domäne Linguistik (Ling): 40 Projekte mit insgesamt 89
- Domäne Musikwissenschaften (MuWi): 25 Projekte mit insgesamt 58 Einzelberichten<sup>10</sup>

#### Prozessoptimierungen im AP

Im Zuge der anstehenden Arbeiten im AP konnten im 1. Projektjahr (2020) zwei Bereiche identifiziert werden, die auf Grund der seit Antragsstellung erfolgten Entwicklungen bezüglich des Prozessmanagements bei den Projektpartnern TIB und IDS möglich sein sollten und die sich im 2. Projektjahr (2021) auch bestätigt hatten:

#### Anpassung Konvertierungsprozess

Sobald die Stichproben definiert waren, sah der weitere Prozess die Vor- und Aufbereitung der zugrundeliegenden Daten für das maschinelle Lernen vor. Bei der Quellenbasis der TIB für den Quelltyp

---

<sup>9</sup> bestehend aus 148 Projekte mit insgesamt 515 Einzelberichten aus TextTransfer II, zzgl. der Stichprobe aus *TextTransfer I* mit 91 Projekten mit insgesamt 391 Einzelprojekten

<sup>10</sup> weitere Details hierzu vgl. TextTransfer II Jahresbericht 2021, Kap. 2.1.3 AP 3 Stichprobe

Projektberichte handelte es sich um PDF-Dokumente, die für die weitere Bearbeitung in ein maschinenlesbares Format, u.a. in das txt-Datenformat, gebracht werden mussten. Der für *TextTransfer II* herangezogene Konvertierungsprozess wurde konsequent aus den Erfahrungen des Pilotprojekts entwickelt.

Der Projektpartner TIB war, in Abweichung zur Antragsstellung, zu Projektbeginn von *TextTransfer II* in der Lage, PDFs automatisiert in das txt-Datenformat umzuwandeln, so dass die Daten anhand eines eigens dafür vom IDS entwickelten Workflows für die weitere Bearbeitung im Rahmen von Fehler: Verweis nicht gefunden durch den Projektpartner IDS aufbereitet werden konnten.

AP3 war zu Projektende planmäßig abgeschlossen.

#### **3.1.4. AP5: Technische und rechtliche Rahmenbedingungen**

Aus den Vorarbeiten von *TextTransfer I (Pilot)* ergaben sich Bedarfe insbesondere hinsichtlich technischer und rechtlicher Rahmenbedingungen, um die Anwendung maschineller Lernverfahren für textuelle Stichproben, aber auch im Speziellen bzgl. der Nutzung der Methode TextTransfer zu optimieren bzw. zu verstetigen.

#### **Standardisiertes Abgabeformat**

Im Pilotprojekt wurde der Bedarf nach einer optimierten Vorbereitung insbesondere des Quelltyps Projektbericht durch ein standardisiertes Abgabeformat identifiziert, um die erheblichen Aufwände einer für Maschinenlesbarkeit notwendigen Datenkonvertierung künftig zu umgehen. An der TIB wurde entsprechend antragsgemäß ein prototypisches Tool entwickelt, mithilfe dessen Projektendberichte – der IDS-Abschlussbericht von *TextTransfer I* fungierte dabei als Beispiel für den Quelltyp Projektbericht– nach offenen, ausgereiften Dokumentenstandards (wie z.B. Markdown und XHTML) erstellt und in einem maschinenlesbaren Format eingereicht werden können. *TextTransfer II* hat sich hierbei, soweit möglich und vorhanden, nach den Vorlagen der Fördermittelgeber richten. Die

TIB hat hierzu die bereits genutzte Publishing Pipeline Fidus Writer auf der Plattform Github verwendet und angepasst<sup>11</sup>, um einen Texteditor für strukturierten Text mit Funktionen zur Erstellung von Templates zur kollaborativen Bearbeitung von Berichten sowie zum Speichern von Überarbeitungen und Metadaten zu entwickeln. Bezugnehmend auf Anforderungen aus *TextTransfer II* hatten die Entwickler von Fidus Writer bereits fehlende Features in der Software nachgerüstet<sup>12</sup>. Künftig ist auch eine automatisierte Einreichung in das DSpace-Repository der TIB vorgesehen. Vorbereitend hierfür hat die TIB bereits ein neues Dspace-Repository unter <https://oa.tib.eu/renate/> aufgesetzt; ein entsprechender Submission-Workflow wurde hierfür entwickelt. Für single-sign-on-Verfahren wurden mögliche Lösungen wie Github (kommerziell), DFN AAI oder Overleaf evaluiert.

Ferner wurde an der TIB ProposalPilot entwickelt<sup>13</sup>. ProposalPilot ist ein Demonstrator (Technologie-Reifegrad 6), der zeigt, wie mit einem besonders schlanken (die Umsetzung erfolgt ausschließlich mit PHP sowie JavaScript, d.h. ohne eigene Datenbank) technischen Ansatz

- strukturierte Forschungsberichte in einem benutzerfreundlichen, barrierearmen Webformular erfasst,
- in JSON, HTML, PDF sowie Metadaten in XML nach den Dublin-Core-Spezifikationen von DSpace 7 ausgegeben,
- sowie optional zur realitätsnahen Nachnutzbarkeit (den Prinzipien von FAIR Data folgend) als Issue in einem frei gewählten GitHub-Repository als Issue<sup>14</sup> gespeichert werden kann.

Fördergeber oder andere Organisationen, die von Proposalpilot ausgehend einen Prozess zur Einreichung strukturierter Forschungsberichte einrichten, können sehr einfach

---

<sup>11</sup> vgl. <https://github.com/TIBHannover/Fidus-Writer-Manual>

<sup>12</sup> vgl. <https://www.fiduswriter.org/2021/07/29/fidus-writer-3-10-with-folders-and-downloadable-document-templates/>.

<sup>13</sup> Vgl. den Open-Source-Quellcode des Projekts unter <https://github.com/TIBHannover/text-transfer-ii-prototype>, sowie den Demonstrator live unter <https://proposalpilot.texttransfer.org/>

<sup>14</sup> Vgl. <https://docs.github.com/de/issues>

- eigene Metadatenfelder in ihrem Eingabeformular ergänzen,
- eine Authentifizierung durch Token ergänzen,
- dem ProposalPilot lokale Schreibrechte auf dem eigenen Server geben. (Letztere Funktion ist im Demonstrator aus Sicherheitsgründen deaktiviert.)

### Rechtliche Rahmenbedingungen

AP5 war zu Projektende planmäßig abgeschlossen. Der Projektpartner TIB war an diesem Arbeitspaket nicht beteiligt.

#### 3.1.5. AP6: Implementierungskonzept

Der Projektverbund hatte sich zum Ziel gesetzt, die Ergebnisse der Methodenentwicklung so aufzubereiten, dass sie perspektivisch in eine konkrete Anwendung überführt werden können. Erste Vorarbeiten auf dem Weg zur Erweiterung einer zunächst als Pilotstudie vorliegenden Methode sollten mit Projektende am IDS bzw. dessen Unterauftragnehmer G&K erfolgen. Aufgabe des Arbeitspaketes war die Identifikation potenzieller Nutzender für die entwickelte Methode. Hierbei waren auch Akzeptanzfragen ihrer Anwendung zu berücksichtigen. Ausgehend vom konkreten Funktionsnachweis sollten weitergehende impact- und tranferrelevante Fragestellungen abgeleitet und Perspektiven zu möglichen Anwendungsszenarien dargestellt werden.

In dem Arbeitspaket sollte entsprechend ein Rollenmodell entworfen werden, das unterschiedliche Rollen im Zusammenhang mit der entwickelten Pipeline kennt und diese in Form eines Prozesses miteinander verbindet. Die Umsetzung wurde am konkreten Beispiel der IDS-internen Anwendung dargestellt.

Eine wesentliche Aufgabe des AP 6 bestand darin, einen Vorschlag für die mögliche organisatorische Implementierung von *TextTransfer* zu entwickeln. Der Startpunkt für den Vorschlag war, dass das BMBF das gemeinsame IDS-/TIB-Entwicklungsprojekt "*TextTransfer*" finanzierte. Als Projektergebnis sollte "*TextTransfer*" – entsprechend der im Bericht erwähnten Spezifikationen - als KI-gestütztes Softwaretool vorliegen. Der Projektträger DLR als damaliger Vertreter des Zuwendungsgebers hatte

großes Interesse daran, dass dieses Projektergebnis in den geförderten Institutionen, idealerweise auch darüber hinaus, aktiv genutzt würde. Hierzu sollte „TextTransfer“ in den beteiligten Einrichtungen implementiert werden.

Im Projektverlauf wurden vier mögliche Implementierungsvarianten erarbeitet.

## Systematisierung möglicher Implementierungskonzepte

### Implementierungstypologie

Aufgrund der entwickelten Anwendung sowie der vielfältigen Anwendungsfelder lässt sich eine Vielzahl, teilweise auch kombinatorisch anzuwendender, Implementierungslösungen finden, welche in der folgende Tabelle systematisch dargestellt werden.

|                        |  |   |
|------------------------|--|---|
| <b>Beratungsmodell</b> | 3. Software PLUS Beratung durch das IDS                      | 4. Software PLUS Beratung als Kooperationsmodell TIB/IDS                      |
| <b>Leistungsart</b>    |  |   |
| <b>Softwaremodell</b>  | 1. Software inkl. Schulung und Implementierung durch das IDS | 2. Software inkl. Schulung und Implementierung als Kooperationsmodell IDS/TIB |
|                        | <i>IDS alleine</i>   | <i>IDS Kooperation</i>  |
|                        | <i>Durchführende(r)</i>                                      |   |



1

Abbildung 1: Systematisierung möglicher Implementierungskonzepte

Vier grundsätzlich mögliche Implementierungsvarianten wurden identifiziert:

1. Software inkl. Schulung anbieten durch den Projektpartner IDS
2. Software inkl. Schulung in Kooperation durch die TIB und das IDS
3. Software inkl. Beratung durch den Projektpartner IDS
4. Software inkl. Beratung in Kooperation durch die TIB und das IDS

33

Mit den dargestellten Möglichkeiten würden auch kultureller und organisatorischer Wandel einhergehen. Management und Business Development-Aktivitäten wären zu erfüllen, die die existierenden u.a. rechtlichen Rahmendbedingungen eines Forschungsinstituts sprengen würden.

AP6 war zu Projektende planmäßig abgeschlossen.

### 3.1.6. AP7: Kommunikationskonzept

Die wesentliche Aufgabe des AP 7 „Kommunikationskonzept“, das durch den Projektpartner IDS zu bearbeiten war, bestand darin, Wege zu entwickeln, um die Arbeiten am Projekt wie auch das Projektergebnis in den anvisierten Zielgruppen bekanntzumachen.

Entsprechend der Vorhabenbeschreibung, die eine generalisierte Bereitstellung der Projektergebnisse vorsah, wurde unter der Federführung des IDS von beiden Partnern eine eigene Webseite für das Projekt aufgebaut, die unter der **Domain *texttransfer.org*** registriert ist.

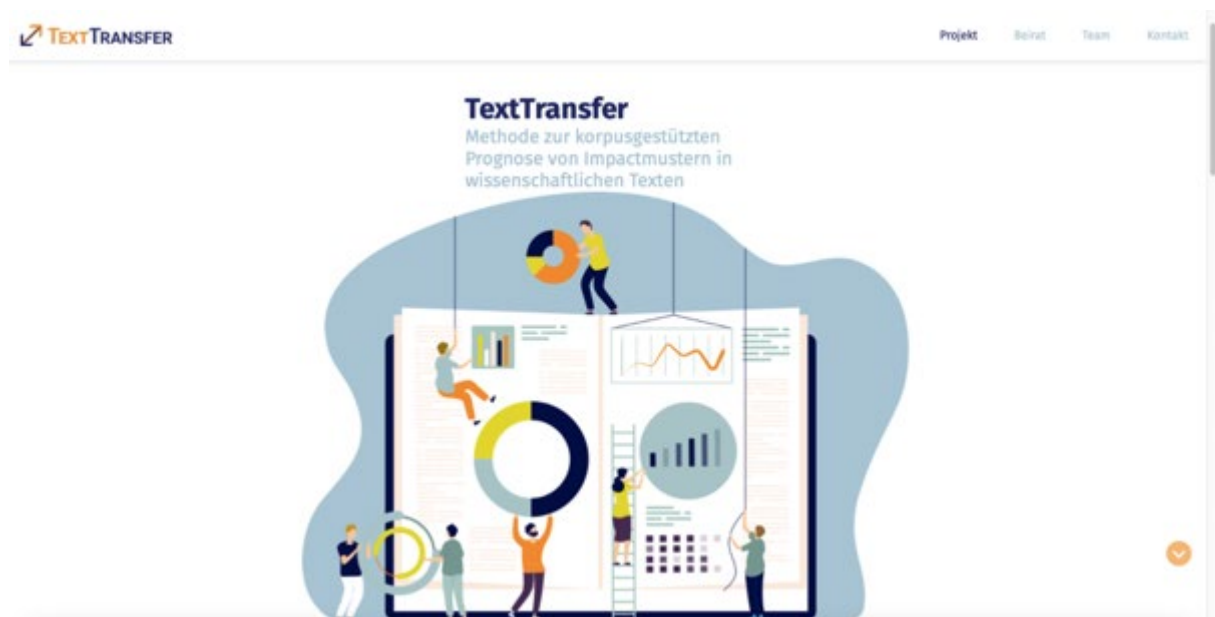


Abbildung 2: Die Webpräsenz des Projektes unter „texttransfer.org“

Die Web-Präsenz diente dazu, die Sichtbarkeit des Projektes zu erhöhen, indem sie u. a. den Ansatz als gefördertes Vorhaben dokumentierte, die unterschiedlichen Player und ihre Rollen vorstellte, als auch

für Interessierte an der Methode eine Kontaktmöglichkeit für weitergehenden Austausch in einfacher Weise zu ermöglichen. Die Web-Seite wurde in deutscher und englischer Sprache angelegt, um die Reichweite zu erhöhen.

Zwecks Konzeption der Website wurde eine spezielle Taskforce zum Thema „Reformulierung Mission *TextTransfer* für die neue Website“ aus Mitarbeiterinnen und Mitarbeitern der Projektpartner TIB und IDS gebildet, die sich mit den Inhalten und dem Design des neu zu gestaltenden Webauftritts befasste. Ziel war es, die Website – sowohl was die Struktur als auch die Inhalte betraf – so zu gestalten, dass die Idee und das Anliegen von *TextTransfer* auch Besucherinnen und Besuchern, die wenig bis kein Hintergrundwissen zum Thema hatten, so verständlich wie möglich vermittelt wurden und somit mögliche Barrieren, sich mit dem Lösungsansatz auseinanderzusetzen, möglichst gering zu halten.

Unter diesem Aspekt konnte für die Formulierung und Erstellung der kritischen textuellen Inhalte – nach einem entsprechenden Briefing durch die IDS-Kolleginnen und -Kollegen – ein Mitarbeiter des Projektpartners TIB gewonnen werden, der sich schwerpunktmäßig mit dem Thema Wissenschaftskommunikation beschäftigt.

## Unser Lösungsansatz

### Mit Textanalyse und maschinellem Lernen zu Impact-Kategorien

In einer ersten Projektphase (TextTransfer I; 2016-2019) wurden exemplarische Projektabschlussberichte mit Hilfe von linguistischen Analysemethoden und des maschinellen Lernens automatisiert ausgewertet. In der derzeitigen zweiten Projektphase (TextTransfer II; 2020-2023) verbessern wir die Genauigkeit und Skalierbarkeit von TextTransfer und optimieren so die bislang prototypisch entwickelte Methode zur automatisierten, nutzergerechten Erschließung von wissenschaftlichen Texten für die wissenschaftliche Praxis. Im Mittelpunkt stehen dabei die messbaren, durch Wissenstransfer erzeugten wirtschaftlichen, technologischen, soziokulturellen, politisch-rechtlichen und/oder ökologischen Wirkungen (Impact).



TextTransfer identifiziert durch maschinell gewonnene, statistische Präzedenzen bestimmte sprachliche Texteneigenschaften und textstrukturelle Zusammenhänge, die auf solche Wirkungen hinweisen, und fasst sie zu verschiedenen Impact-Kategorien zusammen. So können Aussagen über die potenziellen Wirkungen von Forschungsergebnissen getroffen werden. Aktuell orientieren wir uns an den spezifischen Profilen der beteiligten Institutionen und zeigen anhand konkreter Anwendungsfälle, wie formal und inhaltlich heterogene Textdaten analysiert und TextTransfer in die Strukturen und organisatorischen Abläufe der beteiligten Häuser integriert werden können.

Abbildung 3: Unter „texttransfer.org“ wird der Ansatz der Methode TextTransfer leicht verständlich erklärt

Der Projektpartner IDS kümmerte sich um die Erstellung der grafischen Inhalte, das Gesamtlayout sowie die technische Umsetzung der Website.

Die Website wurde in *HTML*, *CSS* und *JavaScript* implementiert. Zusätzlich wurde das *CSS*-Framework *Bootstrap* verwendet. Für die Funktion des Kontaktformulars wurde die Programmiersprache *PHP*

eingesetzt.<sup>15</sup> Die Domain „texttransfer.org“ sowie ein Webhosting-Paket zur Veröffentlichung der Website wurden bei lima-city.de – finanziert durch IDS-Haushaltsmitteln – angemietet.

AP7 war zu Projektende planmäßig abgeschlossen.

### **3.1.7. AP8: Projektmanagement**

Als gesamtverantwortliche Institution hatte das IDS im Verbund sowohl die Projektleitung als auch das Gesamt-Projektmanagement inne und koordinierte somit alle im Projekt zu erbringenden Arbeiten, sowohl der Partner als auch der Unterauftragnehmer. Das Projektmanagement innerhalb eines definierten Aufgabengebietes eines Partners lag beim jeweiligen Projektpartner.

Zu Projektbeginn hatten daher alle Projektpartner entsprechende interne Maßnahmen getroffen, die einen reibungslosen Ablauf des Projektes sicherstellten (u.a. Einrichtung einer internen Infrastruktur mit beispielsweise projektspezifischem Exchange-Server, Zuständigkeiten etc.).

Des Weiteren beinhaltete das Arbeitspaket die Erstellung der Dokumentationspflicht gegenüber dem Projektträger in Form von Zwischenberichten als auch dem Abschlussbericht; alle Berichte wurden unter der Koordination des IDS unter Mitarbeit aller Projektbeteiligten als auch deren Bereitstellung von Informationen und der aufgabenspezifischen Prüfung vorgenannter Berichte erstellt. Die Berichte der beiden beteiligten Partner wurden dem Projektträger jeweils termingerecht vorgelegt.

AP8 war zu Projektende planmäßig abgeschlossen.

## **3.2. Die wichtigsten Positionen des zahlenmäßigen Nachweises**

Das Budget wurde zu 100 % für die Personalkosten verwendet. Dienstreisen haben aufgrund der Corona-Maßnahmen im Projektzeit nicht stattgefunden, das beantragten Budget wurde daher nicht genutzt. Die Mittel wurden stattdessen zur Deckung der erhöhten Personalkosten genutzt.

---

<sup>15</sup> Auf Grund des Umfangs der Skripte, sind diese nicht Teil des Berichts, können jedoch auf Anfrage per Mail an [bopp@ids-mannheim.de](mailto:bopp@ids-mannheim.de) zur Verfügung gestellt werden.

### 3.3. Notwendigkeit und Angemessenheit der geleisteten Arbeit

Projektträger und Projektpartner sind im Vorfeld der Anberaumung des Vorhabens *TextTransfer* zu der Erkenntnis gekommen, dass die unzähligen Ergebnisse bisherige Forschungsarbeiten aller Disziplinen in ihrer verschriftlichen Form eine wertvolle, aber bisher nicht vollumfänglich genutzte Ressource in den Archiven einschlägiger Gedächtnisorganisationen darstellen. Weiterhin stand zu erwarten, dass klassische, analoge Verfahren der Auswertung hinsichtlich verwertbarer Forschungsergebnisse nicht mehr zu ihrer Erfassung hinreichen dürften. Auf dem Wege zur Etablierung einer routinemäßigen Transferkultur in den Wissenschaften war es allen Beteiligten ein Anliegen, die Chancen der Digitalisierung auch in diesem Bereich zu nutzen. Ein automatisiertes Verfahren war zu entwickeln, das in erster Linie die Wissenschaft unterstützt, Transfer- und Impactpotenziale in wissenschaftlichen Texten besser zu identifizieren und so den Wirkungsgrad von Investitionen in die Forschung zu optimieren.

Für einen Funktionsnachweis der Methode *TextTransfer* war daher ein Zusammenspiel von einer transferrelevanten Indikatoren- bzw. Kategorienschemata-Entwicklung, der Annotation von Textquellen sowie der exemplarischen Adaption vorhandener Softwarelösungen basierend auf einem bedarfsgerecht zugeschnittenes Korpus von Forschungsberichten als Stichprobe, die zwecks maschineller Lesbarkeit konvertiert sein musste, nötig.

Mit den an *TextTransfer* beteiligten Instituten und Experten hat sich ein Verbund zusammengefunden, der notwendige Kernfähigkeiten im Korpusaufbau, Text Mining,, Impact Assessment, maschinellem Lernen und Transfereigenschaften von Forschungswissen bündelt - eine Konstellation, die vor dem Hintergrund der Fragestellung im Projekt und der gewählten deutschsprachigen Datenbasis bisher nicht existierte. Eine entsprechende Förderung zur Herstellung notwendiger Verknüpfungen und Kapazitäten war daher notwendig. Diese Kombination von Expertenwissen konnte durch das Projektteam IDS und TIB und den Unterauftragnehmern Görden & Köller GmbH (G&K) und Prof. Dr. Jana Diesner von der School of Information Sciences / The *iSchool* der Universität von Illinois at Urbana-Champaign (UIUC) und ihrer Arbeitsgruppe erbracht werden. Eine solch erfolgreiche Kooperation, die aufgrund des neuartigen Ansatzes und des hohen Innovationsgrads ein hohes Forschungsrisiko birgt,

wäre angesichts zu geringer personeller Ressourcen sowie der nicht hinreichend fachübergreifend breiten Kompetenzen ohne fördernde Maßnahme nicht möglich gewesen.

### **3.4. Voraussichtlicher Nutzen, insbesondere die Verwertbarkeit des Ergebnisses im Sinne des fortgeschriebenen Verwertungsplans**

Ziel der zweiten Förderphase war es, die positiven Ergebnisse von *TextTransfer I* (Pilot) aufzugreifen und konsequent weiterzuführen und die für eine nachhaltige Methodenanwendung unter wissenschaftlichen Realbedingungen in den beteiligten Instituten im Wissens- und Technologietransfer hohe Relevanz haben. Die Methode *TextTransfer* kann somit zu einem Instrument des Wissens- und Technologietransfers werden, das es den beteiligten Instituten auch in Zusammenarbeit mit wissenschaftlichen Dachorganisationen wie der Leibniz-Gemeinschaft erlaubt, eigene schriftliche Erzeugnisse frühzeitig und schnell nach Impactwahrscheinlichkeiten zu kategorisieren und somit die gezielte Suche nach Transfer- und Impactpotenzialen der eigenen Forschung ermöglichen.

Die Projektpartner haben prototypische Anwendungsfälle entwickelt, um die Funktionalität der Methode domänenübergreifend und unter komplexen Bedingungen der wissenschaftlichen Praxis zu demonstrieren, Wege zu ihrer Nutzung in ihrem Alltag aufzuzeigen und Anforderungen an die strukturelle Beschaffenheit wissenschaftlicher Texte mit dem Ziel einer optimierten Maschinenlesbarkeit abzuleiten. *TextTransfer* beinhaltet außerdem Kommunikationsmaßnahmen, um zukünftigen, projektexternen Anwendern bereits während der laufenden Projektphase Einsicht in die Anwendungsfälle und Implementierungslösungen bei den Projektpartnern zu gewähren, die eine eigene Nachnutzung anstreben.

### **3.5. Zum Zeitpunkt der Durchführung des Vorhabens dem ZE bekannt gewordener Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen**

Auf dem Gebiet des Vorhabens sind auf der Eben der Sprachtechnologie und statistischer Modelle nennenswerte Fortschritte erzielt. Dies fällt in den Verantwortungsbereich des Projektpartners IDS Mannheim und wird in dessen Schlussbericht näher dargelegt.

### 3.6. Erfolgte oder geplante Veröffentlichungen des Ergebnisses nach Nr. 6 (BNNest-BMBF 98)

#### 3.6.1. Vorträge

Vorträge über die gemeinsam erzielten Projektergebnisse fanden weitgehend in Verantwortung des Projektpartners IDS statt.

#### 3.6.2. Veranstaltung, Workshops, Kurse

Auch Veranstaltungen, Workshops und Kurse über die gemeinsam erzielten Projektergebnisse fanden weitgehend in Verantwortung des Projektpartners IDS statt.

#### 3.6.3. Publikationen/Poster

- Gabrys-Deutscher, E., and Lütjen, A. (2022). Deutsche Forschungsberichte – eine Sonderform Grauer Literatur in der Technischen Informationsbibliothek (TIB): Rückblick auf über ein halbes Jahrhundert Bestandsaufbau und Ausblick in die Zukunft. O-Bib. Das Offene Bibliotheksjournal Herausgeber VDB, 9(1), 1-13. <https://doi.org/10.5282/o-bib/5768>

27.06.2024 \_\_\_\_\_  
Hannover, den

A handwritten signature in blue ink, appearing to read "Heller".

\_\_\_\_\_  
Lambert Heller (Leiter Teilprojekt TIB)