

Abschlussbericht

B-STA-R: Bilateral Science and Technology Agreements Repository

Eine Vertragsdatenbank für Forschende in den Sozial- und Geisteswissenschaften

1. Allgemeine Angaben

Antragsteller:

Wissenschaftszentrum Berlin gGmbH

Abteilung:

Forschungsgruppe der Präsidentin

Projektleitung:

Nicolas Rüffin

Thema des Projekts:

Überprüfung, Dokumentation und Bereitstellung eines umfassenden Datensatzes zu zwischenstaatlichen Kooperationsabkommen im Bereich Wissenschaft und Technologie

Berichtszeitraum, Förderungszeitraum insgesamt:

01.07.2021 - 31.08.2021

Zugang zum Datensatz:

Rüffin, Nicolas (2021): B-STA-R: A repository for bilateral science and technology agreements. Version 1.0.0. WZB Berlin Social Science Center. Dataset.

<https://doi.org/10.7802/2310>

2. Arbeits- und Ergebnisbericht

Einleitung

Zielsetzung des Projekts war die Bereitstellung eines möglichst umfassenden Datensatzes zu zwischenstaatlichen Abkommen der Wissenschafts- und Technologiekooperation (englisch: Science and Technology Agreements, im Folgenden STA genannt).

Hintergrund dieser Unternehmung war der bis dato vorhandene Mangel an reliablen und kompakt verfügbaren Daten zu dieser Art von bilateralem Abkommen. Dies verblüfft angesichts der Tatsache, dass Themen der Außenwissenschaftspolitik, in denen STA immer wieder als probates Mittel politischer Agency genannt werden, in den vergangenen zehn Jahren erheblich an Aufmerksamkeit gewonnen haben. Bei diesem Typ von bilateralen Abkommen handelt es sich um Vereinbarungen zwischen Staaten, in denen Pläne zur Kooperation in Wissenschaft und Technologie rechtsverbindlich festhalten werden. Dabei kann es sich sowohl um Absichtserklärungen als auch um konkrete Absprachen etwa im Bereich des Schutzes geistigen Eigentums, zum Austausch von Studierenden und Forschenden oder zum Aufbau gemeinsamer Steuerungsgremien handeln. Im Gegensatz zu allen Arten von Verträgen zu ökonomischen oder sicherheitsbezogenen Themen existierte bisher kein zentraler Anlaufpunkt für Studien zu bilateralen Wissenschaftskooperationen. Entsprechende Daten zu Vertragsabschlüssen, Rahmenbedingungen und Inhalten existierten bisher öffentlich zugänglich nur verstreut und wenig systematisiert. Am WZB wurde in den vergangenen Jahren ein umfangreicher Datensatz zu entsprechenden STA aufgebaut, der mithilfe des Projekts aktualisiert, gesichert und in Kooperation mit GESIS für vertiefende Analysen allgemein zugänglich gemacht wurde. Damit schließt der Datensatz B-STA-R eine Lücke in der Verfügbarkeit von Forschungsdaten für die Erforschung internationaler Beziehungen in Wissenschaft und Technologie.

Ablauf des Projekts

Die im Förderzeitraum durchgeführten Arbeiten gliederten sich in drei Phasen. Zunächst wurde das existierende Datenmaterial einer inhaltlichen Prüfung und Ergänzung unterzogen. Da die ersten Eintragungen im Datensatz bereits 2017 vorgenommen wurden, mussten alle Datenquellen und Informationen auf Gültigkeit überprüft sowie gegebenenfalls korrigiert oder aktualisiert werden. Hier zeigte sich, dass die ganz überwiegende Zahl der erfassten Vertragsdaten bereits korrekt erfasst war. In einigen wenigen Fällen war die Originalquelle mit dem entsprechenden Dokument nicht mehr zugänglich, dies betraf etwa zu einem früheren Zeitpunkt erfasste Daten des brasilianischen Forschungsministeriums. In anderen Fällen konnten

hingegen zusätzliche Informationen generiert und Volltexte erschlossen werden. So hat beispielsweise die Volksrepublik China zwischenzeitlich ein umfassendes Repositorium ihrer internationalen Verträge freigeschaltet, aus dem Angaben zu vorhandenen Einträgen verifiziert und teilweise zusätzliche STA ergänzt werden konnten. Als problematisch erwies sich die Frage der rechtssicheren Archivierung von Vertragsdokumenten. Viele nationale Datenquellen schlossen per Disclaimer eine separate Archivierung der Abkommen selbst aus (dies gilt beispielsweise für die im Bundesgesetzblatt veröffentlichten deutschen STA). In enger Abstimmung mit dem Team Forschungsdatenmanagement des WZB wurde daher entschieden, lediglich per Link möglichst exakt auf den Speicherort zu verweisen und das Datum des letzten Seitenaufrufs zu dokumentieren. Nichtsdestotrotz war es im Rahmen der ersten Arbeitsphase möglich, die im ursprünglichen Projektplan genannte Zahl von 1107 Abkommen um 31 weitere Verträge auf nunmehr 1138 zu steigern. Auch die Quote verfügbarer Volltexte konnte von avisierten 61% auf über 75% gesteigert werden, d.h. für über 850 STA liegen nun Verweise auf Volltexte im html- oder .pdf-Format vor. Auf Basis dieses ergänzten Datensatzes wurde eine Klassifikation zur Einschätzung der Reliabilität der Datenbankeinträge entwickelt, die von sehr hoher Reliabilität bei vorhandenen Volltexten bis zu sehr niedriger Reliabilität bei ausschließlicher Nennung entsprechender STA in singulären Sekundärquellen reicht. Des Weiteren wurden bei zahlreichen STA zusätzliche Informationen zum Datum des Abschlusses, zum Inkrafttreten und zu den Vertragslaufzeiten ergänzt.

In der zweiten Projektphase wurde das Datenmaterial in englischer Sprache dokumentiert und umfassend durch das Forschungsdatenmanagement am WZB gegengeprüft. Dabei wurden einzelne Übertragungsfehler behoben und Ungereimtheiten in Dateneinträgen und -beschreibungen geklärt. Des Weiteren wurden in dieser Phase eine Reihe von Schritten unternommen, um die Daten unter Berücksichtigung der FAIR-Prinzipien veröffentlichen zu können. Unter anderem wurden einheitliche Sprachregelungen getroffen, um die spätere Handhabung der Daten für zukünftige Nutzer zu erleichtern. Es wurde entschieden, den Datensatz in nicht-proprietärer Form als .csv-Dateien zu veröffentlichen, um größtmögliche Portabilität zu verschiedenen Anwendungsprogrammen zu ermöglichen. Darüber hinaus wurden ausführliche Metadaten im von GESIS angepassten Schema der Digital Documentation Initiative (DDI) erzeugt.

In der finalen Projektphase wurde der so aufbereitete Datensatz an GESIS zu einer letzten Überprüfung übergeben. Nachdem diese abgeschlossen war, konnte der B-STA-R Datensatz am 17.09.2021 erfolgreich unter einer CC BY 4.0-Lizenz veröffentlicht werden.

Beschreibung der veröffentlichten Daten

Der veröffentlichten Datensatz besteht aus zwei .csv-Dateien, die durch ein Codebuch im .pdf-Format ergänzt werden. Zum einen sind in der Datei "STA_data" die eigentlichen Abkommen als einzelne Dateneinträge vermerkt. Hier finden sich Informationen zum Jahr des Abschlusses, zum Jahr des Inkrafttretens, zur vereinbarten Laufzeit sowie zur Verfügbarkeit von Volltexten, zur Sprache des Dokuments sowie zu den Datenquellen, denen diese Informationen entnommen wurden. Besonderheiten bei einzelnen Abkommen (z.B. etwaige Folgeabkommen oder eingeschränkte Verfügbarkeiten) sind in einer eigenen Spalte vermerkt. Diese Datei enthält Informationen zu 1138 originären bilateralen STA sowie zu einer Reihe von Folgeverträgen. Von diesen 1138 Abkommen weisen rund 78% eine hohe Reliabilität mit Blick auf die Sicherheit der gemachten Angaben im Datensatz auf. Dokumentiert sind in dieser Datei Abkommen im Zeitraum zwischen 1937 und 2020, die zwischen Staaten der G20 und OECD sowie zahlreichen Drittstaaten abgeschlossen wurden.

Die zweite Datei namens "sources_and_searchterms.csv" soll es Nutzern erleichtern, den Prozess der Datenselektion und -sammlung transparent nachvollziehen zu können. In dieser Datei sind die einzelnen Datenquellen gelistet, in denen nach STA gesucht wurde. Des Weiteren enthält diese Datei Informationen zu verwendeten Suchstrategien und Konfigurationen von Filtern, die bei der Suche in den verschiedenen nationalen Repositorien und Quellen eingesetzt wurden.

Das Codebuch enthält eine kurze Einführung ins Thema STA mit Nennung relevanter Hintergrundliteratur. Es erläutert die Motivation zur Datensammlung und die Prozesse des Zustandekommens der Daten, insbesondere zu Selektionskriterien für den Ein- oder Ausschluss einzelner Abkommen. Auch werden hier erste Ideen für eine Weiternutzung der Daten für neue wissenschaftliche Fragestellungen skizziert. Darauf folgt eine detaillierte Beschreibung der einzelnen in beiden .csv-Dateien verwendeten Variablen. Hier ist insbesondere die Erläuterung der Kriterien für die Reliabilitätseinschätzung zu nennen, die unter der Variable "STAVerified" ausgeführt wird.

Die Informationen in allen drei Dateien sind auf Englisch dokumentiert, um einen möglichst großen Anwenderkreis anzusprechen.

Outreach

Um die Bekanntheit - und damit den Kreis potentieller Nutzer - zu erhöhen, wurden seit dem Upload des Datensatzes verschiedene Anstrengungen unternommen, auf B-STA-R hinzuweisen. In Kooperation mit der Kommunikationsabteilung des WZB wurde der Datensatz per Twitter vorgestellt. Die europäische Community der an Themen von internationaler Bildung und

Forschung interessierten Politikwissenschaftler*innen wurde durch einen Beitrag im Blog "Europe of Knowledge"¹ der ECPR Standing Group Knowledge Politics and Policies auf den Datensatz hingewiesen. Auch in der Dezemberausgabe der WZB-Mitteilungen wird ein informativer Beitrag zum Datensatz publiziert werden. Darüber hinaus wird der Datensatz auf der Webseite der European Alliance for Science Diplomacy² aufgeführt, so dass an Außenwissenschaftspolitik Interessierte ihn leicht finden können.

¹ siehe <https://era.ideasoneurope.eu/2021/10/14/b-star-a-new-dataset-to-study-science-diplomacy-and-global-science/>, letzter Zugriff am 28.10.2021.

² siehe <https://www.science-diplomacy.eu/library/repository-for-bilateral-science-and-technology-agreements/>, letzter Zugriff am 28.10.2021.

3. Zusammenfassung

Der B-STA-R-Datensatz enthält Informationen zu 1138 originären bilaterale Abkommen zur wissenschaftlich-technischen Zusammenarbeit sowie zu einer Reihe von Folgeverträgen, die hauptsächlich von Ländern der G20 und der OECD im Zeitraum zwischen 1937 und 2020 abgeschlossen wurden. Die Daten wurden im Zeitraum zwischen 2017 und 2021 erhoben, zwischen Juli und August 2021 zur Veröffentlichung vorbereitet und im September 2021 publiziert. Sämtliche veröffentlichten Informationen stehen interessierten Nutzenden unter einer CC BY 4.0-Lizenz zur Verfügung.

Dokumentiert sind u.a. Informationen zum Jahr des Vertragsabschlusses, zum Inkrafttreten, die initial vereinbarte Vertragslaufzeit sowie zur Existenz von Volltexten der Abkommen. Des Weiteren steht eine umfassende Erläuterung der Suchstrategien in der Generierung des Datensatzes und zu verwendeten Selektionskriterien zur Verfügung. Alle Angaben sind in englischer Sprache gehalten. B-STA-R erlaubt damit wissenschaftliche Untersuchungen im Bereich der Internationalen Beziehungen, Wissenschaftsforschung, der Außenwissenschaftspolitik, sowie der Geschichtswissenschaften.