

Der Beitrag thematisiert Bedeutung und Spezifika von Forschungsdaten der Wirtschafts- und Sozialgeschichte. Weit zurückreichende Zeitreihen werfen vielfältige methodische Probleme auf, die oft eine umfangreiche Dokumentation der Daten erfordern. Mit dem ZBW Journal Data Archive (JDA) und dem Forschungsdaten-Hub Emporion werden zwei relativ neue Angebote vorgestellt, die den Forschenden die notwendigen Ressourcen und Anreize geben sollen, ihre oft mühsam erhobenen Daten der Fachcommunity im Open Access zugänglich zu machen. Dies entspricht zwar fest etablierten Standards des Faches, war jedoch lange mit erheblichen Schwierigkeiten verbunden. Wie wir argumentieren, versprechen offene wissenschaftliche Informationsinfrastrukturen wie das JDA und Emporion in Verbindung mit neuen Methoden der Tabellenerkennung und generell der Künstlichen Intelligenz neue Impulse für die Forschung.

The article addresses the significance and specifics of research data in the fields of economic and social history. Long historical time series pose a range of methodological problems; the solution to many of these involves extensive documentation of the data. The ZBW Journal Data Archive (JDA) and the Emporion research data hub are two relatively new services which provide researchers with the necessary resources and incentives to grant the academic community open access to their often painstakingly collected data. This complies with well-established standards in the field, yet for a long time it was fraught with considerable difficulties. As the authors argue, open scientific information infrastructures such as the JDA and Emporion, combined with new methods of table recognition and artificial intelligence in general, are expected to provide new stimulus for research.

MARK SPOERER, LAETITIA LENEL, CHRISTIAN MATHIEU, SVEN VLAEMINCK

Forschungsdaten in der Schnittmenge von Geschichts-, Sozial- und Wirtschaftswissenschaften

Neue Datenrepositorien für die wirtschafts- und sozialhistorische Forschung

Kleines Fach, große Vielfalt der Forschungsdaten

Spätmittelalterliche städtische Maurerlöhne, Brotpreise, Körpergrößen von frühneuzeitlichen Rekruten, kommunale Steuereinnahmen, Volks- und Viehzählungen, das Bruttoinlandsprodukt usw. – dies ist nur eine kleine Auswahl der vielfältigen Forschungsdaten, die in der wirtschafts- und sozialhistorischen Forschung von Bedeutung sind. Die Forschungsdaten der Wirtschafts- und Sozialgeschichte teilen einige Charakteristika mit anderen Disziplinen, weisen zugleich aber auch Besonderheiten auf, die spezifische Anforderungen an geeignete Repositorien richten.

– *Geringe Masse:* Noch hat die Geschichtsschreibung es nur selten mit Massendaten zu tun, wie sie in anderen empirischen Sozialwissenschaften anfallen (etwa Aktienkurse im Hochfrequenzhandel, Suchmaschinenanfragen, Social-Media-Beiträge). Die Größe der Datensätze beträgt daher oft nur einen Bruchteil derer aus den Wirtschaftswissenschaften oder den angewandten Naturwissenschaften. Dies wird sich in absehbarer Zeit vielleicht ändern, wenn die zeithistorische Forschung beginnt, ebenfalls auf solche Datenquellen zu rekurrieren.

– *Aufwendige Datenerhebung und -zusammenstellung:* Zunehmend, aber letztlich immer noch selten, können Historiker*innen bestehende Datensätze nutzen. Viele Daten müssen entweder aus textlichen Quellen erhoben oder, was noch arbeitsintensiver ist, zusammengestellt werden. Im ersten Fall kann auf publizierte oder archivalisch überlieferte Tabellen zurückgegriffen werden, die entweder durch klassisches Abtippen oder zunehmend auch durch OCR-Software digital verfügbar gemacht werden. Tabellen, so intuitiv sie dem menschlichen Verstand auch erscheinen mögen, stellen bislang eine hohe Hürde für automatisierte Transkriptions-Software (wie etwa Transkribus) dar. Es ist jedoch abzusehen, dass dieses Problem zumindest für publizierte Tabellen in der zweiten Hälfte der 2020er-Jahre gelöst sein wird.¹ Noch aufwendiger ist die Erstellung von Datensätzen, die auf der Aggregation von Einzelbeobachtungen basieren, etwa bei der Übertragung seriell vorliegender standardisierter Sterbeurkunden in eine strukturierte Datenbank.

– *Wechselnde Berichtsgrundlage bei Zeitreihen:* Die statistischen Ämter haben den gesetzlichen Auftrag, Daten für aktuell anstehende politische, soziale und ökonomische Entscheidungsfindungen zu liefern,

(leider) nicht für historische Langzeitbetrachtungen. Die Berechnung wichtiger Kenngrößen wie etwa des Bruttoinlandsprodukts oder des Verbraucherpreisindex werden daher von den statistischen Ämtern immer wieder angepasst. Konsumgüter, die nicht mehr nachgefragt werden (etwa Kassettenrekorder) verschwinden aus dem sogenannten Warenkorb, der dem Verbraucherpreisindex zugrunde liegt, und neue kommen hinzu (z.B. Mobiltelefone). Ein weiteres Beispiel für den Wechsel der Berichtsgrundlage sind Steuerdaten, bei denen berücksichtigt werden muss, ob und wann die ihnen zugrundeliegenden Erhebungsparameter (Freibetrag, Bemessungsgrundlage, Steuersatz etc.) geändert wurden.² Noch misslicher aus historiografischer Sicht ist, dass die statistischen Ämter immer wieder die Definition komplexer Aggregate modifizieren und mit diesen Änderungen nur für ein paar Jahre zurückrechnen. Eine so wichtige Reihe wie das Bruttoinlandsprodukt Deutschlands wird daher selbst vom Statistischen Bundesamt retrospektiv nur bis 1950 ausgewiesen. Die Verknüpfung mit zeitlich weiter zurückreichenden Reihen ist daher ein methodisch ausgesprochen schwieriges Terrain.³

- *Wechselnde Gebietskörperschaftsgrenzen:* Das deutsche Bruttoinlandsprodukt (BIP) ist auch ein gutes Beispiel für die Problematik wechselnder Territorialstände, was gerade in der deutschen Geschichte eine große Herausforderung darstellt: Vor 1871 und insbesondere vor 1806 war Deutschland bekanntlich ein politischer Flickenteppich mit unzähligen Territorien. Seit 1871 hat Deutschland mehrere größere und weitere kleinere territoriale Veränderungen erlebt. Kann das BIP pro Kopf von 1913, das u. a. auch Elsass-Lothringen, die nach dem Zweiten Weltkrieg gegründete DDR und die 1945 Polen zugeschlagenen Gebiete umfasst, mit dem westdeutschen von 1975 verglichen werden? Und lässt, wenn man dem Vergleich das Territorium von 1975 zugrunde legen möchte, die territoriale Zusammensetzung des BIP von 1913 dies 1:1 zu, d.h. kann man die anderen Gebiete überhaupt herausrechnen?
- *Problematische Datenqualität:* Daten werden immer in einem komplexen sozialen Feld konstruiert. Behörden wollen bestimmte Dinge wissen und andere nicht. Historisch lässt sich zudem feststellen, dass mit der Erhebung beauftragte Mitarbeiter*innen hohem Arbeitsaufwand bei der Erhebung der Daten nicht immer positiv gegenüberstanden; und schließlich veröffentlichten Behörden bestimmte Daten gar nicht oder aus politischen Gründen verkürzt, irreführend oder bewusst verfälscht.⁴

Insofern sind historische Daten in besonderem Maße erklärungsbedürftig und benötigen zahlreiche Metadaten. Auch aufgrund dieser spezifischen Anforderungen gab es lange kein Repositorium für historische Forschungs-

daten, obwohl die Wirtschafts- und Sozialgeschichte seit ihren Anfängen eine der klassisch historisch-kritischen Methode verpflichtete, und zugleich in weiten Teilen datenbasierte Disziplin ist. Ein wesentlicher Meilenstein war das bei der GESIS entwickelte Repositorium »histat«, das 2004 online ging und im Herbst 2021 ans Leibniz-Institut für Finanzmarktforschung SAFE wechselte.⁵ »histat« hostet die Daten von 536 Studien mit 436.077 Zeitreihen, nimmt aber keine neuen Inhalte mehr an (vgl. weiter unten).

Die 2020 von der Gemeinsamen Wissenschaftskonferenz im Verbund mit der Deutschen Forschungsgemeinschaft (DFG) initiierte Nationale Forschungsdateninfrastruktur (NFDI) zielt lediglich auf die Vernetzung, Auffindbarkeit und Zugänglichkeit bereits existierender disziplinärer Ressourcen ab. Als hochgradig interdisziplinäres Fach steht die Wirtschafts- und Sozialgeschichte zudem vor der Herausforderung, gewissermaßen die Schnittmenge dreier NFDI-Konsortien – NFDI4Memory, BERD@NFDI und KonsortSWD – zu bilden.⁶

Eine weitere Herausforderung für das Fach liegt in der Gewährleistung von Überprüfbarkeit und Reproduzierbarkeit publizierter Forschungsergebnisse, die auf der Auswertung von Daten basieren. Diese in den vergangenen Jahren verstärkt auf die wissenschaftspolitische Agenda gerückte Forderung zählt grundsätzlich zu den fest etablierten zentralen Standards des Fachs. Parallel dazu machen die Open Science-Erklärungen von Hochschulen, die Verwendungsrichtlinien von Förderinstitutionen sowie die Data Availability Policies von Zeitschriften und Verlagen den offenen Zugang zu Forschungsdaten immer häufiger sogar zur Pflicht – Entwicklungen, die für quantitativ orientierte Fächer wie die Wirtschafts- und Sozialgeschichte von besonderer Relevanz sind.

Die Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte (VSWG), die weltweit älteste wissenschaftliche Zeitschrift in diesem Fachgebiet (und seit Anfang 2024 im Open Access), bietet daher seit Anfang 2019 ihren Autor*innen an, Forschungsdaten hosten zu lassen.⁷ Die Ausgangsüberlegung war, Wissenschaftler*innen einen Anreiz zur Verfügungstellung ihrer im Rahmen eines abgeschlossenen Forschungsprojekts oft mühsam erhobenen Daten auch dann zu geben, wenn es keinen Drittmittelgeber gibt, der dies einfordert. Die VSWG schuf daher eine neue Rubrik »Forschungsdaten« bzw. »Research Data«, in der die Autor*innen ihre Daten in einem kurzen Aufsatz beschreiben können. Aufgrund der eingangs skizzierten vielfältigen idiosynkratischen Besonderheiten vieler Daten ist es dabei wichtig zu erklären, in welchem Zusammenhang sie wo und wie erhoben wurden und insbesondere, worin die Möglichkeiten und Grenzen der (Interpretation der) Daten liegen. Wer immer diese Daten nutzt, ist gehalten, den dazugehörigen Forschungsdatenartikel zu zitieren. Gerade für Nachwuchswissenschaftler*innen, deren Publikations-

liste noch kurz ist, kann dies einen Anreiz darstellen, ihre Forschungsdaten mit anderen zu teilen. Das Angebot, die im Artikel verwendeten Daten hosten zu lassen, kann natürlich auch von Autor*innen regulärer VSWG-Artikel wahrgenommen werden. Als Partner für das Repository für die im betreffenden Artikel beschriebenen bzw. verwendeten Forschungsdaten konnte das ZBW – Leibniz-Informationszentrum Wirtschaft gewonnen werden, das schon seit 2016 mit dem Journal Data Archive (JDA) ein Datenrepositorium für vorwiegend wirtschaftswissenschaftliche Zeitschriften anbietet.

Das Journal Data Archive des ZBW – Leibniz-Informationszentrums Wirtschaft

Das Journal Data Archive ist ein kostenfreier Webservice für wirtschaftswissenschaftliche Fachzeitschriften. Ziel des Datenrepositoriums ist es, die Reproduzierbarkeit publizierter Wirtschaftsforschung zu verbessern und Herausgeber*innen von Fachzeitschriften beim Management publikationsbezogener Forschungsdaten zu unterstützen. Entwickelt wurde die Open Source Software als Ergebnis unterschiedlicher Anforderungen und Bedürfnisse in der Wirtschaftsforschung.

Das Reproduzierbarkeitsproblem in der Ökonomik

Das Themenfeld Reproduzierbarkeit empirischer Forschung ist in den Wirtschaftswissenschaften bereits seit Mitte der 1980er-Jahre in der Diskussion. Seinerzeit hatten die US-Ökonomen Dewald, Thursby und Anderson in einer aufsehenerregenden Studie konstatiert, dass die Ergebnisse vieler empirischer Artikel, die in wirtschaftswissenschaftlichen Top-Zeitschriften publiziert wurden, nicht reproduziert werden konnten: »[...] inadvertent errors in published empirical articles are a commonplace rather than a rare occurrence«. ⁸ Ein durchaus problematischer und in seinen Konsequenzen weitreichender Befund, wie die Forscher ausführten: »The frequency and magnitude of errors in empirical articles raise serious questions regarding the integrity of the refereeing process of professional journals«. ⁹

In der Folge entwickelte sich eine Debatte darüber, wie diesen Missständen sinnvoll zu begegnen sei. Einige Zeitschriften begannen zunächst damit, eine Selbstverpflichtung der Autor*innen einzufordern. Im Fall von Anfragen nach Daten und Programmcode sollten diese, so die Vorgabe, ihre Daten, Berechnungen und weitere relevante Dateien direkt an interessierte Dritte weitergeben. ¹⁰ In der Praxis führte diese Art von Richtlinie allerdings nicht zu einer verbesserten Reproduzierbarkeit, da die Anreize im Wissenschaftssystem anders gesetzt waren. ¹¹ Im Grundsatz ging und geht es bei solchen (Fehl-)Anreizen immer um ein Kosten-Nutzen-Kalkül auf Seiten der Autor*innen: »[...] producing replicable research is time-consuming and if no one is checking, the rational economist will not make the effort to produce replicable research«. ¹²

Aufgeschreckt durch neue Studien, welche die Wirkungslosigkeit der Selbstverpflichtungspolicies zeigten, ¹³ begannen Top-Journals wie die American Economic Review mit der Einführung von Richtlinien, die eine Datenweitergabe an die Redaktion vor Veröffentlichung einer Publikation vorsahen. ¹⁴ Diese Richtlinien erwiesen sich als wesentlich wirkungsvoller und beförderten zugleich die Frage nach geeigneten Infrastrukturen für die Weitergabe und Publikation solcher Forschungsdaten auf die wissenschaftspolitische Agenda.

Kaum Angebote für das Datenmanagement in der Wirtschaftsforschung: Zeit für Neues

Anfang der 2010er-Jahre gab es in diesem Feld in Europa kaum fachlich passende Infrastruktur- und Serviceangebote für wirtschaftswissenschaftliche Fachzeitschriften: Die Forschungsdateninfrastruktur befand sich (mit Ausnahme der Naturwissenschaften) oft noch in den Kinderschuhen. Forschungsdatenrepositorien wie Pangea ¹⁵ oder Dryad ¹⁶ existierten zwar, kamen aber aus den Naturwissenschaften. Andere Angebote waren kommerziell (wie etwa figshare ¹⁷), und die Zukunftsperspektive dieser neuen Services schien unsicher. Für kleinere Fachzeitschriften waren kommerzielle Angebote zudem oft nicht finanzierbar.

Dort, wo Forschungsdaten in der Sozial- und Wirtschaftsforschung bereitgestellt wurden, wie etwa über die beim Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdatenzentren (FDZ), zählte man die Sammlung und Bereitstellung von publikationsbezogenen Forschungsdaten nicht zum engeren eigenen Auftrag. ¹⁸

Verlage als weitere potenzielle Stakeholder verfügten seinerzeit ebenfalls über keine Infrastruktur zur Speicherung und Bereitstellung solcher Forschungsdaten. ¹⁹ Grundsätzlich betonten zahlreiche STM-Verlage auch bereits sehr früh, dass Forschungsdaten (»raw data«) generell öffentlich und frei verfügbar vorliegen sollten. ²⁰ Dies legt nahe, dass auch Verlage die Speicherung und Bereitstellung von Forschungsdaten nicht zu Ihren primären Aufgaben zählten. Praxis der Verlage war es vielmehr, publikationsbezogene Forschungsdaten als Attachments an den Artikel anzugliedern, ²¹ wodurch diese bisweilen auch hinter die Bezahlschranke fielen.

Auch aufgrund fehlender Infrastrukturen für das Management und die Bereitstellung publikationsbezogener Forschungsdaten unterhielten international nur gut zwei Dutzend Fachzeitschriften in den Wirtschaftswissenschaften eigene Datenarchive. ²² Diese waren oft Eigenentwicklungen und Insellösungen, die an Lehrstühlen oder durch universitäre Einrichtungen betrieben wurden. ²³

Zusammenfassend war die Situation in der Wirtschaftsforschung also dadurch gekennzeichnet, dass es einen hohen Bedarf an Infrastrukturlösungen auf Seiten der Wissenschaft gab, aber kaum geeignete, kostengünstige und einfach zu nutzende Infrastrukturangebote

existierten. Diese Lücke sollte mit der Entwicklung des ZBW Journal Data Archives (JDA) gefüllt werden.

Von der Pilotapplikation zum Service

Das JDA existiert in seiner jetzigen Form seit der Jahreswende 2016/2017. Es handelt sich um eine web-basierte Applikation, welche auf der Open Source Software CKAN²⁴ (Comprehensive Knowledge Archive Network) beruht. Die in Python geschriebene Anwendung wird weltweit vorrangig zur Publikation von Daten aus der öffentlichen Verwaltung als Open Data verwendet. Neben der australischen, kanadischen und der US-Regierung nutzt beispielsweise auch die Stadt Berlin CKAN für die Bereitstellung von Daten der öffentlichen Verwaltung.²⁵ Durch die hohe Anzahl an nutzenden Institutionen hat sich eine rege und lebendige User-Community entwickelt, die vielfältige Erweiterungen und Anpassungen der Software teilt und bei Fragen weiterhilft.²⁶

Um als Datenarchiv für Forschungsdaten im Kontext wissenschaftlicher Publikationsprozesse zu fungieren, musste die Applikation zunächst auf die spezifischen Bedürfnisse der akademischen Publikationsworkflows angepasst werden. Hierfür wurde etwa ein granulares Rechtemanagement implementiert, das die klassischen Nutzungsgruppen im wissenschaftlichen Prozess widerspiegelt: Autor*innen, Herausgeber*innen und Gutachter*innen. Für letztere wurde die Möglichkeit geschaffen eine »double-blind« Begutachtung der eingestellten Daten vorzunehmen. Andere wichtige Erweiterungen umfassten die Nutzung von DataCite DOIs für publizierte Forschungsdaten sowie die Möglichkeit, Daten und zugehörige Publikation wechselseitig mittels persistenter Identifikatoren zu verlinken.

Diese Anpassungen wurden in dem vom ZBW-Leibniz Informationszentrum Wirtschaft geleiteten und der Deutschen Forschungsgemeinschaft (DFG) finanzierten Projekt »European Data Watch Extended (EDaWaX²⁷)« (2012 bis 2016) vorgenommen.

Eng begleitet wurde die softwaretechnische Entwicklung im Projekt von den Herausgeber*innen mehrerer wirtschaftswissenschaftlicher Fachzeitschriften. Die Einbeziehung von Fachzeitschriften war für das Projekt zentral, um eine Infrastruktur zu entwickeln, welche den tatsächlichen Anforderungen der Forschungs- und Publikationspraxis entsprach. Diese praxisnahe Entwicklung der Pilotapplikation wurde unter anderem in zwei Workshops realisiert, an denen Vertreter*innen von mehr als einem Dutzend Fachzeitschriften aus der Community teilnahmen.²⁸ Seit Auslaufen der DFG-Förderung wird das Journal Data Archive aus Eigenmitteln weiterbetrieben.

Dateneinreichungen beim ZBW Journal Data Archive

Um die publikationsbezogenen Daten im Journal Data Archive einzureichen, registriert eine Fachzeitschrift zunächst die verantwortlichen Autor*innen in der Applikation. Nach Aktivierung des Accounts können die

Forschungsdaten zu einem Artikel im Journal Data Archive hochgeladen werden.²⁹

Die Autor*innen machen dabei zunächst Angaben zur Gesamtheit der eingereichten Daten (»Collection«), welche den klassischen bibliografischen Angaben zu Publikationen ähneln: So werden neben allen beteiligten Autor*innen auch Titel, Publikationsjahr, ein kurzes Abstract, JEL (Journal of Economic Literature)-Klassifikationen³⁰ und Keywords zur Dateneinreichung abgefragt.

Bei der Entwicklung der Applikation wurde darauf geachtet, so viele Metadaten wie möglich »im Hintergrund« zu erheben. Dies wurde vor allem durch den Einsatz von Normdaten möglich. Per Lookup Service wurden u.a. Informationen zu Personen und Institutionen aus der Gemeinsamen Normdatei (GND) eingebunden, die bei der Identifikation von Autor*innen und deren Instituten die Erstellung von Metadaten und weiteren verknüpften Informationen ohne weitere Eingaben von Userseite ermöglichen. Nach diesen allgemeinen Informationen zur Gesamtheit der eingereichten Forschungsdaten werden dann im nächsten Schritt die Dateien selbst ins Datenarchiv hochgeladen.

Als Metadatenschema findet bislang ein angepasstes Schema der Forschungsdatenregistratur für Sozial- und Wirtschaftsdaten *dalra*³¹ Verwendung, das ab 2025 auf das Schema von DataCite (DataCite Metadata Working Group, 2024) umgestellt wird.

Je nach Workflow und Forschungsdatenrichtlinie der Fachzeitschrift können nun im nächsten Schritt die Daten als Teil des Begutachtungsverfahrens (double-blind peer-review) herangezogen werden. Das Feedback der Gutachter*innen kann genauso über die Applikation abgewickelt werden, wie eine Kommunikation über die Dateneinreichung zwischen Fachzeitschrift und Autorenteam.

Viele Zeitschriften, so bislang auch die VSWG, fordern die Daten erst von den Autor*innen an, nachdem ein grundsätzliches Ergebnis darüber erzielt wurde, ob der Artikel angenommen und publiziert werden soll. Hier wird entsprechend kein dezidiertes Data Review durchgeführt (was aufgrund der Komplexität des Reviews von Daten häufig von vielen Journals und Gutachter*innen nicht leistbar ist), sondern die Daten werden veröffentlicht und können durch die Community selbst begutachtet werden. Durch die Nachnutzung der Daten werden spätestens dann Probleme offenkundig.

Die Redaktion prüft die hochgeladenen Daten und zugehörige Metadaten auf Plausibilität und dahingehend, ob die zeitschrifteneigene Forschungsdatenrichtlinie erfüllt wurde. Ist dies der Fall, können die Daten mit einem DOI versehen und publiziert werden. Abschließend ergänzt die Redaktion noch einige bibliografische Angaben zur Publikation (Jahrgang, Ausgabe, Seitenzahlen und DOI/URL der Publikation), sodass eine wechselseitige Verlinkung von Artikel und Forschungsdaten gegeben ist.

Das ZBW Journal Data Archive als etablierte Infrastruktur

Auch wenn das Publikationsmodell der Daten im JDA von dem Verfahren klassischer Datenrepositorien abweicht, hat das Modell durchaus Vorteile für Fachzeitschriften: Neben der vollständigen Kontrolle über die Publikation der Daten und der Option, direkt Metadaten und weitere Informationen editieren zu können, ist auch die direkte Kommunikation mit den Autor*innen über die Applikation hilfreich. Darüber hinaus ist das Datenarchiv einer Zeitschrift mit dem jeweiligen Cover und einer Beschreibung des fachlichen Schwerpunkts versehen, sodass die Sichtbarkeit der Zeitschrift gesteigert wird.

Dieses Modell liegt quer zu den von Großverlagen oft empfohlenen disziplinären oder allgemeinen Repositorien, hat aber in der Wirtschaftsforschung zahlreiche Nutzende gefunden. Neben der VSWG sind mit der *German Economic Review* (GER) und den Jahrbüchern für Nationalökonomie / *Journal of Economics and Statistics* (JBNST) zwei weitere Impact Factor-Zeitschriften aus dem deutschsprachigen Raum vertreten. Anfang 2021 kam mit dem *Journal of Applied Econometrics* (JAE) ein Top-Journal aus der Wirtschaftsforschung hinzu, welches eines der am längsten bestehenden Datenarchive im Fach besitzt.³² Insgesamt haben derzeit zehn Fachzeitschriften eine Nutzungsvereinbarung für das Journal Data Archive geschlossen, von denen fünf regelmäßig mit dem Service arbeiten. Das ZBW Journal Data Archive beinhaltet inzwischen mehr als 1.800 Dateneinreichungen zu Publikationen und ist damit eines der größten Datenarchive für wirtschaftswissenschaftliche Fachzeitschriften weltweit. Inzwischen wurden auch für rund ein Dutzend Artikel der VSWG die zugehörigen Forschungsdaten veröffentlicht. Darunter finden sich einige besondere Datensätze, wie die umfangreichen Datensätze des Tetradas-Projekts (Tonnage estimates of trade through the Danish Sound, Warenregistrierungen in den dänischen Sundzollregistern 1670–1856).³³

Emporion – das neue Forschungsdaten-Hub für die Wirtschafts- und Sozialgeschichte

Das Journal Data Archive steht Autor*innen nur in Verbindung mit einem Aufsatz in einer der angeschlossenen wirtschaftswissenschaftlichen bzw. wirtschaftshistorischen Zeitschriften zur Verfügung. Vor diesem Hintergrund und in der Absicht, sowohl den eigenen Mitgliedern als auch der gesamten Disziplin eine Plattform für die Open Access-Publikation von Forschungsdaten und Data Papers zur Verfügung zu stellen, hat das DFG-Schwerpunktprogramm »Experience and Expectations. Historical Foundations of Economic Behavior«³⁴ gemeinsam mit der Gesellschaft für Sozial- und Wirtschaftsgeschichte (GSWG)³⁵ sowie der Staatsbibliothek zu Berlin im März 2023 Emporion (griechisch für Markt- und Handelsplatz einer Stadt) lanciert.³⁶

Dieses neue Forschungsdaten-Hub für die Wirtschafts- und Sozialgeschichte ermöglicht die standardkonforme und namentlich den FAIR-Prinzipien gemäße Veröffentlichung von historischen Zeitreihen, Querschnitts- und Paneldaten, Textmining-Analysen sowie von Data Papers zur detaillierten Beschreibung von Struktur und Inhalt auf Emporion oder an anderer Stelle publizierter Daten.³⁷ Emporion richtet sich primär an Forschende der Wirtschafts- und Sozialgeschichte, unabhängig von ihrer institutionellen Affiliation, und steht zudem für Beiträge aus der Unternehmens-, Umwelt- und Technikgeschichte offen. Sein Angebot ist in publikations- wie rezeptionsseitiger Nutzungsdimension kostenfrei. Für die Akzeptanz von Emporion innerhalb der Fachcommunity auch nach Auslaufen des DFG-Schwerpunktprogramms bürgt die Kooperation mit der GSWG – neben dem Wirtschaftshistorischen Ausschuss des Vereins für Socialpolitik die zweite, zudem auch die Sozialgeschichte umfassende disziplinspezifische Fachgesellschaft im deutschsprachigen Raum.³⁸ Während die Staatsbibliothek zu Berlin – auch als Beitrag zum geschichtswissenschaftlichen Teilbereich der Nationalen Forschungsdateninfrastruktur (NFDI4Memory) – den dauerhaften Betrieb von Emporion sowie die formale Kuratierung der dort veröffentlichten Inhalte sicherstellt, verantwortet die Verbundzentrale des Gemeinsamen Bibliotheksverbunds (GBV) das technische Hosting des Repositoriums, das auf Basis des quelloffenen Software Framework »MyCoRe« entwickelt wurde.³⁹ Um das Kernziel von Emporion, die Erhöhung des Anteils im Open Access verfügbarer wirtschafts- und sozialhistorischer Forschungsdaten, zu befördern, galten Barrierefreiheit und Wissenschaftsadäquanz als leitende Designkriterien. Im Unterschied zu »histat« sind alle auf Emporion veröffentlichten Inhalte daher ohne vorherige Registrierung direkt sichtbar, durch Vergabe von DOIs persistent zitierfähig und zudem, den von der DFG vorgegebenen Leitlinien zur Sicherung guter wissenschaftlicher Praxis entsprechend, für mindestens zehn Jahre garantiert archiviert.⁴⁰

In dem Anliegen, Forschenden die Veröffentlichung ihrer Datensätze oder Datenpapiere möglichst komfortabel zu machen, ist Emporion als Publikationsplattform auch via Shibboleth zugänglich, also unter Verwendung der durch eine andere Wissenschaftseinrichtung authentifizierten persönlichen Login-Daten.⁴¹ Durch automatische Meldung neuer Datenpublikationen an das Forschungsinformationssystem »OpenAIRE« entlastet Emporion aus Mitteln der Europäischen Union finanzierte Projekte von ihren entsprechenden Berichtspflichten.⁴² Bei Zweitveröffentlichung eines zuvor bereits im Closed Access erschienenen Datenpapiers reduziert Emporion ebenfalls den Aufwand der Erfassung bibliografischer Metadaten durch ihre Direktübernahme aus Datenbanken wie »Scopus« oder »K10plus«. Nicht zuletzt aufgrund seiner Konformität mit dem international etablierten sozialwissenschaftlichen Metadatenstandard

»Data Documentation Initiative« (DDI) ist die Interoperabilität von Emporion mit anderen bibliothekarischen Nachweissystemen ohne Reibungsverluste gewährleistet – in Hinblick auf Datenimporte wie -exporte.⁴³ Auch in der umgekehrten Dimension von Benutzung erleichtert Emporion den Datenzugang, sind doch seine mit Geo-Koordinaten markierten Inhalte mithilfe einer grafischen Kartensuche zu recherchieren. Schließlich erlauben es mehrere konventionelle und altmetrische Statistikinstrumente, die Rezeption der auf Emporion veröffentlichten Inhalte nachzuvollziehen – unter Einschluss der Interaktionen mit ihnen im Social Web. Indexiert werden die Datensätze und Data Papers neben Google unter anderem von den Suchmaschinen »DataCite Search«⁴⁴ und »BASE«.⁴⁵

Mit Emporion soll einem hochgradig interdisziplinären und gleichermaßen von qualitativen wie quantitativen Methoden geprägten Kleinen Fach ein zentraler Datenmarktplatz zur Verfügung gestellt werden, um auf diese Weise den integrativen Impuls des Schwerpunktprogramms »Experience and Expectations« zu perpetuieren, in dem über sieben Jahre (2016–2023) 31 Projekte aus ganz Deutschland mitwirkten.⁴⁶ Der Staatsbibliothek zu Berlin eröffnet diese Kooperation wiederum die Chance, sich als eigenständige Akteurin und Infrastrukturservicepartnerin der Forschung auf einem weiteren geschichtswissenschaftlichen Feld zu etablieren, dem ihre historischen Sammlungen einen weiten Resonanzraum bieten. Das damit angesprochene Potenzial dokumentiert nicht zuletzt ein unlängst von der DFG bewilligtes Projekt zur Digitalisierung und Open-Access-Transformation der in der Staatsbibliothek zu Berlin in herausragender Zahl verwahrten Jahresberichte deutscher Handelskammern des »langen« neunzehnten Jahrhunderts.⁴⁷ Da diese von der Forschung weitgehend übersehene Quellengattung neben einer verbalen Einschätzung der wirtschaftlichen Lage des jeweiligen Handelskammerbezirks im Berichtsjahr auch einen standardisierten umfangreichen Statistik-anhang umfasst, dürfte Emporion insofern keinen Mangel an künftigen Datenzugängen leiden.

Umgekehrt verspricht Emporion auch neue Impulse für die Forschung, gehören doch die Suche nach den »scarce and scattered numerical traces of the past«, ihre Vereinheitlichung und Digitalisierung zu den aufwendigsten Teilen der Wirtschafts- und Sozialgeschichtsschreibung.⁴⁸ Nach wie vor erfolgt die Veröffentlichung bereits digitalisierter Daten, wenn überhaupt, an ganz unterschiedlichen, für andere Forschende oft kaum nachvollziehbaren Orten. Die Zusammenführung historischer Daten auf einer zentralen und allgemein zugänglichen Plattform kann zukünftige Forschung daher enorm erleichtern und darüber hinaus dazu beitragen, strukturelle Ungerechtigkeiten im Fach abzubauen. Denn die Transkription wirtschafts- und sozialhistorischer Daten ist ebenso zeit- wie kostenintensiv – ob sie nun durch die Forschenden selbst (bzw. ihr Team) oder durch externe

Dienstleistungsunternehmen im Ausland durchgeführt wird. Ein zentraler Umschlagplatz – ein Emporion – für historische Wirtschafts- und Sozialdaten im Open Access ermöglicht es demgegenüber auch Forschenden mit knapperen Ressourcen, empirisch gesättigte Studien durchzuführen. Es ist also nicht unwahrscheinlich, dass offene wissenschaftliche Informationsinfrastrukturen wie Emporion perspektivisch auch die Strukturen und Machtgefüge von Forschung verändern werden – mithin Wandel durch freien Datenhandel.

Ausblick

Teile der Geschichtswissenschaft und insbesondere diejenigen historischen Subdisziplinen, die sich mit sozioökonomischen Fragen und der Geschichte des Alltags beschäftigen, stehen vor einem grundlegenden Wandel. Erstens wird es schon in wenigen Jahren möglich sein, die vielen statistischen Quellenwerke, die seit Beginn des 19. Jahrhunderts publiziert wurden, (semi-) automatisiert einzulesen, und zwar nicht nur als unstrukturierter Fließtext, sondern strukturiert, eben als Tabelle. Dies erlaubt den Zugriff auf zum Teil sehr detaillierte Daten, die zu erheben bislang viel zu arbeitsaufwendig gewesen wäre. Zweitens werden im Zuge der Digitalisierung publizierter Literatur, aber auch archivalischer Quellen massenhaft Forschungsdaten im textlichen Bereich anfallen, die mit Text Mining-Verfahren untersucht werden können. Drittens steht zu erwarten, dass die isolierten Datenrepositorien, von denen hier die beiden neuesten für die Wirtschafts- und Sozialgeschichte vorgestellt wurden, künftig über eine Metastruktur (»data space«) recherchierbar werden.⁴⁹ Viertens, und das ist vielleicht der bedeutendste Fortschritt der letzten Jahr(zehnt)en, stehen im Internet zu keinen oder geringen Kosten KI-Programme zur Verfügung, die die Auswertung solcher Datenbestände erheblich erleichtern werden. Die Transformation von (numerischen und textlichen) Daten in selbst entworfene Datenstrukturen wird die Nutzung statistischer Verfahren drastisch vereinfachen. Schon heute kann man ChatGPT mit geeigneten Prompts instruieren, Skripte, etwa in Python, zu schreiben, die man ohne jede Programmierkenntnisse laufen lassen kann. Der Stellenwert quantitativer Forschung in der Geschichtswissenschaft wird daher deutlich zunehmen. Die in den letzten Jahren geschaffenen neuen fachspezifischen Datenrepositorien stellen jetzt schon die Forschungsdateninfrastruktur dafür bereit.

Anmerkungen

- 1 Vgl. etwa <https://www.zbw.eu/de/forschungsdaten/infrastrukturen/erfahrungen-im-bauen-von-infrastrukturen/digitale-reichsstatistik>; <https://www.bib.uni-mannheim.de/ihre-ub/projekte-der-ub/aktienfuehrer-datenarchiv/>; <https://www.bib.uni-mannheim.de/ihre-ub/projekte-der-ub/aktienfuehrer->

- datenarchiv-ii/; https://www.diw.de/de/diw_01.c.841692.de/projekte/wochenbericht_digital_in_wort_und_zahl_wbdigital.html [alle Zugriff am: 07.11.2024].
- 2 Vgl. etwa Bartels, Charlotte. Top Incomes in Germany, 1871–2014. *Journal of Economic History*. 2019, 79 (4), 669–707. Verfügbar unter: DOI: 10.1017/S0022050719000378
 - 3 Vgl. <https://de.statista.com/statistik/daten/studie/4878/umfrage/bruttoinlandsprodukt-von-deutschland-seit-dem-jahr-1950/> [Zugriff am: 05.11.2024]; ferner Ritschl, Albrecht und Mark Spoerer. Das Bruttosozialprodukt in Deutschland nach den amtlichen Volkseinkommens- und Sozialproduktsstatistiken 1901–1995. *Jahrbuch für Wirtschafts-geschichte*. 1997, (2), 27–54. Verfügbar unter: DOI: 10.1524/jbwg.1997.38.2.27; Burhop, Carsten und Guntram Wolff. A compromise estimate of the net national product and its implications for growth and the business cycle in Germany, 1851–1913. *Journal of Economic History*. 2005, 65 (3), 613–657. Verfügbar unter: DOI: 10.1017/S0022050705000239; Rahlf, Thomas. Volkswirtschaftliche Gesamtrechnung. In: ders., Hrsg. *Deutschland in Daten. Zeitreihen zur Historischen Statistik*. 2. Aufl. Bonn: Bundeszentrale für politische Bildung, 2022, S. 192–205. Verfügbar unter: <https://www.bpb.de/shop/buecher/zeitbilder/515397/deutschland-in-daten/>
 - 4 Vgl. etwa Tooze, Adam J. *Statistics and the German State, 1900–1945. The Making of Modern Economic Knowledge*. Cambridge, et al.: Cambridge University Press, 2001; Fremdling, Rainer und Reiner Stäglin. Verschleierung mit Statistik: Kriegswirtschaftliche Desinformation im Nationalsozialismus. *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*. 2012, 99 (3), 323–335. Verfügbar unter: <https://www.jstor.org/stable/41636878>; Schneider, Michael C. *Wissensproduktion im Staat. Das königlich preußische statistische Bureau 1860–1914*. Frankfurt am Main: Campus, 2013.
 - 5 <https://histat.safe-frankfurt.de/>. Vgl. zur Geschichte von »histat«: Franzmann, Gabriele. The Online-Database histat as an Example for Research-Promoting Infrastructures for Studies in Quantitative Historical Research. *Économies et Sociétés, Série »Histoire économique quantitative«*. 2015, n° 50, 821–856. Verfügbar unter: https://histat.safe-frankfurt.de/assets/Online-DB-histat_Economies-et-Societes_6-2015.pdf
 - 6 Vgl. <https://4memory.de/>, <https://www.berd-nfdi.de/> und <https://www.konsortswd.de/>. Vgl. allenfalls für individual-spezifische Mikrodaten: <https://www.geschichte.uni-halle.de/struktur/hist-data/>
 - 7 Huang, Angela; Chilosi, David und Alexandra Sapoznik. A Source Collection on Urban Annuities, 14th–18th Centuries: An Introduction to the Data. *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*. 2019, 106 (1), 67–80. Verfügbar unter: <http://www.jstor.org/stable/45181715>
 - 8 Dewald, William G.; Thursby, Jerry G. und Richard G. Anderson. Replication in empirical economics: The journal of money, credit and banking project. *American Economic Review*. 1986, 76 (4), 587–603.
 - 9 Dewald et al., 1986, 588.
 - 10 Vgl. McCullough, B. D. Got Replicability? The Journal of Money, Credit and Banking Archive. *Econ Journal Watch: Scholarly Comments on Academic Economics*. 2007, 4 (3), 326–337.
 - 11 Vgl. Feigenbaum, S. und D. Levy. The market for (Ir)reproducible econometrics. *Accountability in Research*. 1993, 3 (1), 25–43.
 - 12 McCullough, B. D. Open Access Economics Journals and the Market for Reproducible Economic Research. *Economic Analysis and Policy*. 2009, 39 (1), 117–126, hier 119.
 - 13 McCullough, B. D. und H. Vinod. Verifying the Solution from a Nonlinear Solver. *American Economic Review*. 2003, 93 (3), 873–892.
 - 14 Bernanke, B. S. Editorial Statement. *American Economic Review*. 2004, 94 (1), 404.
 - 15 <https://www.pangaea.de/> [Zugriff am: 10. Oktober 2024]; vgl. Felden, J.; Möller, L.; Schindler, U.; Möller, R.; Schumacher S.; Koppe R.; Diepenbroek, M. und F. Glöckner. PANGAEA – Data Publisher for Earth & Environmental Science. *Sci Data*. 2023, 10, 347. Verfügbar unter: <https://doi.org/10.1038/s41597-023-02269-x>
 - 16 <https://datadryad.org/stash> [Zugriff am: 10. Oktober 2024]; vgl. Vision, T. J. Open Data and the Social Contract of Scientific Publishing. *BioScience*. 2010, 60 (5), 330–331. Verfügbar unter: <https://doi.org/10.1525/bio.2010.60.5.2>
 - 17 <https://figshare.com/> [Zugriff am: 10. Oktober 2024].
 - 18 Vlaeminck, S. und G. G. Wagner. Zur Rolle von Forschungsdatenzentren beim Management von publikationsbezogenen Forschungsdaten – Ergebnisse einer Befragung von wissenschaftlichen Infrastrukturdienstleistern im Bereich der Sozial- und Wirtschaftswissenschaften. *Zeitschrift für Bibliotheks- und Bibliographie*. 2014, 61 (2), 76–84. Verfügbar unter: <https://doi.org/10.3196/186429501461235>
 - 19 Erst mit dem Kauf der Open Access Plattform Mendeley 2013 und der darauffolgenden Implementierung von Mendeley Data 2016 hat zumindest Elsevier ein eigenes Data Repository am Markt positioniert, vgl. <https://blog.mendeley.com/2016/04/28/mendeley-data-is-out-of-beta/> [Zugriff am: 10. Oktober 2024].
 - 20 STM-Publishers (2007): Brussels Declaration on STM Publishing. [Zugriff am: 10. Oktober 2024]. Verfügbar unter: http://www.stm-assoc.org/2007_11_01_Brussels_Declaration.pdf
 - 21 Vlaeminck, S. Data Management in Scholarly Journals and Possible Roles for Libraries–Some Insights from EDaWaX. *Liber Quarterly*. 2013, 23 (1), 48–79.
 - 22 Duvendack, M.; R. Palmer-Jones und R. Reed, Replications in Economics: A Progress Report. *Econ Journal Watch – Scholarly Comments on Academic Economics*. 2015, 12 (2), 164–191.
 - 23 Eines der am besten umgesetzten Beispiele einer einzelnen Fachzeitschrift ist das Legacy Data Archive des Journal of Applied Econometrics (JAE), welches unter der URL <http://qed.econ.queensu.ca/jae/legacy.html> aufrufbar ist [Zugriff am: 10. Oktober 2024].
 - 24 <https://ckan.org/> [Zugriff am: 10. Oktober 2024].
 - 25 Vgl. <https://daten.berlin.de/datensaetze> [Zugriff am: 10. Oktober 2024].
 - 26 <https://ckan.org/community> [Zugriff am: 10. Oktober 2024].
 - 27 <https://www.edawax.de/about/> [Zugriff am: 10. Oktober 2024].
 - 28 <https://idw-online.de/de/news563132>. Infolge des Feedbacks mussten einige Entwicklungsstränge verworfen werden. Die Zahl der Metadatenfelder wurde beispielsweise so stark wie möglich begrenzt und fiel je nach » resource type « unterschiedlich umfangreich aus. Die Applikation unterscheidet in Data, Text, Code und Other als mögliche resource types. Datensätze benötigen deutlich mehr (und andere) Metadaten für eine etwaige Nachnutzung, als dies beispielsweise bei Text-, Code- oder Bilddokumenten der Fall ist. Das volle Metadaten-schema ist unter <https://journaldata.zbw.eu/info/> verfügbar.
 - 29 Der komplette Ablauf einer Einreichung aus Sicht der Autor*innen ist unter <https://www.youtube.com/watch?v=vyK-44f3X9s> verfügbar [Zugriff am: 10. Oktober 2024].
 - 30 Das JEL Classification System ist unter <https://www.aeaweb.org/econlit/jelCodes.php?view=jel> verfügbar.
 - 31 Koch, U.; Akdeniz, E.; Meichsner, J.; Hausstein, B. und K. Harzenetter. *dalra Metadata Schema. Documentation for the Publication and Citation of Social and Economic Data*. Version:

- 4.0. GESIS – Leibniz Institute for the Social Sciences, 2017. Verfügbar unter: <https://doi.org/10.4232/10.mdsdoc.4.0>
- 32 Das Legacy Data Archive des Journal of Applied Econometrics (JAE), ist unter der URL <http://qed.econ.queensu.ca/jae/legacy.html> weiterhin erreichbar [Zugriff am: 10. Oktober 2024].
- 33 <https://www.uni-bamberg.de/digihist/projekte/tetradas/> [Zugriff am: 10. Oktober 2024]; Scheltjens, Werner. Tetradas: Dataset of Tonnage Estimates of Trade through the Danish Sound, 1670–1856. *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*. 2021, 108 (3), 373–394. Verfügbar unter: DOI: 10.25162/vswg-2021-0012.
- 34 <https://www.experience-expectation.de/>
- 35 <https://www.gswg.eu/>
- 36 <https://emporion.gswg.info/>
- 37 <https://www.go-fair.org/fair-principles/>
- 38 <https://economic-history.committee.socialpolitik.de/>
- 39 <https://www.mycore.de/>
- 40 <https://doi.org/10.5281/zenodo.6472827>
- 41 <https://www.shibboleth.net/>
- 42 <https://www.openaire.eu/>
- 43 <https://ddialliance.org/>
- 44 <https://commons.datacite.org/>
- 45 <https://www.base-search.net/>
- 46 Vgl. Lenel, Laetitia; Nützenadel, Alexander; Streb, Jochen und Ingo Köhler (Hrsg.). *Routledge Handbook of Economic Expectations in Historical Perspective*. London: Routledge, 2025.
- 47 Vgl. <https://blog.sbb.berlin/handelskammerberichte/>
- 48 Gutmann, Myron P., Emily Klancher Merchant, und Evan Robert. »Big Data« in Economic History. *The Journal of Economic History*. 2018, 78 (1), 268–299. Verfügbar unter: <https://doi.org/10.1017/S0022050718000177>
- 49 Vgl. <https://4memory.de/task-areas/task-area-2-data-connectivity/> [Zugriff am: 05.11.2024].

Verfasser*Innen



Prof. Dr. Laetitia Lenel, Juniorprofessorin für Kulturgeschichte des Ökonomischen, Universität Duisburg-Essen, Historisches Institut, Universitätsstraße 12, 45141 Essen, laetitia.lenel@uni-due.de
Foto: Alexandra Roth



Dr. Christian Mathieu, Forschungsreferent, Generaldirektion, Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Potsdamer Straße 33, 10785 Berlin, christian.mathieu@sbb.spk-berlin.de
Foto: privat



Prof. Dr. Mark Spoerer, Inhaber des Lehrstuhls für Wirtschafts- und Sozialgeschichte, Institut für Geschichte, Universität Regensburg, 93040 Regensburg, mark.spoerer@geschichte.uni-regensburg.de
Foto: Roswitha Geiger



Sven Vlaeminck, Wissenschaftlicher Mitarbeiter, Forschungsdaten-Beauftragter, ZBW-Leibniz Informationszentrum Wirtschaft, Neuer Jungfernstieg 21, 20354 Hamburg, s.vlaeminck@zbw.eu
Foto: privat