

## **Beschreibung des Vorhabens – Projektanträge im Bereich „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS)**

### **LIS-Förderprogramm oder Ausschreibung: Infrastruktur für elektronische Publikationen und digitale Wissenschaftskommunikation**

Reiner Diedrichs, Göttingen (Verbundzentrale Göttingen; VZG)  
Wolfram Horstmann, Göttingen (Niedersächsische Staats- und Universitätsbibliothek Göttingen; SUB)  
Isabella Peters & Klaus Tochtermann, Kiel/Hamburg (Leibniz-Informationszentrum Wirtschaft; ZBW)  
Katrin Weller, Markus Strohmaier & Christof Wolf, Köln/Mannheim (GESIS Leibniz-Institut für Sozialwissenschaften)

---

## **Beschreibung des Vorhabens**

### **1 Ausgangslage und eigene Vorarbeiten**

Die wissenschaftliche Kommunikation unterliegt einem Wandel, der sich insbesondere in zwei Aspekten bemerkbar macht: im Publikationsmedium als Grundlage wissenschaftlicher Kommunikation sowie in der Bewertung von wissenschaftlicher Leistung durch die Fachcommunity und andere interessierte Parteien. Beide sind massiv von der Prävalenz des Digitalen und des Internets beeinflusst, wobei vor allem Social-Media-Plattformen den Umgang mit (wissenschaftlichen) Inhalten beeinflussen. Internetnutzende sind nun nicht mehr nur reine Konsumierende von digitalen Inhalten, sondern stellen sie als „Prosumer“ (Toffler, 1980) oder „Produser“ (Bruns, 2008) zunehmend selbst her, und auch Forschende nutzen ausgiebig solche Social-Media-Plattformen, die nicht ausschließlich für wissenschaftliche Tätigkeiten und Kommunikation konzipiert wurden. Der wissenschaftliche Forschungsalltag ist ohne Digitalisierung und Internet kaum mehr denkbar und wird häufig als „Science 2.0“ oder „Open Science“ bezeichnet. Diese Konzepte erkennen die veränderten Nutzungsgewohnheiten und Technologiebedürfnisse an, die durch das Internet oder Social Media in der Wissenschaft entstanden sind.

Mit den neuen Publikationswegen und -formen ergeben sich erweiterte Möglichkeiten und Notwendigkeiten der Messung der Wirkung – des Impacts – eines wissenschaftlichen Ergebnisses. Neben Zitationen und Nutzungsstatistiken bieten sich Social-Media-Plattformen als Basis an, da sie u.U. gleichzeitig die Publikation sowie den Bewertungsmechanismus als Funktionalität bereitstellen. Diese Social-Media-basierten Bewertungen von wissenschaftlichen Inhalten werden als „altmetrics“ (Priem, 2013; siehe auch <http://altmetrics.org/manifesto>), „article-level metrics“ (Fenner, 2013) oder „social media metrics“ (Haustein, Larivière et al., 2014) bezeichnet. Social-Media-Metriken treten als Alternative (daher „alt-metrics“) oder, wie zahlreiche Studien gezeigt haben (u.a. Haustein, Larivière, et al., 2014; Haustein, Peters et al., 2014), als Ergänzung neben die klassischen zitationsbasierten Indikatoren zur Messung des Einflusses wissenschaftlicher Leistung (wie H-Index, Impact Factor, Peer Review) und liefern so zusätzliche Anhaltspunkte zur Qualität und Wirkung einer Publikation in der wissenschaftlichen und allgemeinen Öffentlichkeit (Priem, Piwowar, & Hemminger, 2012). Dies wird von Initiativen wie der San Francisco Declaration on Research Assessment (DORA, 2012) und Forschenden (Holtz & Herb, 2013) begrüßt. Es wird über Zitationen zudem nur der Impact auf andere Autoren abgebildet, was in einer Vernachlässigung der Wirkung wissenschaftlicher Forschung auf die Gesellschaft oder Politik resultiert. Da das Internet und Social Media ohne Gatekeeper funktionieren, wird ihnen das Potential zugeschrieben, die Meinungen verschiedener Öffentlichkeiten widerzuspiegeln und so das Bild von der Wirkung und Relevanz wissenschaftlicher Arbeit zu komplettieren.

Bibliotheken und Informationsinfrastrukturen können ebenfalls von den Social-Media-Metriken profitieren, da durch sie zusätzliche Indikatoren für Qualität oder Aktualität von Literatur zur Verfügung stehen. Diese unterstützen potentiell die Erwerbung und das Bestandsmanagement sowie auf Nutzerseite die Selektion von „lesenswerten“ Inhalten. Denkbar ist die Verbindung von Social-Media-Metriken und Patron Driven Acquisition sowie Initiativen zur Bestandsvervollständigung, um Erwerbungsentscheidungen fundiert – d.h. in diesem Fall crowd-basiert – zu treffen. Die Vorteile der Social-Media-Metriken liegen dabei vor allem in ihrer Generalität und Aktualität, die auch Publikationsformen und -sprachen berücksichtigen, die bislang nicht oder unzureichend von Bibliotheken beachtet wurden. Mit Social-Media-Metriken sind unmittelbare und meistens lizenz- und kostenfreie quantitative Indikatoren für zahlreiche Publikationsformen vorhanden, die eine zügige Anpassung an die Entwicklungen im Social Web und veränderte Nutzungsgewohnheiten erlauben (Adam, 2014).

Die Verlage, die gleichsam als Moderatoren und Vermittler des wissenschaftlichen Diskurses wirken, reagieren bereits zunehmend auf diese Entwicklungen und veröffentlichen neben Zitations- und Nutzungszahlen auch Social-Media-Metriken für ihre Publikationen (z.B. Springer). Dabei greifen sie hauptsächlich auf zwei kommerzielle Aggregatoren zurück: Altmetric.com (Anteilseigner: Macmillan Publishers) und PlumX (Tochtergesellschaft von EBSCO). Open Source-Software (z.B. Lagotto) wird von der Public Library of Science (PLoS) für ihre eigenen Publikationen und von ImpactStory eingesetzt, wobei letztere als Zielgruppe eher Individual-Nutzende hat, die Social-Media-Metriken zur Selbstpräsentation einsetzen wollen.

An dieser Stelle setzt das Projektvorhaben „\*metrics – MEasuring The Reliability and perceptions of Indicators for interactions with sCientific productS“ an und lässt sich von folgender Forschungsfrage leiten: **Welche Aussagekraft haben \*metrics und wie werden sie von den Nutzenden interpretiert?** Als \*metrics werden alle Indikatoren verstanden, die eine Interaktion mit einem wissenschaftlichen Produkt widerspiegeln (d.h., Zitationen, Nutzungs- und Ausleihstatistiken, Social-Media-Metriken, etc.). Da die Bandbreite von Social-Media-Metriken größer ist als bei anderen Indikator-Bündeln, ergibt sich hier eine Konzentration innerhalb des Projektvorhabens. Ziel ist es, die Reliabilität und Wahrnehmung der so generierten Indikatoren zu bestimmen.

Diese Schwerpunktsetzung folgt der Empfehlung der Gutachter zur ersten Einreichung dieses Antrags, die mehrheitlich ein tieferes Verständnis von \*metrics auf Basis einer explorativen Studie angeraten haben. Die Exploration zielt vor allem auf die Erforschung der Herkunft und der Wahrnehmung von \*metrics. Bei der Einschätzung der Reliabilität und Quelle der \*metrics ist man beim Einsatz von Social-Media-Metriken mit drei Herausforderungen konfrontiert:

Zunächst muss Zugang und Verfügbarkeit auf den verschiedenen Social-Media-Plattformen evaluiert werden, die über sog. Application Programming Interfaces (APIs) die Daten (oder auch nur einen Teil davon) zur Verfügung stellen und damit grundlegend den Umgang mit und die Aussagekraft von Social-Media-Metriken prägen. Ähnliche Schwierigkeiten sind bei Nutzungsstatistiken zu erwarten, etwa bei der Berechnung der Views. Auch wenn das DFG-geförderte Projekt Open-Access-Statistik zu einer gesteigerten Implementierung von Standards zur Bestimmung der Nutzung elektronischer Dokumente bei Repositorien und teils auch Open-Access-Journalen führte, werden diese Standards bei weitem nicht flächendeckend und in allen Publikationsdiensten (z.B. Git) angewandt. Diese fehlende Nutzung existierender Standards erschwert jedoch die Vergleichbarkeit der Downloadstatistiken und die Eliminierung maschineller Zugriffe (Herb et al., 2009), sodass die Statistiken oft ein verzerrtes Bild der tatsächlichen Nutzung wiedergeben. Twitter regelt bspw. die Zugriffsraten auf die API und stellt nur 1% des gesamten Tages-Twittervolumens zum Download zur Verfügung, für den Rest muss eine beträchtliche Summe gezahlt werden. Dies wirkt sich unmittelbar auf tweet-basierte Social-Media-Metriken aus, da die Eigenschaften der im Download fehlenden Tweets unbekannt sind. Somit ist auch keine korrekte Hochrechnung auf das Gesamtvolumen möglich; Aussagen über alle Twitter-Nutzende oder Tweets können nicht getroffen werden.

Als zweites werden die \*metrics von den Aggregatoren beeinflusst, da sie die gesuchten Daten (z.B. eine Menge an DOIs) auf den (Social-Media-)Plattformen lokalisieren, herunterladen und zusammenfassen. Dabei sind die verwendeten Such- und Aggregationsalgorithmen oftmals intransparent, so dass nicht ohne weiteres nachvollzogen werden kann, wie die Werte der \*metrics zustande kommen. Die Erstellung der Social-Media-Metriken erfolgt zudem häufig nur identifier-basiert und ohne die Hinzunahme von textuellen Informationen zum gesuchten Dokument (z.B. bibliographische Angaben). Deswegen und weil dieselben Publikationen unterschiedliche Identifier haben können, ist auch hier keine Vollständigkeit der \*metrics zu erwarten und man wird bei der Nutzung der \*metrics, die Ausprägung der Social-Media-Interaktionen (z.B. Anzahl der Tweets) in Korrelation mit der tatsächliche Menge von gefundenen Identifiern (i.e. Abdeckungsgrad) eher unterschätzen. Im Projekt werden daher (bibliothekarische) Verfahren erprobt (z.B. Deduplizierung, Textmining), um eine möglichst vollständige Datengrundlage für die \*metrics-Generierung herzustellen und den Anteil an fehlenden Informationen bei den Anbietern von \*metrics beziffern zu können.

Eine dritte Herausforderung ergibt sich aus der Frage nach dem Urheber der Interaktion mit dem wissenschaftlichen Produkt und typischen Verhaltensformen, die sich in den \*metrics widerspiegeln. Insbesondere das Verständnis der Motivation zur Nutzung der Plattformen und ihrer Funktionalitäten spielt hier eine Rolle und macht eine Einordnung und Evaluation der \*metrics erst möglich. Diese Aspekte sind speziell für Forschende und wissenschaftliche Publikationen noch nicht geklärt, wobei disziplinspezifische und demographische Ausprägungen, wie bei der wissenschaftlichen Social-Media-Nutzung (Dzeyk, 2013; Mazarakis & Peters, 2015; Mohammadi et al., 2015), zu erwarten sind. Das Wissen um

die Herkunft der \*metrics ist in Social-Media-Umgebungen sowie bei Nutzungsstatistiken ungleich wichtiger als bei der Herkunft von Zitaten, da nutzergenerierter Content im Social Web keiner Qualitätsprüfung unterzogen wird. Werden Social-Media-Metriken dann als Proxies für die Güte von Forschung angesehen, geschieht dies gewissermaßen auf Basis ungeprüfter Qualität. Das Projektvorhaben will diese Leerstelle schließen.

Auch die Wahrnehmung der Social-Media-Aktivitäten (z.B. liken) ist für die Evaluation der Social-Media-Metriken sowie ihrer Visualisierungen und Aggregationen (z.B. ResearchGate Score) entscheidend. Welche Auswirkungen haben verschiedene Visualisierungstypen (z.B. Ranglisten, ImpactStory's Perzentile, 'Donuts' bei altmetric.com) und Scores auf die Entscheidungen und zukünftigen Handlungen der Nutzenden? Helfen diese Informationen bei der Einschätzung der Relevanz oder Qualität einer wissenschaftlichen Publikation oder ist der Interpretationsaufwand zu hoch? Da letztlich die Nutzenden den \*metrics ihren Stellenwert zuweisen (sehr wahrscheinlich disziplin- und gruppenspezifisch, z.B. gender-basiert), ergibt sich mit der Untersuchung an dieser Stelle ein weiterer Schwerpunkt des Projekts.

Da die kommerziellen \*metrics-Anbieter noch nicht offen und transparent agieren, so dass das Auffinden und die Aggregation der \*metrics von den Nutzenden nicht fundiert nachvollzogen und bewertet werden können, soll das Projektvorhaben auch dazu beitragen, Sensibilität für Offene Metriken und Daten – also open \*metrics – zu schaffen und Anreize für die Auseinandersetzung mit der Thematik geben. Hier soll insbesondere die Bibliothekscommunity einbezogen werden, um z.B. im Rahmen von Informationskompetenzschulungen auf den sinnvollen Umgang mit und Einsatz von \*metrics hinzuweisen. Dieses Vorgehen dient auch der Erhöhung der Sichtbarkeit des Projekts und macht es als Akteur in der Diskussion und (Weiter-) Entwicklung von \*metrics für alle Interessengruppen wichtig. Im Projekt werden die Entwicklungen der kommerziellen \*metrics Anbieter nicht kopiert, sondern durch die Expertise der Projektteilnehmer ergänzt.

Das Projekt \*metrics profitiert von den Vorarbeiten und Erfahrungen in der Zusammenarbeit der Projektpartner SUB und VZG aus dem DFG-Projekt Open-Access-Statistik (OA-Statistik). OA-Statistik ist eine Infrastruktur zur zentralen Auswertung und Bereitstellung von standardisierten Nutzungszahlen digitaler Dokumente und wird seit 2014 als funktionaler Dienst durch die Verbundzentrale des GBV angeboten. Als Katalogisierungs- und Dienstleistungszentrum für wissenschaftliche und öffentliche Bibliotheken und für Forschungseinrichtungen in sieben Bundesländern, beschäftigt sich die VZG tagtäglich mit Datenvorhaltung, -deduplizierung und der Entwicklung neuartiger Bibliotheks- und Informationsdienstleistungen. Für eine internationale Vergleichbarkeit der Nutzungsstatistiken kooperierte OA-Statistik eng mit internationalen Partnern. Dazu war bzw. ist der Projektpartner SUB in verschiedenen internationalen Gremien Mitglied bzw. leitet eine internationalen Arbeitsgruppe. Darüber hinaus hat die SUB die wissenschaftliche Leitung bei dem EU-Projekt OpenAIRE, in dem Nutzungszahlen eine wichtige Rolle spielen. Durch ZBW-Expertise im Bereich Social Media und Altmetrics in der Person von Prof. Dr. Isabella Peters wird das bereits aus dem Projekt OA-Statistik vorhandene inhaltliche und technische Know-How für den vorliegenden Antrag maßgeblich erweitert. Die Kompetenz liegt vor allem in den eigenen Forschungsarbeiten zur Aussagekraft und Theoretisierung von \*metrics für verschiedene Publikationstypen (Haustein, Peters et al., 2014; Peters et al., 2015), zu den Funktionalitäten und der Reliabilität von \*metrics-Aggregatoren (Jobmann et al., 2014), und zur wissenschaftlichen Nutzung von Social Media (Mazarakis & Peters, 2015) sowie in der starken Einbindung in die wissenschaftliche \*metrics-Community (z.B. ASIS&T SIGMetrics ). Die ZBW ist zudem Initiator und aktiver Teilnehmer des Leibniz-Forschungsverbands „Science 2.0“. Mit GESIS konnte ein weiterer Projektpartner gewonnen werden, der sich ebenfalls im Forschungsverbund Science 2.0 an der Erforschung von wissenschaftlicher Kommunikation beteiligt. GESIS verfügt mit der Abteilung Computational Social Science über fundierte Expertise in der Analyse von Social-Media-Daten und anderen Internetdaten für wissenschaftliche Fragestellungen (z.B. aus Twitter [Lietz et al., 2014], Wikipedia [Laufer et al., 2015] oder Reddit [Singer et al., 2014]). Die große inhaltliche und technische Expertise der Projektpartner sowie die erfolgreiche Zusammenarbeit in vorherigen Projekten ermöglichen eine umfangreiche und tiefe wissenschaftliche Bearbeitung des Themas.

## 1.1 Projektbezogene Publikationen

### 1.1.1 Veröffentlichte Arbeiten aus Publikationsorganen mit wissenschaftlicher Qualitätssicherung, Buchveröffentlichungen sowie bereits zur Veröffentlichung angenommene, aber noch nicht veröffentlichte Arbeiten

JOBMANN, A. ; HOFFMANN, C.P. ; KÜNNE, S. ; PETERS, I. ; SCHMITZ, J. ; WOLLNIK-KORN, G. (2014): *Altmetrics for large, multidisciplinary research groups: Comparison of current tools*. Bibliometrie - Praxis und Forschung, 3. URL: <http://www.bibliometrie-pf.de/article/view/205/258>.

HAUSTEIN, S. ; PETERS, I. ; SUGIMOTO, C.R. ; THELWALL, M. ; LARIVIERE, V. (2013): *Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature*. Journal of the Association for Information Science and Technology, 65(4), 656–66. DOI: [10.1002/asi.23101](https://doi.org/10.1002/asi.23101).

MAZARAKIS, A. ; & PETERS, I. (2015). *Quo Vadis German Scholarly Communication in Economics? An Analysis on the Use of Social Media*. Economics. In Proceedings of Workshop on The Future of Scholarly Communication in Economics, Hamburg, Germany. URL: <http://hdl.handle.net/10419/110679>.

PETERS, I. ; KRAKER, P. ; LEX, E. ; GUMPENBERGER, C. ; GORRAIZ, J. (2015). *Research Data Explored: Citations versus Altmetrics*. Proceedings of the 15th International Conference on Scientometrics and Informetrics, Istanbul, Turkey. URL: <http://arxiv.org/abs/1501.03342>.

WELLER, K. ; DRÖGE, E. ; PUSCHMANN, C. (2011). *Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences*. In M. Rowe, M. Stankovic, A.-S. Dadzie, & M. Hardey (Eds.), Making Sense of Microposts (#MSM2011), Workshop at Extended Semantic Web Conference (ESWC 2011), Crete, Greece (pp. 1–12). CEUR Workshop Proceedings Vol. 71. URL: <http://ceur-ws.org/Vol-718/>

WELLER, K. (2015): *Social Media and Altmetrics: An Overview of Current Alternative Approaches to Measuring Scholarly Impact*. In: Welp, I. M.; Wollersheim, J.; Ringelhan, S.; Osterloh, M. (Eds.), Incentives and Performance – Governance of Research Organizations (pp. 261-276). Springer. URL: <https://www.performancemanagement.wi.tum.de/fileadmin/w00bkk/www/Buch/manuscripts/Weller.pdf>

### 1.1.2 Andere Veröffentlichungen

HAEBERLI-KAUL, J. et al.: *Standardisierte Nutzungsstatistiken für Open-Access-Repositoryn und -Publikations-dienste*. DINI Schriften 13-de [Version 1.0], September 2013. URN: <urn:nbn:de:kobv:11-100212755>.

HERB, U. ; BEUCKE, D.: *Die Zukunft der Impact-Messung: Social Media, Nutzung und Zitate im World Wide Web*. In: Wissenschaftsmanagement. Zeitschrift für Innovation, 19 (2013), Nr. 4, S. 22–25. Preprint unter URL: <http://dx.doi.org/10.5281/zenodo.7696>.

WELLER, K. ; STROHMAIER, M. (2014) (Hrsg.). Social Media in Academia. *Special issue of IT – Information Technology*, 56(5). URL: <http://www.degruyter.com/view/j/itit.2014.56.issue-5/issue-files/itit.2014.56.issue-5.xml>

## 2 Ziele und Arbeitsprogramm

### 2.1 Voraussichtliche Gesamtdauer des Projekts

1. April 2016 bis 31. März 2018 (24 Monate)

### 2.2 Ziele

Das Hauptanliegen des Projektvorhabens ist es, ein tieferes Verständnis von \*metrics, also Nachweisen der Interaktion mit wissenschaftlichen Produkten, zu entwickeln, insbesondere im Hinblick auf ihre Aussagekraft und wie sie von Stakeholdern wahrgenommen werden. Dabei spielen die Betrachtung der Qualität der den \*metrics zugrundeliegenden Daten (Validität und Reliabilität), ihres Entstehungskontextes sowie ihrer Fähigkeit, Wissenschaft vor dem Hintergrund der disziplinären Spezifika zweckgemäß zu bewerten, eine entscheidende Rolle. Nutzende von \*metrics (z.B. Forschende, Forschungsförderer), Informationsinfrastruktureinrichtungen und Bibliotheken, sowie \*metrics-Anbieter profitieren von den Projektergebnissen, ermöglichen sie doch den informierten Einsatz von \*metrics und Kenntnis über ihre Interpretationsmöglichkeiten und Limitationen.

Ein öffentlich gefördertes Projekt wie \*metrics garantiert zudem Transparenz und den offenen Zugang zu den Projektergebnissen, was durch die Vormachtstellung der kommerziellen Anbieter in diesem Be-

reich in der Regel nicht der Fall ist. Dadurch trägt das Projektvorhaben zu einer Öffnung und besseren Nachvollziehbarkeit der \*metrics-Initiativen bei, gestaltet die Diskussion um das Thema mit, schließt Allianzen mit kommerziellen \*metrics-Anbietern und bezieht wichtige Zielgruppen (z.B. Forschende, die Repositorien-Community sowie Informationsinfrastrukturen) in die Entwicklung ein.

- **Projektziele**
- **Analyse und Implementierung (AP 1+2)**
- **Community-Aufbau (AP 3)**

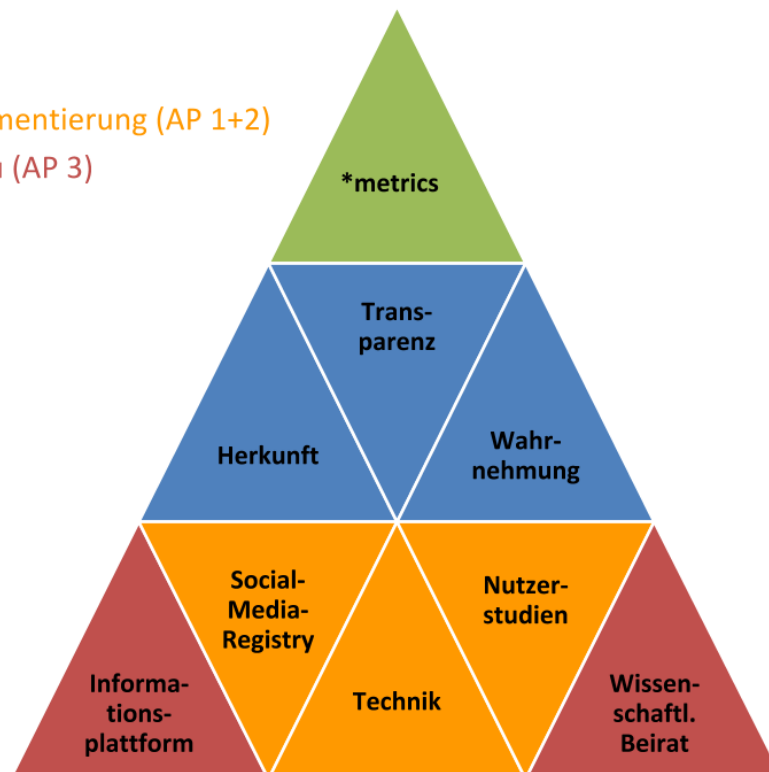


Abbildung 1: Projektorganisation und Zusammenspiel von Projektzielen, Analyse und Implementierung, Community-Vernetzung und Arbeitspaketen zur Beantwortung der Forschungsfrage: Welche Aussagekraft haben \*metrics und wie werden sie von den Nutzenden interpretiert?

Die Projektorganisation und das Zusammenspiel von Projektzielen (blau), Analyse und Implementierung (orange) und Community-Vernetzung (rot) sind in Abbildung 1 sowie im Gantt-Diagramm in Anlage 1 dargestellt. Projektziele sind die Analyse und Evaluation der Herkunft und Wahrnehmung der \*metrics, die in ihrer gemeinsamen Betrachtung Schlussfolgerungen zu der Aussagekraft von \*metrics zulassen. Die Auseinandersetzung mit diesen Schwerpunkten im Projekt führt zu einer größeren Transparenz und einem besseren Verständnis der \*metrics.

Daher fußt das Projekt auf zwei komplementären Ansätzen zum methodischen Vorgehen, die die Expertise der Projektgruppe optimal kombinieren und sich in den Arbeitspaketverantwortlichkeiten widerspiegeln. Der erste Ansatz umfasst die Nutzerstudien, die insbesondere von den wissenschaftlich-methodischen Kompetenzen der Projektteilnehmer ZBW und GESIS und ihrem Zugang zu den Zielgruppen sowie den eigenen Content-Plattformen (z.B. EconStor und SSOAR) profitieren, und findet in Arbeitspaket 1 „Status Quo und explorative Nutzer- und Nutzungsstudien“ statt. Ergänzend kommt eine Analyse des Status Quo der Social-Media-Plattformen und \*metrics hinzu, das sogenannte Social-Media-Registry, das die Basis für beide Ansätze bildet und in Kooperation der Projektmitglieder mit der wissenschaftlichen Fachcommunity erstellt wird. Der zweite Ansatz enthält die Bewertung der technischen Grundlage der \*metrics, die auf das verbundübergreifende, bibliothekarische Fachwissen der VZG aufbaut und in Arbeitspaket 2 „Technologie-Analyse zur Sammlung und Bereitstellung von \*metrics“ durchgeführt wird.

Flankiert werden die methodischen Arbeiten von einem wissenschaftlichen Beirat sowie vom Community-Vernetzung durch die SUB, der Öffentlichkeitsarbeit zum Projekt, der Verknüpfung mit (internationalen) Initiativen zu \*metrics (z.B. COAR, NISO) und dem Aufbau einer Informationsplattform, die alle Aktivitäten im Projekt bündelt (Arbeitspaket 3 „Community-Vernetzung“). Das Ziel ist es hier, frühzeitig Ergebnisse des Projekts in die Community zu tragen, um Feedback von Stakeholdern zu erhalten und dazu die hervorragende Vernetzung der Projektpartner zu nutzen. Die enge Verzahnung der Arbeitspakete, die nur durch Hauptverantwortlichkeiten (siehe Anlage 1) geprägt, jedoch immer in Kooperation durchgeführt werden, unterstützt die ganzheitliche Bearbeitung der Forschungsfragen.

## 2.3 Arbeitsprogramm und Umsetzung

### AP 1: Status Quo und explorative Nutzer- und Nutzungsstudien

Dieses Arbeitspaket widmet sich der Exploration der in der Wissenschaft populären Social-Media-Dienste, ihren Nutzenden und der Frage, wie diese Social-Media-Dienste und -Funktionalitäten verwenden. Damit bietet es den inhaltlichen Rahmen für die Interpretation und Bestimmung der Aussagekraft und Validität der \*metrics. Die aggregierten Ergebnisse dieses Arbeitspakets stellen die funktionalen und fachlichen Spezifikationen (im Sinne eines Pflichtenhefts) der \*metrics dar, die die Basis für die Untersuchungen und Weiterentwicklungen im Arbeitspaket 2 „Technologie-Analyse zur Sammlung und Bereitstellung von \*metrics“ bilden. Die Spezifikationen werden im kontinuierlichen engen Austausch mit AP 2 formuliert. Darüber hinaus eignen sich die Ergebnisse dazu, die Diskussion um die Standardisierung von \*metrics, wie z.B. bei der NISO<sup>1</sup>, inhaltlich zu fundieren und zu vervollständigen.

#### AP 1.1: Social-Media-Registry für die Wissenschaft erstellen

Es wird ein zweisprachiges (deutsch/englisch) Social-Media-Registry erstellt, das insbesondere die Funktionalitäten der Social-Media-Dienste beleuchtet (wie z.B. bei Peters & Heise, 2014) sowie deren Nutzergruppen erfasst. Die Bereitstellung des Registry erfolgt über die Informationsplattform \*metrics (vgl. AP 3.1).

Da sich \*metrics immer auf ganz bestimmte Funktionalitäten beziehen (z.B. Retweets) und damit ihre Aussagekraft bestimmen, müssen zunächst der Umfang und die zugrundeliegenden Mechanismen der Funktionalitäten bekannt sein. Dabei werden auch die Vor- und Nachteile bei der Nutzung der Dienste im wissenschaftlichen Kontext sowie die rechtlichen Rahmenbedingungen (auf Basis der Geschäftsbedingungen) verdeutlicht. Dies ist notwendig, da Forschende momentan mit diversen Social-Media-Diensten konfrontiert sind, die einerseits primär wissenschaftliche Arbeitsabläufe unterstützen (z.B. Literaturverwaltungssysteme wie Mendeley) und andererseits eher aus dem privaten Bereich stammen (z.B. soziale Netzwerke wie Facebook), aber dennoch zum wissenschaftlichen Arbeiten genutzt werden (vgl. dazu auch die Umfrage vom Leibniz-Forschungsverbund Science 2.0<sup>2</sup>). Die rechtlichen Aspekte betreffen zum Beispiel Fragen zur Weiterverwendung von Indikatoren aus Social-Media-Diensten (z.B. Tweets; Beurskens, 2014) und ihrer Bereitstellung (im Sinne von Open Metrics).

Die Arbeiten an dem Registry laufen von Projektbeginn an und müssen durch die aktiven Entwicklungen in dem Gebiet fortwährend angepasst werden. Die Arbeiten sollen durch die Community und den Beirat kommentiert bzw. ergänzt werden, um Expertise von außerhalb des Projekts einzubeziehen (vgl. AP 3.1 und AP 3.2). Neben den direkten Auswirkungen dieses Arbeitspakets auf die weiteren Arbeitspakete der Gruppe „Status Quo und explorative Nutzer- und Nutzungsstudien“, bestehen enge Querverbindungen zum AP 2.1, welches einen Überblick über die auf dem Markt befindlichen \*metrics-Aggregationssoftware gibt und aufzeigt, wo \*metrische Indikatoren derzeit eingesetzt werden (z.B. Verlage, Repositorien, Wissenschaftsevaluatoren, Bibliotheken, Lebensläufe). Weiterhin fließen die Ergebnisse aus AP 1.1 in die technischen Anpassungen und Weiterentwicklungen der Aggregationssoftware (AP 2.1) und der Software zur Ansprache von APIs ein (AP 2.3).

**Schwerpunkt:** Monat 1 bis 20

**Meilensteine::** **M1:** Monat 6 – a) Erste Vorstellung des Social-Media-Registry zwecks Community-Feedback (u.a. Präsentation auf Konferenz, im Workshop - AP 3.1); b) Erster Report über funktionelle und fachliche Spezifikationen der betrachteten Social-Media-Plattformen (für AP 2.1)

**M2:** Monat 20 – Finaler Report über funktionelle und fachliche Spezifikationen (aus AP 1.1-1.4) sowie Standardisierungsempfehlungen

<sup>1</sup> Pressemitteilung, 20.06.2013: „NISO to Develop Standards and Recommended Practices for Altmetrics“, [http://www.niso.org/news/pr/view?item\\_key=72efc1097d4caf7b7b5bdf9c54a165818399ec86](http://www.niso.org/news/pr/view?item_key=72efc1097d4caf7b7b5bdf9c54a165818399ec86), abgerufen am 14.10.2015.

<sup>2</sup> [http://www.goportis.de/fileadmin/downloads/aktuelles/Bericht\\_escience\\_2\\_0\\_Hochschulsample\\_Download.pdf](http://www.goportis.de/fileadmin/downloads/aktuelles/Bericht_escience_2_0_Hochschulsample_Download.pdf)

## AP 1.2: Nutzung von Social-Media-Plattformen verstehen

Social-Media-Dienste bieten nicht nur eine Vielzahl von Funktionalitäten an, sie werden auch unterschiedlich häufig für unterschiedliche Zwecke und aus unterschiedlichen Gründen genutzt, z.B. Mendeley für die Literaturverwaltung, Blogs für Publikation und Diskussion von Forschungsergebnissen (Shema, Bar-Ilan, & Thelwall, 2012) und Twitter zur Selbstrepräsentation (Bowman, 2015). Bei dem momentanen Gebrauch von \*metrics würde ein wissenschaftlicher Artikel jeweils für die Speicherung bei Mendeley sowie für die Erwähnung im Blog eine Erhöhung um 1 auf der Skala zur Messung des erzielten Einflusses erhalten – auch wenn die Speicherung für den Nutzenden vielleicht einen ganz anderen Stellenwert hat als die Erwähnung im Blog. Der \*metrics-Aggregator Plum versucht diese Problematik mit einer Kategorisierung der Indikatoren anzugehen<sup>3</sup>.

Um den Einfluss der Social-Media-Dienste sowie ihrer Funktionalitäten auf die Aussagekraft der \*metrics zu verstehen, werden aufbauend auf AP 1.1 sowie eigenen Vorarbeiten (u.a. Siegfried, Mazarakis, & Peters, 2014) Nutzer- und Nutzungsstudien in den beiden Zielgruppen Wirtschafts- und Sozialwissenschaften durchgeführt. Die Nutzerstudien umfassen Online-Umfragen sowie qualitative Methoden (z.B. Fokusgruppengespräch) und liefern durch Methoden-Triangulation einen detaillierten Überblick über die durchgeführten Aktionen (z.B. likes), die über die von den Social-Media-Diensten bereitgestellten Funktionalitäten ermöglicht werden. Die Ergebnisse geben zudem Aufschluss darüber, welche Social-Media-Dienste eine, ggf. disziplinspezifische, kritische Masse an Nutzenden erreicht haben und daher nachhaltig als Basis für \*metrics verwendet werden können. Die Schwerpunkte der Studien bilden Befragungen zur Verwendung der Social-Media-Dienste und ihrer Funktionalitäten (Wofür wird das genutzt? z.B. zum persönlichen Informationsmanagement) sowie zu den Gründen der Nutzung (Warum wird das genutzt? z.B. zur Selbstrepräsentation) bzw. Nicht-Nutzung (z.B. Werbung wird nicht retweeted; siehe auch AP 1.3). Ergänzt wird dieses Vorgehen durch die Auswertung von Nutzungsstatistiken<sup>4</sup>, die z.B. über die Logfiles der Content-Partner wie EconStor zur Verfügung gestellt werden und implizit Hinweise auf Nutzungsgewohnheiten geben (siehe auch AP 2.5). Wo möglich sollen auch weitere implizite Informationen zur Nutzung von Social-Media-Funktionalitäten herangezogen werden: Stellvertretend für social-media-basierte \*metrics werden Tweets betrachtet, die anhand von grundlegenden Statistiken (z.B. Retweets) ausgewertet und mit linguistischen Analyseverfahren (Natural Language Processing) untersucht werden. Die Auswahl der Social-Media-Plattformen für die erweiterten Analysen wird dabei vor allem durch die verschiedenen Nutzergruppen der Content-Partner sowie aus dem Umfeld der Projektpartner bestimmt (z.B. Forschende aus der BWL und VWL bei EconStor und aus den Sozialwissenschaften bei SSOAR), was einen Vergleich der Nutzungsgewohnheiten verschiedener Disziplinen erlaubt.

Ziel des APs ist zu erfahren, welchen Einfluss das Nutzungsverhalten auf die \*metrics hat, ob stabile Verwendungspraktiken nachzuweisen sind und wie \*metrics sinnvoll von Stakeholdern für unterschiedliche Zwecke (z.B. Relevanzbestimmung von Content) eingesetzt werden können. Durch statistische und linguistische Analysen entstehen Hinweise darauf, in welchem Rahmen der Verwendungszweck und die zugrundeliegende Motivation von \*metrics zuverlässig erfasst werden können, was für die Entwicklung künftiger automatischer Prozeduren zur Verarbeitung solcher Informationen essenziell ist und daher auch in AP 2 berücksichtigt werden muss.

**Schwerpunkt:** Monat 1 bis 12

**Meilensteine:**

- M3:** Monat 6 – Durchführung und Auswertung Explorationsumfrage (online) zur Verwendungsweise von Social-Media-Plattformen und -Funktionalitäten
- M4:** Monat 8 – a) Durchführung und Auswertung qualitativer Befragungen zur Verwendungsweise von Social-Media-Funktionalitäten und Anpassung der Motivationsumfrage; b) Erster Report über funktionelle und fachliche Spezifikationen zur Verwendung der Social-Media-Plattformen
- M5:** Monat 10 – Auswertung impliziter Nutzungsinformationen
- M6:** Monat 12 – a) Durchführung und Auswertung Vertiefungsumfrage (online) zur Nutzungsmotivation; b) Veröffentlichung der Ergebnisse in geeigneter Form

<sup>3</sup> <http://www.plumanalytics.com/metrics.html>

<sup>4</sup> Beispielsweise durch den Dienst Open-Access-Statistik: <https://www.gbv.de/Verbundzentrale/serviceangebote/oas-service/open-access-statistik-service>

### AP 1.3: Urheber von \*metrics kennen

Social-Media-Plattformen stehen allen Internetnutzenden zur Verfügung, so dass in einem umfassenderen Maße als bisher wissenschaftliche Kommunikation mit verschiedenen Öffentlichkeiten (z.B. Politik, Forschende, interessierte Laien) betrieben werden kann. Dies wird von den meisten Forschenden als starker Motivationsfaktor zur Nutzung von Social Media angesehen (Mazarakis & Peters, 2015). Die Kommunikation über Social Media ist zudem dadurch geprägt, dass keine Gatekeeper regulierend eingreifen und die Qualitätsbewertung der Inhalte damit auf die Nutzenden übertragen wird (Haustein, Larivière et al., 2014).

Die Frage nach der Provenienz der \*metrics liegt auf der Hand: Wer ist der Urheber und wie relevant ist sein Urteil? Die Antwort auf diese Frage scheint für viele Forschende ausschlaggebend zu sein, wenn über den Stellenwert von \*metrics diskutiert wird, vor allem weil häufig ein Mangel an fachlichem Austausch moniert wird (Mazarakis & Peters, 2015). Die Reputation des Urhebers spielt bei der Akzeptanz von \*metrics also eine entscheidende Rolle und wird bspw. bereits von dem Sozialen Netzwerk ResearchGate Page-Rank-ähnlich in seinem ResearchGate Score<sup>5</sup> berücksichtigt.

Das AP beschäftigt sich daher mit den Urhebern der \*metrics und bewertet die Aussagekraft der \*metrics hinsichtlich dieses Gesichtspunkts. Im engen Austausch mit AP 1.1 werden zunächst die Nutzenden verschiedener Social-Media-Plattformen charakterisiert (u.a. demographisch, disziplinar; siehe auch Mohammadi et al., 2015) sowie die technische Zugänglichkeit dieser Informationen überprüft, um geeignete Gruppen für die weiteren Analysen auswählen zu können (z.B. Doktoranden). Über soziale Netzwerkanalyse werden strukturelle Informationen über die Zusammensetzung und Topologie der Nutzerschaft gewonnen und ihr Einfluss auf die Wahrnehmung von \*metrics (siehe AP 1.4) ausgewertet. Analog zu den populären Zitationsdatenbanken (wie Web of Science), die vor dem Auszählen der Zitationen eine Quellenprüfung (z.B. Zeitschriften) unternehmen und nur bestimmte Dokumenttypen auswerten (z.B. Reviews), sollen solche Auswahlverfahren auch auf die Quellen der \*metrics angewendet werden, um die Sinnhaftigkeit eines solchen Vorgehens in Social-Media-Umgebungen zu prüfen. Die Untersuchungen erfolgen in Abstimmung mit AP 1.4 und sind wichtig geworden, da Anbieter von \*metrics bereits auf diese Weise arbeiten (z.B. altmetric.com<sup>6</sup>).

**Schwerpunkt:** Monat 3 bis 20

**Meilensteine:** **M7:** Monat 7 – Erster Report über funktionelle und fachliche Spezifikationen zur Urheberschaft

**M8:** Monat 16 – Charakterisierung Urheber von \*metrics und Bestimmung Relevanz von Social-Media-Quellen

### AP 1.4: Wahrnehmung von \*metrics erfassen

\*metrics werden häufig dafür eingesetzt, eine möglichst objektive Bewertung von Wissenschaft, Zeitschriften oder Forschenden durchzuführen und deren Relevanz zu bestimmen. Damit haben \*metrics eine Signalfunktion, die Entscheidungen auf Seiten der \*metrics-Nutzenden auslösen soll (z.B. wird ein Artikel heruntergeladen oder nicht). \*metrics-Aggregatoren bieten oftmals nicht nur die reinen Nutzungszahlen zu den wissenschaftlichen Publikationen an, sondern versuchen auch, diese in einen Kontext einzubetten, um die Interpretation und Einordnung der Werte zu ermöglichen. Eine Rangfolge nach Höhe der \*metrics ist dabei die einfachste Darstellungsoption. ImpactStory zeigt die Social-Media-Metriken und Zitationszahlen bspw. in Perzentilen an, altmetric.com visualisiert die verschiedenen \*metrics durch einen farbkodierten „Donut“ sowie einen Score. Darüber hinaus kann sowohl die geographische als auch die disziplinäre Herkunft der Nutzenden auf einer Landkarte dargestellt werden.

Obwohl all diese Bemühungen darauf abzielen, die \*metrics besser verständlich und nachvollziehbar zu machen, konnte West (2014) zeigen, dass auf Seiten der Nutzenden große Unsicherheiten bei der Interpretation solcher Darstellungen bestehen.

Da mit diesem AP weitgehend Neuland in der \*metrics-Forschung betreten wird, wird ein elaboriertes, vierstufiges Untersuchungsdesign eingesetzt:

<sup>5</sup> <http://www.researchgate.net/publicprofile.RGScoreFAQ.html>

<sup>6</sup> "Each category of mention contributes a different base amount to the final score. For example, a newspaper article contributes more than a blog post which contributes more than a tweet", <https://www.altmetric.com/whatwedo.php>.



- 1) In Kooperation mit AP 1.2 und AP 1.3 werden den Nutzenden die verschiedenartig visualisierten \*metrics und korrespondierende wissenschaftliche Produkte zur Evaluation präsentiert. Die Evaluation wird mit mehrstufigen Conjoint-Analysen und monetärer Wertschätzung durchgeführt, um den Einfluss der \*metrics und anderer Eigenschaften des wissenschaftlichen Produkts (z.B. Veröffentlichungsdatum, Dokumenttyp) auf das Präferenzurteil des Nutzenden zu ermitteln.
- 2) Die Conjoint-Analysen werden um Eye-Tracking-Studien ergänzt, um von den Probanden nicht zu verbalisierende Präferenzen in Visualisierungen oder Anordnungen der \*metrics ableiten zu können (z.B. über Heat Maps)<sup>7</sup>. Gwizdka (2014) zeigt, dass nicht-relevante Dokumente eher gescannt als gelesen werden, das Projekt „Seeing Data“ hat elf Faktoren ermittelt, die die Auseinandersetzung mit Daten-Visualisierungen beeinflussen<sup>8</sup>. Basierend auf den Nutzer-Netzwerk-Analysen und den Inhaltsanalysen aus AP 1.2 und 1.3 kann hier zudem für einzelne Fallstudien die Wahrnehmung bestimmter Gruppen analysiert werden. Bei GESIS gibt es Vorarbeiten zur Wahrnehmung von Genderunterschieden in Wikipedia (Wagner et al., 2015). Auch in anderen Online-Umgebungen wurden geschlechtsspezifische Wahrnehmungsunterschiede beobachtet (Sugimoto et al., 2013).
- 3) Social-Media-Datenanalysen runden das experimentelle Setting ab. Basierend auf den Ergebnissen der Conjoint-Analysen und Eyetracking-Studien lassen sich Hypothesen für den Verlauf der Entwicklungen von social-media-basierten \*metrics ableiten. Mit Hilfe von Quasi-Experimental Designs (QED) können diese auf kausale Zusammenhänge überprüft werden, indem man beispielsweise verschiedene Einheiten mit zunächst gleichen Ausgangskriterien in ihrer Entwicklung im Zeitverlauf vergleicht. Hierbei handelt es sich um Methoden, die in der Social-Media-Forschung bereits vereinzelt angewandt werden (z.B. von Cheng et al., 2014), die jedoch für die \*metrics Forschung noch einen Neuigkeitswert haben: lassen sich beispielsweise Entwicklungen von Retweethäufigkeiten oder Bookmarking-Zahlen auf Grund bestimmter Eigenschaften vorhersagen? Die genaue Ausgestaltung der Analysen im Quasi-Experiment-Design erfolgt erst im Projektverlauf, so dass in vorherigen Projektphasen gewonnene Erkenntnisse zur Nutzungsgewohnheit berücksichtigt werden können.
- 4) Der Einsatz von qualitativen Methoden (z.B. Interviews) zielt auf die Ermittlung von jenen Eigenschaften der wissenschaftlichen Produkte, die in weiteren Conjoint-Analysen abgefragt werden sollten. So werden qualitative Merkmale mit quantitativen Ausprägungen in Beziehung gesetzt.

Die Ergebnisse ermöglichen ein tieferes Verständnis der Wahrnehmung von Indikatoren und \*metrics auf Seiten der Nutzerschaft (ggf. disziplin- und plattformspezifisch) sowie zukünftige Entwicklungen im Bereich Visualisierung und Kontextbildung. Diese können auch als Grundlage für maschinelle Testverfahren (z.B. im Machine Learning) verwendet werden, um auf lange Sicht die Qualität wissenschaftlicher Produkte und die Ausprägung der \*metrics automatisiert zusammenzubringen. So werden auch Anforderungen an \*metrics deutlich, die frühzeitig an die Stakeholder zurückgespiegelt werden sollen, um (technische) Umsetzung und Diskussion zu befördern (siehe AP 3.1).

Als Anwendungs- und Rekrutierungsumgebung für die \*metrics-Visualisierungen und Testnutzenden, fungieren die Projektpartner mit EconStor<sup>9</sup>, SSOAR<sup>10</sup> und GoeScholar<sup>11</sup> (vgl. AP 2.5). Die jeweilige Community wird um Feedback zu den Funktionen und der Darstellung gebeten (vgl. AP 1.2 und AP 2.5). Es sind auch technische Arbeiten nötig, wie z.B. das Anpassen von in AP 2.1 evaluierter Software zur \*metrics-Aggregation sowie weiterer Software zur Abfrage der Social-Media-Schnittstellen (siehe AP 2.3) und zur Lieferung und Aufbereitung der Daten für die Vorhaltung in Datenbanken, um sie wissenschaftlich nutzen zu können.

<sup>7</sup> Die ZBW verfügt über ein eigenes Usability-Labor am Standort Hamburg, das für diese Zwecke genutzt werden kann. Die Anschaffung weiterer Soft- und Hardware zur Durchführung der Studien erfolgt aus ZBW-Mitteln.

<sup>8</sup> siehe <http://blogs.lse.ac.uk/impactofsocialsciences/2015/07/22/seeing-data-how-people-engage-with-data-visualisations/>

<sup>9</sup> ZBW betreibt mit EconStor das größte deutsche Repository, welches als fachlicher Publikationsserver von über 250 wirtschaftswissenschaftlichen Einrichtungen zur Verbreitung ihrer Veröffentlichungen genutzt wird. Momentan umfasst Econstor mehr als 89.000 Publikationen der Volks- und Betriebswirtschaft.

<sup>10</sup> Das Social Science Open Access Repository (SSOAR) wird seit 2007 von GESIS betrieben. Hier werden frei verfügbare sozialwissenschaftliche Publikationen im Volltext (i.d. R. im Rahmen des Zweitveröffentlichungsrechts) gesammelt und recherchierbar gemacht. Aktuell umfasst das Repository über 34.000 Open-Access-Publikationen.

<sup>11</sup> GoeScholar ist ein Open-Access-Repository und stellt die Forschungsergebnisse der Universität Göttingen damit frei zur Verfügung. GoeScholar bildet alle Disziplinen und Wissenschaftsbereiche der Universität Göttingen ab.

**Schwerpunkt:** Monat 6 bis 24

**Meilensteine::**

- M9:** Monat 11 – a) Visualisierung \*metrics und Implementierung bei Content-Partnern (auf Basis der vorhandenen Aggregationssoftware), b) Konzeption Conjoint Analysen und Befragungen
- M10:** Monat 13 – a) Durchführung Experimente und Ergebnisanalyse; b) Erster Report über funktionelle und fachliche Spezifikationen zur Darstellung von \*metrics
- M11:** Monat 16 – Einbindung aktualisierter Daten bei Content-Partnern und Eye-Tracking-Studien
- M12:** Monat 21 – Durchführung Experimente und Ergebnisanalyse
- M13:** Monat 24 – Publikation der Ergebnisse in geeigneter Form

## **AP 2: Technologie-Analyse zur Sammlung und Bereitstellung von \*metrics**

Parallel zur explorativen Phase in AP 1 werden die zur Zeit auf dem Markt verfügbaren Aggregationssoftwarelösungen zur Sammlung und Bereitstellung von \*metrics hinsichtlich ihrer zu erwartenden Leistungsfähigkeit bewertet sowie ausgewählte Open-Source-Aggregationssoftwarelösungen testweise installiert und hinsichtlich ihrer Funktionalität evaluiert. Dazu werden die Repositorien der Content-Partner als Datengrundlage (Dokumente und Metadaten) genutzt und deren Dokumente und Metadaten mittels der zu untersuchenden Aggregationssoftwarelösungen aufbereitet. Die ausgewählten Aggregationssysteme können so hinsichtlich Quellen, Bereitstellung, Funktionen, Skalierbarkeit, Zahlenvergleich etc. beschrieben und analysiert werden. Da sehr wahrscheinlich nicht alle Aggregationssoftwarelösungen ihre Berechnungsgrundlagen sowie die verwendete Datengrundlage vollumfänglich offenlegen, wird zusätzlich zu den oben genannten Analysen, die Datengrundlage der einzelnen Repositorien angereichert (durch Match-and-Merge der Daten aus den anderen Repositorien sowie weiterer Daten, z.B. CBS<sup>12</sup>-Daten etc.) und bei jedem Arbeitsschritt wiederholt mittels der verschiedenen Aggregationssoftwarelösungen aufbereitet, um die jeweiligen Änderungen in den Ergebnissen protokollieren und nachvollziehen zu können. Zusätzlich sollen die einzelnen Datenmengen aus den Repositorien der Content-Partner sowie des CBS mittels Match-and-Merge zu einem großen Datenpool (Dokumente und Metadaten) zusammengeführt und wiederum mittels der einzelnen Aggregationssoftwarelösungen aufbereitet werden.

Sofern die Datenbasen (Erwähnungen von Dokumenten in verschiedenen Social-Media-Diensten aufgrund bestimmter Metadaten) der einzelnen Aggregationssoftwarelösungen offenliegen, werden auch diese gegenseitig mittels Match-and-Merge durch die Datenbasen der anderen Aggregationssoftwarelösungen angereichert. Die Datengrundlagen werden dann wiederum durch die so angereicherten Aggregationssoftwarelösungen bearbeitet und die Ergebnisse miteinander verglichen. Durch diese Prozesse wird eine Vergleichbarkeit der auf dem Markt befindlichen Aggregationssoftwarelösungen möglich. Zusätzlich entsteht ein großer Datenpool (Dokumente und Metadaten) aus den angereicherten Datengrundlagen der Content-Partner sowie der CBS-Daten in Form eines „Metadaten-Register“.

Im finalen Schritt werden die einzelnen Softwarekomponenten im Hinblick auf die Ergebnisse der Exploration aus AP 1 („Pflichtenheft“) angepasst, zusammengesetzt und als Proof-of-Concept zur Lauffähigkeit gebracht. Das Ziel von AP 2 ist es also, auf Basis der vorhandenen und angepassten Aggregationssoftwarelösungen Informationen über die Reproduzierbarkeit, Vollständigkeit und Reliabilität der \*metrics zu sammeln. Die daraus gewonnenen Ergebnisse fließen zurück in den Explorationsprozess und geben damit auch darüber Auskunft, welche Schlussfolgerungen hinsichtlich der \*metrics-Auswertungen nicht gezogen werden können, da weder die technischen Gegebenheiten der Aggregationssoftware noch die verfügbaren \*metrics-Daten dies zulassen. Sämtliche während der Projektlaufzeit entwickelten Methoden und Softwarekomponenten sowie die aggregierten Daten, unter Berücksichtigung der Lizenzbedingungen der Ausgangsdaten, werden der Allgemeinheit frei zugänglich zur Verfügung gestellt.

<sup>12</sup> Das CBS der Firma OCLC ist die technische Plattform zum Betrieb des GVK (Gemeinsamer Verbundkatalog). Der GVK ist mit über 51 Millionen Titeldatensätzen und 97 Millionen Besitznachweisen die größte Datenbank zur Katalogisierung in Deutschland.

Die Unterarbeitspakete dieses APs sollen als jeweils dynamische Arbeitsprozesse verstanden werden, die stets ineinandergreifen. Somit ist die Aufzählung nicht als Abfolge von Arbeitsschritten zu verstehen, sondern als Zusammenfassung von Arbeiten mit ähnlichem Charakter.

### **AP 2.1: Vorhandene \*metrics-Aggregationssoftware sichten und evaluieren**

Der Bereich der alternativen Metriken und der Social-Media-Dienste zeichnet sich zurzeit unter anderem durch sehr intensive Entwicklungen hinsichtlich der verwendeten Metriken und der jeweils eingesetzten Software aus. Entsprechend wächst auch das Angebot an Software kontinuierlich, die die unterschiedlichen Social-Media-Dienste abfragt und die einzelnen Indikatoren zu \*metrics aggregiert. Entsprechend lässt sich zum Zeitpunkt der Antragsstellung keine aussagekräftige Übersicht erstellen. Zu Beginn der Projektlaufzeit wird deshalb auf Grundlage des in AP 1.1 beschriebenen Social-Media-Registry der Markt gesichtet und die aktuellen Systeme hinsichtlich Quellen, Bereitstellung, Funktionen, Skalierbarkeit etc. weiter analysiert. Ein besonderes Augenmerk wird dabei auf die Funktionalität der vorhandenen Lösungen hinsichtlich der jeweiligen Crawling- und weiteren technischen Eigenschaften gelegt. Es wird außerdem berücksichtigt, wo \*metrische Indikatoren (Metadaten, die für die Zusammenführung von Dokumenten verwendet werden können) sowie \*metrics-Aggregationssoftwarelösungen derzeit eingesetzt werden (z.B. Verlage, Repositorien, Bibliotheken).

Auf Grundlage dieser Übersicht werden ausgewählte Open-Source-Lösungen testweise installiert und im Hinblick auf ihre Funktionalität und Skalierbarkeit evaluiert. In Abstimmung mit den Ergebnissen aus der Exploration in AP 1 werden weitere Evaluationskriterien entworfen, anhand derer die installierten Open-Source-Lösungen detailliert bewertet und insbesondere im Hinblick auf ihre Einsatzmöglichkeiten im Bibliotheks- und Repository-Umfeld geprüft werden. Gleichzeitig informiert AP 2.1 die in AP 1 durchzuführenden Arbeiten über die technischen Möglichkeiten der Software, sodass diese z.B. in den Umfragen (siehe AP 1.2) berücksichtigt werden können. Darüber hinaus werden notwendige Erweiterungen beschrieben, um die in AP 1 formulierten funktionellen und fachlichen Spezifikationen erfüllen zu können, sowie deren Implementierungsaufwand abgeschätzt.

Auf Grundlage dieser Analysen wird die Entscheidung zugunsten einer oder mehrerer der vorgestellten Open-Source-Softwarelösungen fallen, die im Folgenden als Grundlage für die Bereitstellung einer transparenten Datenbasis dienen.

**Schwerpunkt:** Monat 1 bis 13

**Meilensteine:** **M14:** Monat 3 – Arbeitspapier: Übersicht und Analyse vorhandener Aggregationssoftware

**M15:** Monat 9 – Arbeitspapier: Evaluierung ausgewählter Open-Source-Lösungen

**M16:** Monat 13 – Handlungsanweisung: Notwendige Erweiterungen und Implementierungsaufwand für AP 2.3

### **AP 2.2: Möglichst vollständige Datengrundlage schaffen**

Dokumente werden gewöhnlich in sehr unterschiedlicher Weise in Social-Media-Kanälen referenziert bzw. sind unter verschiedenen Adressen erreichbar. So stellt die Nutzung von URL-Kürzungsdiensten, die Verlinkung auf die Metadaten des Dokuments oder seine verschiedenen Repräsentationsformen eine Herausforderung für die Messung von Zugriffszahlen auf Dokumentenebene dar. Zusätzlich bestehen sogar oft für dasselbe Dokument verschiedene eindeutige Identifier (DOI, URN, PURL, Handle, ISBN etc.), oder das Dokument besitzt keinerlei Identifier, sondern es werden nur Titel und/oder Autor genannt. Die gängigen Lösungen führen diese verschiedenen Repräsentationsformen nicht hinreichend zusammen oder erledigen dies – wie der kommerzielle Anbieter Plum Analytics – mit intransparenten Verfahren und genügen damit nicht dem Kriterium der Reproduzierbarkeit.

Zudem ist zu beachten, dass nicht jede Plattform alle Dokumente referenziert bzw. per API ausliefert. Auch die metrischen Indikatoren der einzelnen Social-Media-Plattformen sowie die Granularität der gewonnenen Daten weichen teilweise stark voneinander ab. Herauszufinden, inwiefern die Daten technisch miteinander verglichen werden können und inwieweit die jeweiligen Abfragemetriken erweitert werden müssen, ist ein wesentlicher Bestandteil dieses APs und steht damit im engen Zusammenhang mit dem eher inhaltlich ausgerichteten AP 2.4.

Es müssen Werkzeuge zur Verfügung stehen, welche die \*metrics in ausreichender Menge und Qualität bei den unter AP 1.1 benannten Diensten abfragen, um eine aussagekräftige Auswertung durchfüh-

ren zu können. Hier sollen die unter AP 2.1 designierten Open-Source-Softwarelösungen auf ihre nachnutzbaren Funktionalitäten geprüft und gegebenenfalls weiterentwickelt werden. Die technische Grundlage dieser Abfragemimiken bilden größtenteils die APIs der einzelnen Aggregationsdienste (REST, Streaming, etc). Hier wäre es mit überschaubarem Aufwand möglich, eigene Abfragen für den zu entwickelnden Dienst zu implementieren und die vorhandene Software dadurch zu erweitern (AP 2.3). Als Anwendungsfälle sollen die Publikationen der Content-Partner EconStor, SSOAR und GoeScholar als Datengrundlage für die Zusammenführung der Daten aus den Social-Media-Diensten verwendet werden. Bei Bedarf kann diese Datengrundlage um weitere Content-Partner ergänzt werden, um z.B. die Anwendbarkeit auf andere Dokumenttypen wie Datensets zu überprüfen.

Ist diese Datengrundlage erstellt, werden diese Daten (Dokumente und Metadaten) in das Katalogisierungssystem der VZG in einen separaten Bereich eingespielt. In diesem Arbeitspaket wird ein Mapping-Register entwickelt, das für eine beliebige Referenz eines Dokuments alle anderen existenten Referenzierungsformen von Manifestationen dieses Dokuments auflistet. Dazu ist die Erstellung eines bibliografischen Indexes notwendig, der Metadaten bestehender Artikeldatenbanken und durch Webcrawling gewonnene Referenzierungsformen vereinigt. Als Basis werden, zusätzlich zu den im CBS vorhandenen Metadaten, die frei zugänglichen bzw. für Bibliotheken kostenfrei nutzbaren Metadaten aus Cross-Ref, BASE, Culturegraph und weiteren Quellen genutzt. So entsteht eine zentrale „Datendrehzscheibe“ für den zu entwickelnden Dienst, in welcher mit den vorhandenen Mapping- und Merging-Algorithmen die eingesammelten Daten zusammengeführt werden können. Im CBS stehen ausgereifte Softwarelösungen zur Deduplizierung und Zusammenführung von beliebigen Daten zur Verfügung. Die VZG als Partner von OCLC hat Zugriff auf den gesamten Sourcecode dieser Software und kann entsprechend Änderungen daran vornehmen. Beispielsweise bietet die Software die Möglichkeit, mehrstufig gewichtete Strings zu vergleichen bzw. zusammenzuführen, so dass abgekürzte und ähnliche Schreibweisen des selben Begriffs erkannt werden können. Damit ist dann auch eine Texterkennung von bspw. Titeln in Tweets realisierbar, wenn kein Identifier vorhanden ist. Im Anschluss steht ein Datenpool aus Repository-Dokumenten, \*metrics etc. allen Interessierten zur Nachnutzung zur Verfügung, sofern das rechtlich möglich ist.

Dieses AP zielt auf die Erstellung einer vollständigen Datenbasis als Grundlage für \*metrics-basierte Untersuchungen. Von der Deduplizierung der wissenschaftlichen Produkte und Zusammenführung aller Referenzierungsformen profitieren nicht nur die \*metrics-Anbieter, sondern vor allem die Forschenden und Autoren, da eine vollständige Abdeckung der Identifier bei der Suche auch zu vollständigeren und damit valideren \*metrics führt.

**Schwerpunkt:** Monat 3 bis 24

**Meilensteine:** **M17:** Monat 12 - a) Handlungsempfehlung: Vergleichbarkeit von Social-Media-Metriken und notwendige Algorithmen-Erweiterungen; b) Erstellung erster belastbarer Datengrundlage für \*metrics-basierte Untersuchungen

**M18:** Monat 14 – Erstellung eines ersten Registers für das Mapping von Dokumenten

**M19:** Monat 18 – a) Arbeitspapier: Prüfung und Weiterentwicklung der Open-Source-Softwarelösungen auf ihre nachnutzbaren Funktionalitäten; b) Anwendungsfälle EconStor, SSOAR und GoeScholar

### **AP 2.3: Technische Anpassungen und \*metrics-Bereitstellung durchführen**

In diesem Arbeitspaket werden die in AP 2.1 und 2.2 entwickelten Handlungsanweisungen zusammen mit den Erfordernissen aus den anderen Arbeitspaketen (insbesondere AP 1.4) schrittweise umgesetzt, um die Ziele des Projekts zu erreichen. Sämtliche Arbeiten an der Software erfolgen von Anfang an in GitHub, damit andere an die Entwicklungen anknüpfen können. Bereits vorhandene Software zur \*metrics-Aggregation (siehe AP 2.1) und zur Ansprache von Web-Schnittstellen weiterer Social-Media-Plattformen (z.B. Webometric Analyst<sup>13</sup>) sowie entsprechender Code (z.B. über GitHub) sollen soweit wie möglich nachgenutzt werden. Durch diese technischen Anpassungen und Erweiterungen wird als Ergebnis eine hinreichend vollständige \*metrics-Datengrundlage generiert und mittels einer abfragbaren Open-Source-Datenbank (z.B. CouchDB) der interessierten Community bereitgestellt.

<sup>13</sup> <http://lexiurl.wlv.ac.uk>

**Schwerpunkt:** Monat 4 bis 24

**Meilensteine:**

- M20:** Monat 12 – Bereitstellung erster \*metrics-Datengrundlage
- M21:** Monat 22 – Dokumentation des Codes und des Workflows
- M22:** Monat 24 – Entwicklung und Bereitstellung a) angepasster Software; b) hinreichend vollständiger \*metrics-Datengrundlage; c) hinreichend vollständiges Mapping-Register

#### **AP 2.4: Abdeckungsgrad und Ausprägung der \*metrics bei Aggregatoren überprüfen**

Social-Media-Dienste stellen offene Schnittstellen (APIs) zur Verfügung. \*metrics-Aggregatoren greifen auf diese zu und stellen die Suchergebnisse ihren Kunden gebündelt zur Verfügung. Daher ist man an dieser Stelle mit zwei Bottlenecks konfrontiert. Der erste Bottleneck betrifft die Ausgabe der \*metrics durch die Social-Media-Plattformen. Twitter stellt z.B. zwei Zugangswege bereit, die den Abdeckungsgrad und die Ausprägung der \*metrics maßgeblich beeinflussen. Während der (kostenpflichtige) Twitter Firehose-Zugang<sup>14</sup> 100% der verfügbaren Tweets passend zu den Suchparametern ausgibt, werden über die (kostenlos zugängliche) Twitter Streaming API<sup>15</sup> nur maximal 1% des Tweetvolumens eines gewissen Zeitraums zurückgegeben. Morstatter et al. (2013) zeigen, dass über die Streaming API ca. 43% der Tweets abgerufen werden können (i.e. Abdeckungsgrad). Im Hinblick auf die Ausprägung, d.h. Intensität, der \*metrics liefert die Streaming API schlechtere Ergebnisse als zufällig ausgewählte Tweets aus dem Firehose.

Der zweite Bottleneck findet sich auf Seiten der \*metrics-Aggregatoren, da eine Intransparenz bezüglich der Algorithmen zum Auffinden der wissenschaftlichen Produkte innerhalb der Social-Media-Plattformen sowie bei der Zusammenfassung der \*metrics besteht. Die Nutzenden wissen nicht, ob die Aggregatoren alle Repräsentationen eines wissenschaftlichen Produktes bei der Suche berücksichtigen. Eigene und weitere Vorarbeiten (u.a. Chamberlain, 2013; Jobmann et al., 2014; Peters et al., 2015; Zahedi, Fenner, & Costas, 2014) konnten bereits bestätigen, dass verschiedene \*metrics-Aggregatoren eine unterschiedliche Anzahl an wissenschaftlichen Produkten in den Social-Media-Diensten auffinden und sich die Werte der \*metrics ebenfalls unterscheidet. Ein Forschungsprojekt zur Vergleichbarkeit der \*metrics-Aggregatoren wird aktuell von altmetric.com gefördert<sup>16</sup>.

Das AP verfolgt zwei Ziele. Zum einem erfolgt ein systematischer Vergleich der \*metrics-Aggregatoren (aus AP 2.1), um die Konsistenz der damit erzielten \*metrics zu überprüfen. Den Schwerpunkt der Analyse bildet die Vollständigkeit der Suche der Aggregatoren, also der Abdeckungsgrad, der durch die technischen Arbeiten in AP 2.2 fundiert wird. Von anderen Aggregatoren (bspw. altmetric.com, ...) bereitgestellte Datensätze an Social-Media-Metriken für Repositories werden mit den eigenen erhobenen \*metrics verglichen. Da in AP 2.3 ein möglichst vollständiger Datenraum mit allen Repräsentationen der wissenschaftlichen Produkte hergestellt und für die Suche genutzt wird, verspricht das Vorgehen profunde Erkenntnisse über die Reliabilität der \*metrics-Aggregatoren. Weitere Ergebnisse sind:

- a) Mit den Daten aus AP 2.3 werden strukturelle und inhaltliche Informationen darüber gewonnen, welche permanenten Identifier besonders gut von den Aggregatoren abgedeckt werden und welche Publikationstypen häufig in Social-Media-Diensten aufzufinden sind (z.B. im Repository um archivierte Version mit Handle vs. Non-Open-Access-Verlagsversion mit DOI). Die inhaltlichen Analysen grenzen das AP damit auch von o.g. altmetric-Forschungsprojekt ab.
- b) Der Vergleich der \*metrics-Aggregatoren soll darüber hinaus die Konsistenz der \*metrics über verschiedene Zeiträume betrachten, wobei hier vor allem die Plausibilität der \*metrics (sowohl Abdeckung als auch Ausprägung) eine Rolle spielt.

Die stichprobenartige Gegenüberstellung hinsichtlich Vollständigkeit, Überlappung und Ausprägung der \*metrics von potentiell auffindbaren wissenschaftlichen Produkten (aus AP 2.2), über \*metrics-Aggregatoren (aus AP 2.3) sowie über Social-Media-Dienste selbst gefundener Publikationen ist das zweite Ziel des APs. Damit wird die Tradition der Untersuchungen zur Validität und Fehlerrate der populären Zitationsdatenbanken wie Web of Science oder Scopus sowie Google Scholar aufgegriffen (García-Pérez, 2010; Jacsó, 2006; Meho & Rogers, 2008) und das Vorgehen auf die für das Projekt aktuellen Gegebenheiten übertragen.

<sup>14</sup> <https://dev.twitter.com/streaming/firehose>

<sup>15</sup> <https://dev.twitter.com/streaming/overview>

<sup>16</sup> <https://altmetricsconf.wordpress.com/2014/12/17/1am-altmetrics-project-funding-awarded>

Die Ergebnisse tragen maßgeblich zur besseren Nachvollziehbarkeit der aggregator- und social-media-basierten \*metrics bei, lassen Schätzungen über eine Fehlerrate zu und helfen, die Aussagekraft der \*metrics zu bestimmen sowie informierte Entscheidungen über den Einsatz von \*metrics-Aggregatoren z.B. in Bibliotheken zu treffen.

**Schwerpunkt:** Monat 12 bis 19

**Meilensteine:** **M23:** Monat 15 – Vergleich der \*metrics-Aggregatoren  
**M24:** Monat 17 – Analyse der inhaltlichen Ergebnisse  
**M25:** Monat 19 – Vergleich verschiedene Datenbasen zur Abdeckung und Höhe der \*metrics

### **AP 2.5: Integration bei Content-Partnern testen (Proof-of-Concept)**

Auf Grundlage der Ergebnisse aus den vorhergehenden APs soll eine Funktionalität entwickelt werden, welche die \*metrics-Daten prototypisch<sup>17</sup> den Content-Partnern EconStor, SSOAR und GoeScholar zur Verfügung stellt (z.B. Pop-up-Fenster). Ebenfalls prototypisch sollen Discovery-Systeme (VuFind-PlugIn) und Bibliothekskataloge die Möglichkeit erhalten, die gewonnenen Daten anzubieten, wobei die Form der \*metrics-Daten zunächst geklärt werden muss. Geplant ist derzeit ein Widget, welches die Daten anzeigt, oder alternativ ein eigenes Browser-Fenster oder eine Mouseover-Funktionalität. Die Entwicklungen dazu werden die Ergebnisse aus AP 1.4 aufgreifen. Unter Einbeziehung der Datenzugangsmöglichkeiten (z.B. ob Daten per API oder beispielsweise als Wochenlieferung den jeweiligen Diensten zur Verfügung gestellt werden) wird ein angemessenes Zeitintervall zwischen Crawlingzeitpunkt, Skalierbarkeit der Software und Daten-Aktualität erarbeitet (aus AP 2.2). Darüber hinaus stellt AP 2.5 sicher, dass eine zusätzliche Testumgebung für Nutzerstudien (z.B. bei Experimenten zur Wahrnehmung von \*metrics in AP 1.4) bereitsteht.

**Schwerpunkt:** Monat 17 bis 24

**Meilensteine:** **M26:** Monat 18 – Testintegration bei EconStor, SSOAR und GoeScholar  
**M27:** Monat 23 – Prototypische Einbindung des PlugIn in VuFind

### **AP 3: Community-Vernetzung und Projektmanagement**

Es widmen sich erst seit kurzem systematische Untersuchungen dem genaueren Verständnis von \*metrics (i.e. Haustein, Sugimoto, & Larivière, 2015). Die Vielzahl an \*metrics-Aggregationssoftware und Social-Media-Plattformen macht zudem eine Auseinandersetzung mit der Thematik mühsam. Interessierte Personengruppen müssen die vorhandenen Informationen zum Thema noch an verschiedenen Stellen (z.B. auch über Blogs) aufwändig zusammensuchen. Daher ist ein Ziel des Projekts, dass interessierte Communities an einer zentralen Stelle und über verschiedene Dialogformate informiert werden. Darüber hinaus ist das Projektmanagement in diesem Arbeitspaket angesiedelt.

#### **AP 3.1: Projektergebnisse von \*metrics in die Community kommunizieren und offenen Diskurs führen**

Als ein zentraler Output des Projekts ist die Informationsplattform \*metrics geplant, die über den aktuellen Stand im Projekt informiert und die internationalen \*metrics-Diskussionen bündelt. Die Plattform ist als Anlaufstelle für die verschiedenen an der Thematik interessierten Communities vorgesehen und wird zweisprachig (deutsch/englisch) angeboten. Damit kann die deutschsprachige Bibliotheks-Community erreicht werden, die Tools für ihre Informationssysteme (z.B. Repositorien) nutzen möchte. Gleichzeitig dient die Plattform als Anlaufstelle für die internationale \*metrics-Community. Dabei werden bereits bestehende Kooperationen der beteiligten Projektpartner berücksichtigt und beispielsweise die Repository-Community über die Confederation of Open Access Repositories (COAR)<sup>18</sup> oder die \*metrics-Community über die Beteiligung an der NISO Altmetrics Working Group „A: Definitions & Use Cases“<sup>19</sup> mit eingebunden. Ebenso können die Kontakte zu den Beiratsmitgliedern genutzt werden, um

<sup>17</sup> Sollten die in AP 1.4. durchgeführten Untersuchungen als Desiderat ergeben, weitere Visualisierungsmöglichkeiten bereitzustellen, könnten diese in einem Folgeantrag erarbeitet werden.

<sup>18</sup> <https://www.coar-repositories.org/> und im Speziellen die Interest Group „Usage Data and Beyond“.

<sup>19</sup> Isabella Peters und Daniel Beucke sind Mitglieder in der Arbeitsgruppe; [http://niso.org/topics/tl/altmetrics\\_initiative](http://niso.org/topics/tl/altmetrics_initiative).

Schnittstellen zu anderen Plattformen zu aktivieren. Auf nationaler Ebene findet eine enge Abstimmung mit dem Kompetenzzentrum Bibliometrie<sup>20</sup> statt. Interessierte können sich einen detaillierten Überblick über die derzeit angebotenen \*metrics-Aggregatoren verschaffen und die verschiedenen Funktionen vergleichen. Darüber hinaus wird das Social-Media-Registry in die Informationsplattform integriert, die mit umfangreichen Möglichkeiten zur Kommentierung ausgestattet ist, um Anmerkungen aus der Community zu erfassen.

Es ist angedacht, die \*metrics-Informationsplattform als Teilbereich der COAR-Website zu realisieren. Zwei Projektpartner beteiligen sich aktiv an den Arbeiten in der COAR-Community (z.B. Leitung der Interest Group „Usage Data and Beyond“), wodurch die nachhaltige Pflege der Plattform gesichert wäre. Die COAR-Geschäftsstelle ist an der SUB Göttingen angesiedelt, so dass eine Abstimmung schnell und unkompliziert erfolgen könnte.

Eine intensive Verbreitung der Projektergebnisse über verschiedene Social-Media-Kanäle ist vorgesehen. Es werden aktiv Informationen zum Thema und zum Stand im Projekt nach der Idee des Open Project (Dokumentation des Outputs, aktuellen Forschungs- und Entwicklungsstand) im eigenen Blog bereitgestellt. Die relevanten Publikationen werden in einer Mendeley-Gruppe gesammelt, der Projektoutput wird in Social-Media-Diensten wie SlideShare, Zenodo etc. publiziert und anderen Nutzenden zur Verfügung gestellt. Das Corporate Design für das Projekt wird konzipiert und umgesetzt, so dass es für verschiedene Templates (Website, Präsentationen, Poster) nachgenutzt werden kann. Ziel ist es, das Thema \*metrics über die ohnehin bereits mit dem Thema vertrauten Einrichtungen und Personen hinaus in die weiter gefasste Öffentlichkeit zu tragen. Dies umfasst vor allem Forschende, Informations- und Infrastruktureinrichtungen sowie weitere Personen, Communities und Einrichtungen, die sich mit Wissenschaft und Forschung sowie deren Evaluation befassen (Hochschulverwaltungen, wissenschaftliche Fachgesellschaften, Bibliotheken etc.). Zudem stellt das Projekt mit der Informationsplattform der \*metrics-Community eine zentrale Anlaufstelle zur Verfügung, die Theorie und Praxis in den Austausch bringt.

**Schwerpunkt:** Monat 1 bis 24

**Meilensteine:**

- M28:** Monat 2 – Start des offenen Diskurses über verschiedene Social-Media-Kanäle
- M29:** Monat 3 – Erstellung des Corporate Design
- M30:** Monat 4 – Launch der Informationsplattform \*metrics (Vorstellung z.B. bei der COAR-Jahrestagung in Wien Anfang April 2016)
- M31:** Monat 24 – Ableitung von Empfehlungen aus Projektergebnissen und Veröffentlichung DINI-Schrift (Deutsche Initiative für Netzwerkinformation e. V.)

### **AP 3.2: Forschungsfragen gemeinsam mit der Community innerhalb von Workshops bearbeiten**

Neben den virtuellen Angeboten (AP 3.1) an die Community ist ein aktiver Face-to-Face Austausch mit Interessierten geplant. Dazu wird zu Beginn und am Ende der Projektlaufzeit jeweils ein Workshop veranstaltet, der über das \*metrics und die Aktivitäten und Informationsangebote des Projekts informiert.

Ein Workshop ist im ersten Viertel der Projektlaufzeit geplant. Zielgruppe sind alle am Thema \*metrics Interessierten aus Bibliotheken (u.a. Repository-Community), Universitäten (u.a. Verwaltungen, Controlling, Forschungsinformationssysteme-Community), Verlagen und aus der Wissenschaft selbst, die \*metrics für ihre Arbeit nutzen (wollen). Dabei wird den Zielgruppen das Thema \*metrics vorgestellt und bei Bedarf auch vertiefend vermittelt. Neben dem aktuellen Stand der Forschung wird auch die praktische Umsetzung, wie sie innerhalb des Vorhabens geplant ist, vorgestellt. Ein weiterer Schwerpunkt werden die Arbeiten am Social-Media-Registry sein, die bei dem Workshop zur Diskussion gestellt werden, um ggf. weitere Entitäten für die Erfassung im Registry zu erhalten. Der zweite Workshop gegen Ende der Projektlaufzeit spricht die selbe Zielgruppe an und stellt vor allem die Ergebnisse des Projekts detailliert vor bzw. zur Diskussion, informiert über den aktuellen Forschungsstand und initiiert Kooperationsnetze für die Nachhaltigkeit und Fortentwicklung der Services. Hier werden die Beiratsmitglieder – speziell im Hinblick auf die Einordnung der Projekt-Ergebnisse – Beiträge beisteuern. Während des zweiten Workshops ist zudem ein Hackathon geplant, der insbesondere die technische Community ansprechen soll. Ziel eines solchen Hackathons ist das gemeinsame Arbeiten bzw. Entwickeln

<sup>20</sup> <http://www.bibliometrie.info/>

von Software auf Grundlage der vorliegenden Daten, Erkenntnisse und genutzter Software. Die besten Entwicklungen werden ausgezeichnet und ggf. in die weitere Arbeit integriert.

**Schwerpunkt:** Monat 5 bis Monat 7 und Monat 17 bis Monat 19

**Meilensteine:** **M32:** Monat 6 – Erster Projektworkshop  
**M33:** Monat 18 – Zweiter Projektworkshop

### AP 3.3: Projektmanagement

In diesem Arbeitspaket erfolgen die Planung, Koordinierung und Steuerung des Projekts über die gesamte Laufzeit. Das Projektmanagement stimmt auf operativer Ebene die Arbeitspakete mit den Projektteams ab und kontrolliert kritisch deren Ausführung. Durch regelmäßige Berichte ist das Projektmanagement über die Fortschritte in den einzelnen Arbeitspaketen informiert, um so sämtliche Maßnahmen zur systematischen Erkennung, Analyse, Bewertung, Überwachung und Kontrolle von Risiken durchzuführen.

Das Projektmanagement hat unmittelbaren Kontakt zu interessierten Communities, die ggf. Projektergebnisse übernehmen wollen, sowie zu den Mitgliedern des wissenschaftlichen Beirats, der zusätzlich zur Informationsplattform regelmäßig über die Entwicklungen im Projekt informiert wird. Rückmeldungen werden an die Projektbeteiligten weitergegeben. Die Mitglieder des Beirats sind Vertreter aus Forschung und Praxis: Dr. Martin Fenner (Technischer Direktor bei DataCite), Dr. Juan Gorraiz, Dr. Christian Gumpenberger (beide Bibliometrie Team der Universität Wien), Dr. Stefanie Haustein (EBSI – Université de Montréal), Lambert Heller (Leiter des Open Science Lab der TIB Hannover), Prof. Dr. Stefan Hornbostel (Leiter des Instituts für Forschungsinformation und Qualitätssicherung), Stacy Konkiel (altmetric.com), Heather Piwowar (ImpactStory), Prof. Dr. Wolfgang Semar (Hochschule für Technik und Wirtschaft Chur, Schweiz).

Im Rahmen dieses Arbeitspakets wird die gesamte Projektkommunikation koordiniert. Für die interne Kommunikation werden Mailinglisten zur Verfügung gestellt und administriert. Als Arbeitsplattform wird ein Wiki genutzt, das durch das Projektmanagement verwaltet und strukturiert wird. Ein weiterer Arbeitsschwerpunkt ist die Organisation und regelmäßige Durchführung von Projekttreffen und Videokonferenzen. Diese werden durch das Projektmanagement vor- und nachbereitet.

**Schwerpunkt:** Monat 1 bis Monat 24

**Meilensteine:** **M34:** Monat 12 – Zwischenbericht  
**M35:** Monat 24 – Abschlussbericht

## 2.4 Maßnahmen zur Erfüllung der Förderbedingungen und Umgang mit den Projektergebnissen

Das Ziel der Förderung ist die wissenschaftlich fundierte Durchdringung der Aussagekraft und Wahrnehmung von \*metrics. Projektergebnisse werden daher in wissenschaftlichen Publikationsorganen veröffentlicht und erhobene Daten und angepasste Software – soweit dies datenschutzrechtlich gestattet ist – zur Nachnutzung bereitgestellt. Im Einzelnen:

- ✓ Für die Entwicklung der prototypischen Infrastruktur sowie die Bereitstellung des Quellcodes samt Dokumentation aus AP 2 bei GitHub ist die VZG zuständig.
- ✓ Alle Projektergebnisse werden als Teil der Informationsplattform \*metrics präsentiert. Die SUB zeichnet sich für die Nachhaltigkeit dieser Plattform zuständig.
- ✓ Die Projektpublikationen werden sowohl in der Projektdokumentation abgebildet als auch in OA-Repositorien der jeweiligen Autoreninstitutionen abgelegt. Darüber hinaus werden die Publikationen über diverse Social-Media-Kanäle verbreitet. Verantwortlich dafür sind die SUB und die ZBW.
- ✓ Die Ergebnisse der Studien werden für Publikationen und Präsentationen verwendet und durch GESIS als Forschungsdaten auf Repositorien wie Zenodo bzw. GESIS' Datorium bereitgestellt.
- ✓ Die Ergebnisse des Projekts fließen zudem in die Arbeitsgruppen der DINI (AG EPub und AG FIS), einschließlich der Weiterentwicklung des DINI-Zertifikats, sowie in die Arbeiten zur NISO Alternative



Assessment Metrics (Altmetrics) Initiative ein. Die Projektpartner sind größtenteils Mitglieder in den Arbeitsgruppen (AG EPub: SUB, VZG, ZBW; AG FIS: SUB und NISO Arbeitsgruppe: SUB, ZBW).

- ✓ Der Austausch mit den verschiedenen Fachcommunities wird durch SUB, ZBW und GESIS sichergestellt.

### 3 Literaturverzeichnis

Das ausführliche Literaturverzeichnis befindet sich in Anlage 3.

## 5 Voraussetzungen für die Durchführung des Vorhabens

### 5.1 Angaben zur Dienststellung

Reiner Diedrichs (Direktor, VZG), Wolfram Horstmann (Direktor, SUB Göttingen), Isabella Peters (Professorin, ZBW Kiel/Hamburg; Christian-Albrechts-Universität zu Kiel), Klaus Tochtermann (Direktor, ZBW Kiel/ Hamburg; Professor, Christian-Albrechts-Universität zu Kiel), Christof Wolf (Präsident, GESIS; Professor, Universität Mannheim), Markus Strohmaier (Abteilungsleiter, GESIS; Professor, Universität Koblenz-Landau), Katrin Weller (Teamleiterin Social Analytics and Services, GESIS)

### 5.2 Zusammensetzung der Projektarbeitsgruppe

Jeweils aus Hausmitteln der eigenen Institutionen finanziert:

- GESIS:** Prof. Dr. Markus Strohmaier (Abteilungsleitung Computational Social Science), Katrin Weller (Teamleiterin Social Analytics and Services)
- SUB:** Margo Bargheer (Gruppenleitung Elektronisches Publizieren), Dr. Birgit Schmidt (Scientific Manager), Dr. Arnulf Timm (Abteilungsleiter Zentrale Informationsversorgung)
- VZG:** Justine Haeberli-Kaul (Abteilung Bibl. Dienste), Dr. Andreas Krausz (Abteilung Anwendungsbetreuung)
- ZBW:** Prof. Dr. Isabella Peters (Arbeitsgruppe Web Science), Kaltrina Nuredini (Arbeitsgruppe Web Science)

### 5.3 Zusammenarbeit mit anderen Institutionen und anderen Wissenschaftlerinnen und Wissenschaftlern

#### Institutionen oder Wissenschaftlerinnen und Wissenschaftler, mit denen für dieses Vorhaben eine konkrete Vereinbarung besteht

- ✓ Von den Mitgliedern des Beirats besteht eine Zusage, dass sie das Projekt im Rahmen der Beiratsmitgliedschaft aktiv unterstützen.
- ✓ Die Confederation of Open Access Repositories (COAR) unterstützt das Projekt bei seinem Vorgehen und hat dies in einem Letter of Support bestätigt (siehe Anlage 5.1).
- ✓ Der Dienstanbieter Altmetric LLP unterstützt das Projekt mit einem Zugang auf seine gesamten Altmetrics-Daten und Bestätigt es in einem Letter of Support (Anlage 5.2).
- ✓ Die TIB Hannover unterstützt das Projekt und hat die Absicht, Testdaten und Erfahrungen des Projekts nachzunutzen (siehe Letter of Intent in Anlage 5.3).
- ✓ Die DINI Arbeitsgruppe Elektronisches Publizieren unterstützt das Projekt mit ihrer Expertise.

#### 5.3.1 Institutionen, Wissenschaftlerinnen und Wissenschaftler, mit denen in den letzten drei Jahren gemeinsame Projekte durchgeführt wurden

Im Folgenden werden die Institutionen, Wissenschaftlerinnen und Wissenschaftler genannt, die im Rahmen des Projektantrags inhaltlich relevante Zusammenhänge aufweisen:

- GESIS:** FoKoS Uni Siegen; Forschungsdatenzentrum der Bundesagentur für Arbeit (BA) im Institut für Arbeitsmarkt- und Berufsforschung (IAB); GWDG; Rat für Sozial- und Wirtschaftsdaten Berlin; SOEP Berlin; WeST Uni Koblenz; ZBW
- SUB:** Akademie der Wissenschaften zu Göttingen; HU Berlin/CMS; Max Planck Digital Library; SULB Saarbrücken; UB Bielefeld; UB Konstanz; UB FU Berlin; UB Osnabrück; UB Stuttgart; VZG
- VZG:** BBAW; BSB München; BSZ; DNB; GESIS; hzb; HeBis; HU Berlin; IntraData GmbH; MPDL; Staatsbibliothek zu Berlin; Staatliche Museen zu Berlin; Stiftung Preußischer Kulturbesitz; SUB Göttingen; SUB Hamburg; SULB Saarbrücken; TIB/UB Hannover; UB Frankfurt a. M.; UB Freiburg; UB Köln; UB Regensburg; UB Stuttgart; ZB MED; ZBW
- ZBW:** DIW; GESIS; IfW; Kompetenzzentrum der Universität Trier; INNO-tec – Institut für Innovations-

forschung, Technologiemanagement und Entrepreneurship der LMU München; Rat für Sozial- und Wirtschaftsdaten Berlin; SOEP Berlin; WZB Wissenschaftszentrum Berlin für Sozialforschung

## **5.4 Erklärungen zur Erfüllung der Förderbedingungen**

### **5.4.1 Bereitstellung der Projektergebnisse**

Wir verpflichten uns...

1. ...die aus dem Projekt resultierenden Publikationen ebenso wie einschlägige Dokumentationen im Open Access verfügbar zu machen und Dritten zur umfassenden Nachnutzung bereitzustellen.
2. ...den Quellcode der im Projekt entwickelten Software nach den Prinzipien von Open Source zu dokumentieren und für die Nachnutzung durch Dritte verfügbar zu machen.

### **5.4.2 Regeln guter wissenschaftlicher Praxis**

Wir verpflichten uns, mit der Einreichung des Antrags auf Bewilligung einer Sachbeihilfe bei der DFG, die Regeln guter wissenschaftlicher Praxis einzuhalten.

\*Dies ist eine gekürzte Form der Antragsunterlagen ohne Angaben zu Personal- und Finanzaufwendungen.