„Quo vadis offene Wissenschaft in Berlin und Brandenburg“

Bericht über die vierte Runde der Veranstaltungsreihe im Rahmen der International Open Access Week 2024/25

Lea Satzinger

Vom 21. Oktober 2024 bis 11. Februar 2025 fand die vierte Veranstaltungsreihe für Berlin-Brandenburg im Kontext von Open Access unter dem Motto „Quo vadis offene Wissenschaft“ statt. An den insgesamt vier Terminen nahmen circa 300 Interessierte teil. Die Veranstaltungsreihe bot ein breites Themenspektrum von Open-Access-Monitoring über Open Access für Objekte, Künstlicher Intelligenz bis hin zu offener Datenkultur. Sie wurde vom Open-Access-Büro Berlin, der Vernetzungs- und Kompetenzstelle Open Access Brandenburg, dem Bibliotheksverbund KOBV, dem Institut für Bibliotheks- und Informationswissenschaft (IBI) der Humboldt-Universität zu Berlin und der Technischen Hochschule Wildau ausgerichtet.

21. Oktober 2024: Open-Access-Monitoring

„Kenne die Daten besser als der Verlag“ – dies habe laut Jochen Johannsen, Direktor der Universitätsbibliothek der RWTH Aachen und Mitglied der DEAL-Gruppe, bei den DEAL-Verhandlungen eine große Rolle gespielt. Das alte Subskriptionssystem sei bei den angesetzten Kosten pro Einrichtung in hohem Maße willkürlich gewesen; das neue Paper-Charge-Modell weise hingegen weitaus größere Transparenz und Gleichheit auf. Erreichbar sei dies durch sehr genaue Daten der teilnahmeberechtigten Einrichtungen gewesen, die der DEAL-Verhandlungsgruppe zur Verfügung standen.

Ein offen nutzbares deutschlandweites Monitoring von Publikationen wird seit 2018 sukzessive durch den Open-Access-Monitor ermöglicht, der als Datenquellen unter anderem das Web of Science, Scopus sowie OpenAlex nutzt. Abgebildet werden im Monitor bisher wissenschaftliche Zeitschriftenartikel, wobei eine Ausweitung auf weitere Dokumententypen wie conference proceedings – die auch bereits Teil von Transformationsverträgen sind – in Planung sei, erläuterte Irene Barbers, Leiterin des Fachbereichs Literaturerwerbung in der Zentralbibliothek des Forschungszentrums Jülich.[1]

OpenAlex stelle, so Najko Jahn, Data Analyst und stellvertretender Leiter der Gruppe Wissen als Gemeingut an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen, eine vielversprechende Entwicklung dar. Im Unterschied zu kommerziellen Datenbanken, die vor allem auf Algorithmen für die Datenbereinigung setzen, sei die Verbesserung von Angaben durch direkte Rückmeldungen der Einrichtungen beim offenen System OpenAlex einfacher möglich. Wie inklusiv OpenAlex ist, sei aktuell Gegenstand einer Untersuchung des Kompetenznetzwerks Bibliometrie. Die Skalierbarkeit offener Systeme sei ein entscheidender Aspekt für deren Erfolg und im Fall von OpenAlex noch abzuwarten.

Jochen Johannsen griff diesen Punkt auf und verdeutlichte, dass offene Daten für Hochschulen und deren Bibliotheken von großer Bedeutung seien. Von Hochschulleitungen würden neben einem detaillierten Blick auf eigene Leistungszahlen zum Vergleich auch Daten anderer Einrichtungen gewünscht. Dabei gelte das Prinzip der Reziprozität: Einrichtungen profitieren gegenseitig, wenn sie eigene Daten teilen und die anderer nachnutzen können. Für Publikations- und Kosten-Reportings auf Landesebene wie openaccess.nrw seien offene Daten essentiell.

1 Irene Barbers ist am 8. Januar 2025 verstorben. Die bibliothekarische Community betrauert diesen Verlust: https://www.fz-juelich.de/de/zb/aktuelles/meldungen/2025/nachruf-irene-barbers-2025-01-10

Auch bei der Berliner Landesinitiative sei der Aufbau eines offenen Monitoringsystems zur Erfassung und Sichtbarmachung erfolgt, erläuterte Maxi Kindling als Leiterin des Open-Access-Büros Berlin.[2] Im oa.atlas werden Open-Access-Policies & -Services der Einrichtungen erfasst, und das System mittlerweile auch deutschlandweit genutzt. Von Heinz Pampel, dem Moderator der Session, tätig am Institut für Bibliotheks- und Informationswissenschaft der HU Berlin, wurde an dieser Stelle erfragt, ob beim Monitoring nicht auch ein breiteres Portfolio als rein Open Access in den Blick genommen werden sollte. Good Practices sollten, so Maxi Kindling, sichtbar gemacht werden, allerdings sei es in der Praxis ein weiter Weg der Umsetzung an allen Einrichtungen, umfassende Daten in Hochschulbibliographien oder Forschungsinformationssystemen abzubilden. Najko Jahn ergänzte, dass ein qualitativer, explorativer Ansatz zur Erhebung von Daten genutzt werden könnte, jedoch kein umfassendes Monitoring, da Open Science in seinen Fragestellungen sehr breit ausgerichtet sei. Für Open Access seien lokale und globale Trends ermittelbar. Umgesetzt sei dies bereits im Hybrid Open Access Dashboard, das in konsortialen Verhandlungen sowie auch von Seiten der Verlage Nachnutzung erfahre und sich aus offenen Daten speise.

Die Auseinandersetzung mit Zahlen zur Open-Access-Transformation nehme, so Jochen Johannsen, zu, sowie auch die nötige Sprechfähigkeit gegenüber Präsidien, Rektoraten oder Fakultäten. Auch für die strategische Steuerung bei Gesprächen mit Hochschulleitungen seien Daten essentiell, beispielsweise bei der Implementierung von Informationsbudgets, ergänzte Irene Barbers. Auch Maxi Kindling sieht dies für die Kommunikation auf politischer Ebene am Beispiel der Open-Access-Quoten für Berliner Einrichtungen gegeben.

20. November 2024: Open Access für Objekte

Die zweite Veranstaltung der „Quo vadis"-Reihe nahm Open Access für nichttextuelle Medien in den Blick. Auf dem Podium diskutierten Friederike Kramer (Universität der Künste Berlin), Ben Kaden (Vernetzungs- und Kompetenzstelle Open Access Brandenburg, Projekt WiNoDa) und Melanie Seltmann (Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, Datenkompetenzzentrum QUADRIGA); moderiert wurde die Veranstaltung von Sophie Kobialka (Fachhochschule Potsdam, Projekt WiNoDa).

Zum Einstieg wurde die Frage erörtert, von welchen Inhalten bei nichttextuellen Medien gesprochen wird. Friederike Kramer erläuterte, dass an Kunstuniversitäten mit diversesten Datenformaten gearbeitet werde, die nicht immer in Verbindung mit Forschungstexten stünden, beispielsweise Installationen in der bildenden Kunst. Im Projekt QUADRIGA werde, so Melanie Seltmann, unter anderem mit Fokus auf bewegtem Bild gearbeitet und Korpusbildung sowie Annotation seien Schwerpunkte. Ben Kaden beschrieb, dass das Projekt WiNoDa sich vor allem mit Sammlungen in Museen befasse, wobei digitalisierte Objekte und damit Dateiformate und Softwarelösungen eine zentrale Rolle spielten.

Die Open-Access-Konzepte sind abhängig vom Format entsprechend breit angelegt. Im Kontext bewegtes Bild stelle die Verfügbarkeit einer geringen Zahl passender Daten eine Herausforderung für die Erstellung von Open Educational Resources dar. Wichtig seien rechtliche Erlaubnisse für den Zugriff, die Bearbeitung sowie die Veröffentlichung angepasster Inhalte. In musealen Sammlungen hingegen sei Public Engagement bereits mitgedacht und es werde vor allem überlegt, wie didaktisch Verständnis aufgebaut werden könne und was die passende Zugangsform sei, um das Zielpublikum optimal zu erreichen. An Kunstuniversitäten spiele die schöpferische Leistung als wirtschaftliche Grundlage der Kunstschaffenden eine zentrale Rolle und es sei ein gewisses Spannungsfeld der Nachnutzung von Inhalten anderer und der Beschränkung eigener Inhalte gegeben. Bei dieser Fragestellung sei ein Generationenwechsel zu beobachten, mit zunehmendem Blick auf das Profitieren von geteilten Inhalten mit umfassenden Möglichkeiten der Nachnutzung.

Ein schwieriges Thema stelle aufgrund der breiten Vielfalt nichttextueller Inhalte die als nächste aufgeworfene Frage des Monitorings dar. Für Kunstschaffende sei Nachnutzung frei geteilter Inhalte einerseits ein positiver Faktor, werde andererseits aber auch als verpasstes Potenzial für monetäre Möglichkeiten gesehen. Schwierig sei, Leistungsmessungen vom Publikationsoutput abhängig zu machen, da nichttextuelle Medien in den gängigen Nachweissystemen und darauf aufbauenden landesweiten Open-Access-Reportings nicht verzeichnet seien. Von den Diskutantinnen und Diskutanten wurde zudem erörtert, dass es an konkreten Fragestellungen fehle: Die reinen Zahlen seien teils weniger von Interesse; relevant sei, in welcher Form das Objekt nachgenutzt werde, beispielsweise als Download oder für einen Remix.

Abschließend wurde ein Blick auf verfügbare Infrastrukturen und Plattformen geworfen. Für das multimediale Publizieren stehe nicht die eine Plattform als Goldstandard zur Verfügung; herausfordernd sei, dass bestimmte Medienkomplexitäten wie vernetzte dynamische Dokumente oder umfassende Annotationsap-

2 Maxi Kindling ist seit 01.01.2025 an der Universitätsbibliothek der TU Berlin tätig.

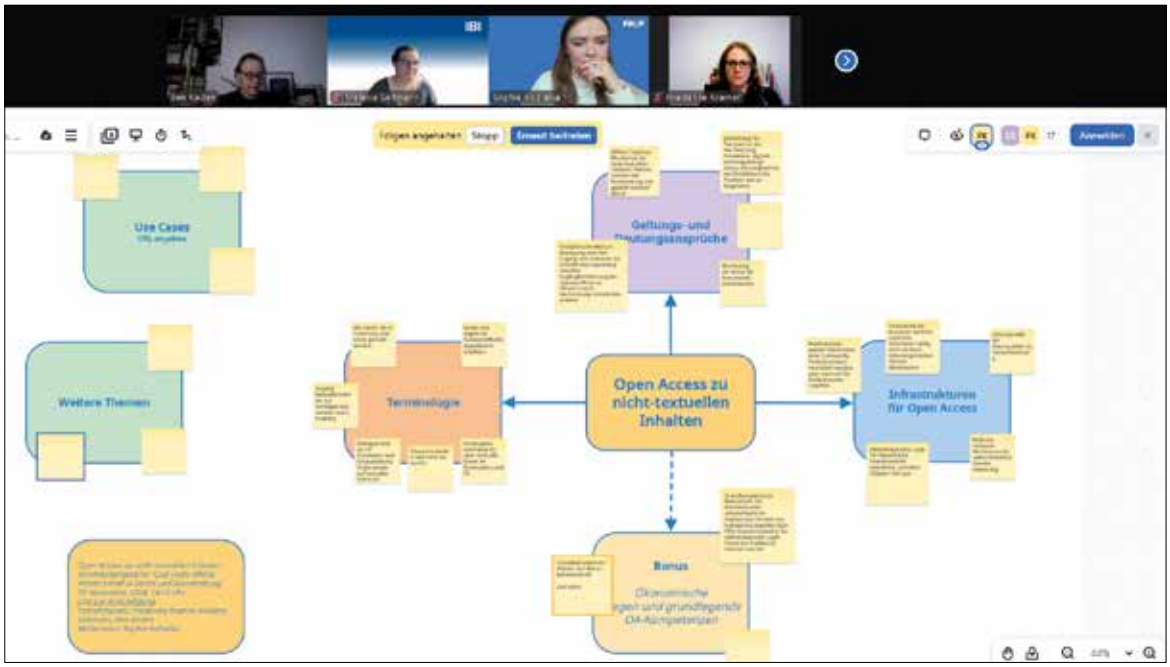

Abbildung 1: Die Diskutantinnen und Diskutanten des zweiten Veranstaltungstermins sowie das in der Session befüllte Whiteboard; von links nach rechts: Ben Kaden (VuK Open Access Brandenburg), Melanie Seltmann (Institut für Bibliotheks- und Informationswissenschaft der HU Berlin), Sophie Kobialka (Fachhochschule Potsdam; Moderation), Friederike Kramer (Universität der Künste Berlin)

parate in gängigen Systemen nicht abbildbar seien. Genutzt würden institutionelle, fachspezifische oder übergreifende Repositorien wie media/rep oder Zenodo sowie kommerzielle Angebote wie YouTube. Für Kulturerbe-Daten seien Plattformen wie die Deutsche Digitale Bibliothek, museum-digital oder Europeana von Bedeutung. Häufig würden von Publizierenden abhängig vom Zweck der Veröffentlichung – Hinterlegung, Auffindbarkeit, Nachnutzung, Langzeitarchivierung – diverse Wege genutzt und darauf basierend eine oder auch mehrere Plattformen gewählt. Relevant sei, dass Beratungsmöglichkeiten, beispielsweise an den lokalen Bibliotheken, bekannt seien und bereits frühzeitig im Publikationsprozess genutzt würden.

18. Dezember 2024: KI und Open Access

Auch KI und die damit verbundenen rechtlichen Fragestellungen spielen im Kontext von Open Access und Open Science eine zunehmende Rolle. Diesen Aspekten widmete sich der dritte Teil der Veranstaltungsreihe. Frank Seeliger (Technische Hochschule Wildau) begrüßte auf dem Podium Judith Ludwig (Technische Informationsbibliothek), Fabian Rack (FIZ Karlsruhe, iRights.Law) sowie Christof Schöch (Universität Trier). Die Panelisten wurden zu kurzen Statements mit anschließender Diskussion und Fragen aus dem Publikum eingeladen.

Den Beginn machen durfte Judith Ludwig, die Unterschiede im Rechtsrahmen für analoge und digitale Werke aufzeigte. Im analogen Raum bestehe ein klarer Rechtsrahmen, wohingegen im digitalen Raum die Welt noch sehr im Umbruch sei und an vielen Punkten große Unsicherheit herrsche, da Regelungen noch nicht „durchreguliert" seien. Relevant sei vor allem das 2018 verabschiedete UrhWissG als „Dach" für Bibliotheken und dessen § 60a bis § 60e, die Erlaubnisse, Möglichkeiten und Grenzen für Nutzungen festlegen. Die unklare Rechtslage fördere dabei einen Chilling-Effekt: Einrichtungen würden das Risiko von Rechtsverstößen minimieren und Nutzung unterlassen, was zu einer weiteren Manifestation der Unklarheit und letztlich Verkleinerung der Spielräume führe. Gutes Beispiel hierfür seien Lizenzverträge, die den Einsatz von KI-Methoden untersagen, und gesetzliche Vorgaben, die Text- und Data-Mining erlauben, aber unklar lassen, inwieweit der Einsatz und das Training generativer KI hier als einbezogen betrachtet werden dürfen. Einrichtungen rät Judith Ludwig, Open Science aktiv zu fördern, Allianzen zu bilden und restriktive Geschäftspraktiken aktiv in Frage zu stellen sowie den Rechtsrahmen zu verstehen, zu nutzen und zu vermitteln.

Als zweites gab Fabian Rack einen kurzen Überblick über die CC-Lizenzen (Creative Commons) zur Freigabe urheberrechtlich geschützter Inhalte. Nutzungsvorbehalte gegenüber KI stellen bei CC einen gewissen Widerspruch dar, da es keine No-AI-CC-Lizenz gibt, sprich das Einfließen in KI-Modelle immer möglich ist.[3] Weitere

3 siehe auch https://creativecommons.org/2023/08/18/understanding-cc-licenses-and-generative-ai/

Regelungen, beispielsweise zum Datenschutz, können einer Nutzung jedoch entgegenstehen. Für KI-Output sprach Fabian Rack die Empfehlung aus, diesen unter CC0 freizugeben.

Als abschließendes Statement ging Christof Schöch auf die Bedeutung von KI und Open Science für die Digital Humanities ein. Forschung werde in diesem Bereich vor allem zu nicht urheberrechtlich geschützten Materialien betrieben, da Texte des 20. Jahrhundert vielfach rechtlich nutzbar, aber nicht weiter verbreitbar seien. Dies stelle bei Korpora für die Reproduzierbarkeit und Nachnutzbarkeit ein Problem dar. Verwendet würden daher sogenannte „abgeleitete Texte“: Der ursprüngliche Volltext wird durch Transformationsmethoden verändert und im Informationsgehalt reduziert, beispielsweise durch Anpassung der Wortreihenfolge, sodass urheberrechtlich Relevantes nicht mehr enthalten ist. Der Output genießt somit keinen Werkschutz; er ist für Menschen nicht lesbar und durch KI nicht rekonstruierbar. Möglich ist jedoch die Anwendung einer Reihe von Standardmethoden der Digital Humanities wie die Zählung der Häufigkeit von Wortarten. Für die Digital Humanities seien bei Korpora neben Volltexten auch qualitätsvolle Metadaten gewünscht – und hier betonte Christof Schöch die hohe Metadatenqualität der Bibliotheken, wobei diese nicht immer aus Sicht der maschinellen Lesbarkeit gedacht sei, zum Beispiel bei der Codierung des Geschlechts. Schließlich warb Christof Schöch dafür, dass Bibliotheken nicht rein auf geäußerte Bedarfe der Wissenschaft reagieren sollten, sondern die Forschung proaktiv mit neuen Lösungen im Rahmen von KI-Anwendungen unterstützen.

Frank Seeliger leitete an dieser Stelle zur Diskussion über und fragte die Panelisten, welche neuen Bibliotheks-Services denkbar wären. Als Beispiel nannte Judith Ludwig ORKG.ask der TIB, das wissenschaftliche Fragestellungen auf Basis von CC-lizenzierten Inhalten der universitären Repositorien beantworte. Abschließend nahm der Moderator noch Rückgriff auf einen Vortrag von Georg Rehm zu europäischen Sprachmodellen und stellte die Frage, ob die EU in Sachen KI auf Beschränkung, China und die USA hingegen auf Entwicklung fokussiert seien. Fabian Rack hob positiv hervor, dass der AI Act KI-Anwendungen in Risikoklassen unterteile, wies jedoch auch auf den daraus resultierenden hohen Umsetzungs- und Beratungsaufwand hin. Judith Ludwig stellte die Regelungen in der EU und den USA gegenüber und formulierte, absichtlich etwas zugespitzt: In der EU sei alles, was nicht ausdrücklich erlaubt sei, verboten; in den USA hingegen würden die Fair-Use-Prinzipien gelten und Freiheiten einräumen; werden diese Freiheiten ausgereizt, würden gegebenenfalls Verfahren angestrengt, wobei auch im KI-Kontext die nähere Klärung des Rechtsrahmens in den USA abzuwarten sei. Bemerkbar sei ein Brüssel-Effekt, sprich die Übernahme von Rechtsrahmen und Regulierungsmaßnahmen der EU in außereuropäischen Märkten.

11. Februar 2025: Offene Datenkultur

Die abschließende Veranstaltung der Reihe war dem Thema „Offene Datenkultur“ mit Fokus auf institutionellen und Sammlungsdaten gewidmet. Julia Boltze-Fütterer vom Kooperativen Bibliotheksverbund Berlin-Brandenburg lud die Panelisten zu Impulsvorträgen und Diskussion ein.

Abbildung 2: Die Diskutantinnen und Diskutanten des vierten Veranstaltungstermins (von links oben nach rechts unten): Alexander Winkler (digiS – Forschungs- und Kompetenzzentrum Digitalisierung Berlin), Anna Gnyp (Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland, Projekt SODa), Julia Boltze-Fütterer (Kooperativer Bibliotheksverbund Berlin-Brandenburg; Moderation), Tobias Weiß (Janusz-Korczak-Bibliothek Pankow), Jörg Lehmann (Staatsbibliothek zu Berlin), Sophie Kobialka (Fachhochschule Potsdam, Projekt WiNoDa)

Die Nachnutzung von Daten wurde von Alexander Winkler vom digiS – Forschungs- und Kompetenzzentrum Digitalisierung Berlin als zentraler Aspekt in den Mittelpunkt seines Vortrags gestellt. Systeme mit qualitätsvollen Metadaten seien vielfach vorhanden, jedoch sei die Nutzbarkeit der Systeme verbesserungsfähig. Als Beispiel griff Alexander Winkler den Katalog VD17 heraus, der für Datenabfragen über eine SRU-Schnittstelle verfügt. Für deren Nutzung seien jedoch bibliotheksinformatische Grundkenntnisse zu XML und dem PICA-Format notwendig, weshalb zusätzlich auch ein JSON-Dump bereitgestellt wurde. Hürden wie hoher Zeitaufwand oder benötigte Fachkenntnisse würden häufig von einer Nachnutzung abschrecken. Einrichtungen könnten dem durch nutzerzentrierte Angebote wie die Bereitstellung einer guten Dokumentation, interaktiver Notebooks und Anwendungsbeispielen entgegenwirken. Im Sinne der FAIR-Prinzipien sollte zudem eine offene Lizenzierung der Daten erfolgen sowie Standards für Formate und Vokabulare zum Einsatz kommen.

Auch Jörg Lehmann von der Staatsbibliothek zu Berlin betonte die Verfügbarkeit umfassender Datensätze mit qualitätsvollen Metadaten. So stellt die Staatsbibliothek zu Berlin unter anderem OCR-erschlossene Volltexte, Bilder und Katalogdaten zur Verfügung, die über diverse Schnittstellen – SRU, OAI-PMH, IIIF – mit Hilfe von Tutorials für die Programmiersprachen Python und R nachnutzbar sind. Relevant sei, KI-Datensets im Hinblick auf Biases und statistische Schiefen zu kontextualisieren, wobei man hier noch auf reine Beschreibungen angewiesen sei, da Metriken für Biases bisher nicht existieren.

Einen Blick auf Open Data aus Sicht der Öffentlichen Bibliotheken wagte Tobias Weiß von der Janusz-Korczak-Bibliothek Pankow, die ihre Ausleihstatistik seit 2023 online zur Verfügung stellt. Openness sei für Öffentliche Bibliotheken ein umfassenderes Thema, etwa im Hinblick auf die Ausrichtung von Angeboten auf diverse Nutzergruppen. Jedoch fehle es bisher an entsprechend definierten Stellenprofilen und damit Know-how, da andere Themen wie die Weiterentwicklung technischer Infrastruktur im Fokus stünden. Dementsprechend ist die Stadtbibliothek Pankow innerhalb des Verbundes der Öffentlichen Bibliotheken Berlins bisher die einzige Daten-publizierende Einrichtung.

Im folgenden Impuls nahm Anna Gnyp von der Koordinierungsstelle für wissenschaftliche Universitätssammlungen in Deutschland und vom Projekt SODa die Sammlungen der Universitäten und Hochschulen in den Blick. Auch diese seien mit Hürden bei der Verfügbarkeit geeigneter Infrastruktur und personeller Ressourcen mit Datenkompetenzen konfrontiert. Zusätzlich hinzu träten Unsicherheiten im Umgang mit sensiblen Objekten wie Humanpräparaten und damit verbunden Fragestellungen des Urheber- und Persönlichkeitsrechts sowie der Provenienzforschung. Fehlende Standards für beispielsweise 3D-Digitalisierung und mangelnde Sichtbarkeit der Sammlungen seien weitere Herausforderungen. Insgesamt existieren in Deutschland circa 1.200 wissenschaftliche Universitätssammlungen, für die das Projekt SODa bis Ende 2026 den Aufbau eines Datenkompetenzzentrums zur Erhöhung der digitalen Zugänglichkeit anstrebt.

Neben Hochschulen verfügen auch Museen über umfassende naturkundliche Sammlungen und objektzentrierte Daten. Diese stehen, so Sophie Kobialka von der Fachhochschule Potsdam, im Projekt WiNoDa im Mittelpunkt. Am Projekt beteiligt ist unter anderem das Museum für Naturkunde Berlin mit einer umfangreichen Sammlung von über 30 Millionen physischen und digitalen Objekten. WiNoDa wird bis Ende 2026 den Kompetenzaufbau zu Open Data fördern und stellt Selbstlernmodule als Open Educational Resources zur Verfügung. Zudem ist eine Bedarfsanalyse zu Open Science und Public Engagement geplant, um die Nutzbarkeit und Erforschung naturwissenschaftlicher Sammlungen zu erhöhen, Akteure zu vernetzen und eine Einbindung der Gesellschaft zu erreichen.

In der Abschlussdiskussion wurde hervorgehoben, dass die Kontexte Forschender von den Einrichtungen mitgedacht werden müssen. Umgekehrt seien Forschende dazu aufgefordert, ihre Bedarfe gegenüber Institutionen zu äußern. Vorhandene Angebote sollten beworben und somit breit wahrgenommen werden.

Fazit & Ausblick

Die „Quo vadis"-Reihe startete im Rahmen der Open Access Week unter dem Motto „Community over Commercialization" und bot wieder ein gelungenes Forum für die Reflektion und den Austausch zu aktuellen Fragestellungen. Mit einem breiten Spektrum an Themen machte die Reihe deutlich, dass Open Science in all seinen Facetten für öffentliche und wissenschaftliche Einrichtungen mittlerweile eine zentrale und weiter zunehmende Rolle spielt. Neben dem Fokus auf Open-Access-Finanzierung und -Monitoring rücken auch weitere Aspekte von Open Science wie Open Data und Open Educational Resources tiefer in den Blick. Zusätzlich wird das Voranschreiten von Open Science auch durch aktuelle Entwicklungen wie Künstliche Intelligenz beeinflusst, sodass der Themenbereich weiterhin dynamisch bleiben wird. Für eine Fortsetzung der Reihe ab Herbst 2025 ist daher wieder ein breit gefächerter Austausch zu erwarten.

Lea Satzinger
stellv. Leitung Medienerwerbung und -erschließung, ThULB Jena
ORCID: https://orcid.org/0000-0001-8763-9840