

Herausforderung Datenqualität – Forschung im digitalen Wandel

Interdisziplinäre Konferenz des Rates für Informationsinfrastrukturen (RfII)
mit Unterstützung der VolkswagenStiftung am 27.–28. Februar 2020 in Hannover

Zusammenfassender Konferenzbericht

Der Rat für Informationsinfrastrukturen (RfII) legte im September 2019 unter dem Titel „Herausforderung Datenqualität“ Empfehlungen zur Zukunftsfähigkeit von Forschung im digitalen Wandel vor. Hierauf aufbauend veranstaltete er im Februar 2020 mit Unterstützung der VolkswagenStiftung eine zweitägige interdisziplinäre Konferenz mit Repräsentanten aus Wissenschaftspolitik, Forschungsförderung und verschiedenen wissenschaftlichen Disziplinen im Schloss Herrenhausen in Hannover. Diese Veranstaltung sollte zu einem übergreifenden Diskurs zur Qualität von (digitalen) Forschungsdaten anregen.

Neben wissenschaftspolitischen Keynotes von MARTINA BROCKMEIER, der ehemaligen Vorsitzenden des Wissenschaftsrates, PETRA GEHRING, der Vorsitzenden des Rates für Informationsinfrastrukturen sowie RALF WEHRSPHORN, Vorstand für den Bereich Technologiemarketing und Geschäftsmodelle der Fraunhofer Gesellschaft, umfasste das Programm unter anderem Sessions zu Herausforderungen für die Datenqualität und das Datenqualitätsmanagement in den Bereichen Modellierung, Simulation und KI und der empirischen Forschung in den Sozialwissenschaften. In einer weiteren Session wurden Datenqualitätsprobleme in der Wirtschaft sowie Praxis und Herausforderungen für eine globale Forschungsdateninfrastruktur in den Lebenswissenschaften diskutiert. Eine offene Paneldiskussion mit Vertretern aus Wissenschaftspolitik, Forschung und dem wissenschaftlichen Infrastrukturbereich rundete das Portfolio ab. Moderiert wurden die einzelnen Programmteile von LARS BERNARD, KURT KREMER, STEFAN LANGE, STEFAN LIEBIG, WOLFGANG MARQUARDT und JAN-MARTIN WIARDA.

In einer Evening Lecture verdeutlichte der ehemalige Generaldirektor der Europäischen Organisation für Kernforschung (CERN), ROLF-DIETER HEUER, wie die Forscher des CERN über Jahrzehnte eine ebenso nachhaltige wie Community-gestützte Datenqualitätskultur etabliert haben und was andere wissenschaftliche Communities von diesem Beispiel lernen könnten. Eingerahmt wurden diese Beiträge durch Grußworte von GEORG SCHÜTTE, Generalsekretär der VolkswagenStiftung, STEFAN LIEBIG, dem stellvertretenden Vorsitzenden des Rates für Informationsinfrastrukturen, MICHAEL MEISTER, dem Parlamentarischen Staatssekretär im Bundesministerium für Bildung und Forschung, sowie von einer die Tagung beschließenden Zusammenfassung der Erträge durch PETRA GEHRING. Alle Vortragenden waren sich einig, dass Datenqualität – und die hierfür notwendigen Kenntnisse, Infrastrukturen und organisatorischen wie rechtlichen Rahmenbedingungen – im Zeitalter der umfassenden Digitalisierung von Wirtschaft und Gesellschaft zu einem der zentralen Themen der institutionellen Strukturen, der Selbststeuerung aber auch des Selbstverständnisses von Wissenschaft im 21. Jahrhundert geworden ist. Der hierfür erforderliche „Kulturwandel“ in allen akademischen Disziplinen müsse weiter gefördert und von kluger wissenschaftspolitischer Rahmen- und Anreizsetzung begleitet werden. Vortragende und Teilnehmer der Tagung waren einhellig der Meinung, dass der Rat für Informationsinfrastrukturen mit seinem Positionspapier und dieser Veranstaltung ein sehr begrüßenswertes Signal gesetzt habe, das Thema Datenqualität weithin sichtbar zu machen. Ein Signal, das freilich erst den Beginn eines umfassenden globalen Diskurses über Datenqualität in der Wissenschaft markieren kann.

Zu den einzelnen Themenfeldern der Tagung

Keynotes

Aus der Perspektive der Keynote Speaker wurde festgestellt, dass das Thema Datenqualität schon immer in der Wissenschaft (zumeist im Bereich der Methodenkritik) virulent gewesen sei, aber inzwischen auch angesichts der gestiegenen Erwartungshaltung aus Politik und Gesellschaft in Bezug auf offen zugängliche (Forschungs-) Daten deutlich an Dynamik und Dringlichkeit gewonnen habe.

Da die Problemlösungsfähigkeit wissenschaftlicher Forschung heute häufig an den Schnittstellen verschiedener Fächer liege und zahlreiche Forschungsfragen nur noch interdisziplinär gelöst werden könnten, stelle dies hohe Anforderungen an die Dokumentation und „Erklärbarkeit“ von Daten für fachexterne wissenschaftliche Anwendungen. Aber auch innerhalb der einzelnen Disziplinen sei der Mangel an Reproduzierbarkeit von Studien – und eine fehler- oder lückenhafte Dokumentation von Forschungsdaten sowie der Verfahren und Geräte, mit denen sie gewonnen und analysiert werden – zu einem gravierenden Problem geworden.

Von der heute auch auf politischer Seite erhöhten Aufmerksamkeit für Datenqualität könne die Wissenschaft insgesamt und vor allem langfristig profitieren: mit den FAIR-Prinzipien und ihrer Aufnahme in den neuen Kodex der Deutschen Forschungsgemeinschaft (DFG) seien bereits erste Leitlinien gezogen worden. Die wissenschaftlichen Gemeinschaften – so der Tenor – werden den Prozess aktiv gestalten können, wenn sie sich über weitere notwendige Kriterien für Datenqualität verständigen und diese darüber hinaus transparent machen, um sowohl das Vertrauen u.a. der Öffentlichkeit gegenüber wissenschaftlichen Daten (wieder-) zu gewinnen, als auch die Reputation der Ergebnisse zu erhöhen. Innerwissenschaftlich sei eine solche Transparenz ohnehin geboten, um die Nachvollziehbarkeit der Resultate und Nutzbarkeit der Daten sicherzustellen. Dazu gehöre auch, die Daten in ihren Erhebungskontext einzuordnen, was eine frühzeitige Dokumentation bereits ab den ersten Entwicklungsschritten erfordere. Allein dies setze in vielen Disziplinen eine veränderte Selbstreflexion bis hin zu einem expliziten „Kulturwandel“ voraus, wie MARTINA BROCKMEIER betonte. Hinzu komme, dass sich mit der Nutzung der Daten über die Fachgrenzen hinaus neue Anforderungen an die hierbei zu verwendende Sprache stellten. Die Definition einer gemeinsamen Sprache bzw. Ontologie werde daher als eines der dringendsten Desiderate erachtet. Der damit verbundene hohe personelle – sowohl quantitativ als auch im Hinblick auf Fachkompetenz und Verantwortungsbewusstsein – und technische Aufwand sollte als Investition in die Zukunft betrachtet werden, die von einer umfassenden Neubewertung von Tätigkeiten in der Forschung und ihrer Förderung flankiert werde – einschließlich der hierauf bezogenen Reputationssysteme – so PETRA GEHRING. Es sei zu überprüfen, ob neben der Bereitstellung der für das Gelingen dieser Transformation notwendigen finanziellen Ressourcen weitere Anreize gesetzt werden müssten. So regte beispielsweise RALF WEHRSPORN an, die qualitativ hochwertige Kuratierung von Datensätzen mit monetären *Incentives* (äquivalent andere symbolische Anreize) zu fördern, die sich an der faktischen Nachfrage orientierten. Dem liege die These zugrunde, dass besonders gut kuratierte Datensätze bevorzugt von Dritten nachgefragt und häufiger abgerufen würden als qualitativ schlechter eingeschätzte Datensätze. Jeder wissenschaftlich (oder auch wirtschaftlich) einschlägige Abruf bzw. Download könne dann mit einer spezifischen Auszahlung verbunden werden, die letztlich auch reputationssteigernde Effekte für die Kuratoren und für die Daten anbietende Einrichtung nach sich ziehen könnte.

Insgesamt waren sich die Tagungsteilnehmer einig, dass hohe Qualitätsansprüche an Daten und Datenmanagement dem Eigeninteresse der Wissenschaft entsprechen. Daher sollte die Sicherstellung

von Datenqualität und der hierfür notwendigen Ressourcen für alle Beteiligten vor Ort ein selbstverständliches Anliegen sein. Keinesfalls sollten entsprechende Gestaltungsfragen ausschließlich an Expertenzirkel – sprich: in die Nationale Forschungsdateninfrastruktur (NFDI) oder die European Open Science Cloud (EOSC) delegiert werden.

Session 1

Gerade im Rahmen der besonders datenintensiven Forschung, z. B. im Rahmen wissenschaftlicher Simulationen und Modellbildungen, stellen sich große Herausforderungen an die Qualität von Daten. So machte einleitend WOLFGANG MARQUARDT darauf aufmerksam, dass sich der eigentliche Informationsgehalt einer wissenschaftlichen Aussage nicht per se aus der Größe der Datenbasis ergebe, sondern aus deren Qualität, die durchaus auch mit der zugrunde gelegten Methode bzw. den Methoden in Verbindung zu setzen sei. Er betrachtete in diesem Zusammenhang Co-Designs von Studien als den idealen Weg, wobei diese wegen des entsprechenden Mehraufwands von den zur Verfügung stehenden Ressourcen abhängig seien. Dies griff BODO ROSENHAHN mit der Feststellung auf, dass eine erste Fehlerquelle bereits in der Datenauswahl bestehen könne, d.h. der Definition und Dokumentation der Auswahlkriterien und der hierfür verwendeten Metrik. Zusätzliche Relevanzgewinne dies, sobald auf externe bzw. bereits vorliegende Daten zurückgegriffen werde. Hier könne es zu zusätzlichen Verzerrungen („Datendrift“) kommen, wenn der ursprüngliche Forschungskontext nicht angemessen berücksichtigt werde bzw. wegen fehlender Metadaten nicht berücksichtigt werden könne. Aber auch im Zuge der Datenverarbeitung seien unabhängig von der Datengröße regelmäßige Überprüfungen – mit Dokumentation – notwendig, ob sich unverhofft Datenabweichungen oder auch (Schein-) Korrelationen ergeben, die die Validität und Aussagekraft der Ergebnisse oder auch ihre Reproduzierbarkeit beeinträchtigten. Einflussfaktoren seien hier beispielsweise Datenquellen und ihre jeweilige Bewertung, z.B. im Hinblick auf zeitliches und räumliches Auflösungsvermögen, wie PETER PELZ am Beispiel der Modellierung von Stadtentwicklung im Globalen Süden zeigen konnte. Ähnliches ist auch bei maschinellem Lernen und im Zusammenhang mit künstlicher Intelligenz zu beachten, da sich Sachverhalte, die in der Codierung fehlerhaft angelegt wurden – z.B. geschlechts- oder gruppenbezogener „Bias“ in der Programmierung – dabei ohne Kontrolle und Datenkuratierung im weiteren Verlauf potenzieren und „verselbstständigen“ könnten. In der Klimaforschung, so berichtete GABRIELE GRAMELSBERGER aufgezeigt, finde seit langem eine Selbstevaluierung durch die weltweit vernetzte Community statt. Dabei sei nicht nur die Themenstellung von evident globalem Interesse. Auch die Erarbeitung der Forschungsdaten finde bereits seit Jahrzehnten durch ein gezielt ausgehandeltes Data Sharing statt, wie sich anhand der International Panel on Climate Change-Reports (IPCC) unter dem Dach der United Nations (UN) darlegen lasse. Hier würden die teilnehmenden Klimamodelle (ca. 20 bis 40 ausgesuchte Modelle) erst nach aufwendiger Vorabprüfung aufgenommen und in bestimmten zeitlichen Abfolgen jeweils einer Vielzahl von weiteren Prüfungen unterzogen. Auf diese Weise erreichten die veröffentlichten Daten eine Qualität, die international sowohl in der Wissenschaft als auch in der Politik und in der Öffentlichkeit für Vertrauen und Wertschätzung gesorgt habe. Durch diese übergreifende Beteiligung von Wissenschaftlerinnen und Wissenschaftler an der Datenbereitstellung, -überprüfung und -veröffentlichung über Institutionen-, Länder- und Fachgrenzen hinweg seien Klimamodelle als Best Practice-Beispiele für Community-Modelle zu verstehen.

Was hier in einer globalen wissenschaftlichen Community funktioniert, sei freilich in anderen Fällen, in denen noch mehr disziplinäre Kulturen eingebunden werden müssten, schwieriger zu realisieren – umso mehr, wenn sich für eine erfolgreiche Kooperation neue Kommunikationsstrukturen erst

etablieren müssten, die weit über die Verständigung auf einen gemeinsamen Sprachgebrauch hinausgehen. Dies ließ sich auch der *Lecture* von ROLF-DIETER HEUER entnehmen, der einen Einblick in die avancierte Entwicklung des Datenqualitätsmanagements am CERN gab, die ebenfalls von einer globalen wissenschaftlichen Community getragen und verantwortet wird. Als vorteilhaft würden sich hier allerdings die spezifischen Voraussetzungen der Großforschung erweisen, in der sich die Qualitätsinteressen der globalen Community auf eine überschaubare Zahl an Experimentaleinrichtungen von Weltgeltung und deren Weiterentwicklung fokussieren lassen. Nichtsdestotrotz war und ist die (Weiter-) Entwicklung der Datenqualitätskultur am CERN eine enorm fordernde Managementaufgabe, die auch die Bedingungen der Archivierung von Forschungsdaten, Versuchsanordnungen und der jeweils verwendeten technischen Parameter von Hard- und Software stets im Blick haben müsse. Heuer machte unter anderem darauf aufmerksam, dass ein solcher Prozess auch Zeit brauche: Datenintensive Institutionen wie das CERN hätten ihr beispielgebendes Datenqualitätsmanagement über Jahrzehnte aufgebaut und verbessert. Andere Einrichtungen – auch die aufzubauende NFDI – bräuchten das Rad nicht neu zu erfinden und könnten beim CERN gute Anhaltspunkte finden. Gleichwohl müsse die Wissenschaftspolitik hier Zeit für Lerneffekte und die richtige Kalibrierung von dezentraler und zentral gesteuerter Entwicklung einplanen.

Session 2

Die Beiträge der zweiten Session widmeten sich der Frage, welche Chancen und Herausforderungen mit der Bereitstellung und Nutzung von Daten im Internet für die Forschung in den Sozial- und Wirtschaftswissenschaften verbunden sind. Hierzu legten FRAUKE KREUTER und JULIA INGRID LANE aufeinander aufbauend dar, welche vielfältigen Möglichkeiten und methodischen Herausforderungen sich für sozialwissenschaftliche Forschungsfragen aus der Flut von Datenmaterial im Internet und durch Mobilfunkdaten ergeben. Auch hier sei festzuhalten, dass die ubiquitäre Verfügbarkeit von Daten (z.B. durch die stetige quantitative und qualitative Weiterentwicklung der Sensorik in Mobiltelefonen) ohne entsprechende Reflexion der wissenschaftlichen Fragestellung sowie der hierauf bezogenen Auswahlkriterien und Messmethoden keinen wissenschaftlichen Nutzwert entwickeln könnten. Vielmehr sei die Transparenz und hiermit einhergehend die Explizierbarkeit des Generierungsprozesses von Daten einschließlich der Hardware, mittels derer sie erzeugt und prozessiert werden, gerade auch im Zuge von KI und maschinellem Lernen von essentieller Bedeutung. Ohne entsprechende Reflexion und Dokumentation von Datenquellen und ihren Einfluss auf die Datenqualität könnten Ergebnisse keine Validität beanspruchen und blieben nicht reproduzierbar. Um die entsprechenden Wechselwirkungen kritisch zu überprüfen, müssen in den Sozialwissenschaften neue Messinstrumente und neue, auch automatisierte Methoden entwickelt und angewendet werden. Auch seien neue Formen der Dokumentation erforderlich, die in der Gesamtheit oftmals nur in Kooperationen und Teams mit unterschiedlichen Kompetenzen realisiert werden könnten.

Ein Best Practice-Beispiel sahen Kreuter und Lane in der von ihnen mitgetragenen Coleridge-Initiative, die sich die Verbesserung, Verknüpfung und Qualitätssicherung sozialwissenschaftlicher Datensätze zum Ziel gemacht habe. Die Initiative bringe Datenproduzenten – insbesondere administrative Einrichtungen in den USA – und Forschende zusammen und biete einen sicheren Raum, in dem amtliche Daten datenschutzrechtlich unbedenklich und qualitätsgesichert zur Lösung von sozialwissenschaftlich und gesellschaftlich relevanten Fragestellungen verknüpft werden könnten. Zugleich werde ein Metadatenystem entwickelt, mit dem einzelne Datensätze verglichen werden könnten – auch hinsichtlich der Studien, die mit ihnen schon durchgeführt und der Forschungsergebnisse, die mit ihnen bereits erzielt worden seien.

Die exponierte Bedeutung einer gut definierten und dokumentierten Fragestellung in den Sozialwissenschaften wurde auch von JANA DIESNER aufgegriffen. Sie wies auf die Einflüsse von unbewussten Verzerrungen („bias“) gerade in Untersuchungen zu sozialen Interaktionen hin, die großen Einfluss auf die Bewertung der erhobenen Daten ausübten. In diesem Zusammenhang könne bereits der bewusste oder unbewusste Sprachgebrauch über den empirischen oder normativen Gehalt einer Fragestellung entscheiden und ihren Qualitätsanspruch verändern. Da Sprache außerdem immer mehrdeutig sei und selbst Namen nicht eindeutig zuzuordnen seien, bedürfe die entsprechende Codierung gezielter Aufmerksamkeit und gegebenenfalls neuer Herangehensweisen, z.B. mit Hilfe des „natural language processing“ (NLP), um bessere Datenqualität zu gewährleisten und dauerhaft zuverlässige Ergebnisse anbieten zu können.

Session 3

Nicht nur in der Wissenschaft, auch in Wirtschaftsunternehmen ist das Bewusstsein für die Bedeutung von qualitativ hochwertigen Daten gestiegen. In diesem Umfeld werde Datenqualität zumeist mit dem Ziel der „fitness for use“ definiert, berichtete CHRISTINE LEGNER. Dies führe allerdings zu Schwierigkeiten, wenn z.B. eine Datenbank aktualisiert werde oder sich der Verwendungszusammenhang ändere. Daher habe sich gerade die Wiederverwendbarkeit als ein entscheidender Faktor bei der Analyse des Kosten- und Nutzenpotentials von Data-Governance-Programmen und Datenqualitätskonzepten herausgestellt. Allerdings stelle sich sowohl innerbetrieblich als auch bei der Verknüpfung von Daten mit Dritten in der Wirtschaft eine neue Herausforderung in Bezug auf die Verteilung von Datenverantwortung und Datensouveränität („ownership“ vs. „stewardship“). Daher hätten Forschung und Unternehmen zusammen Standardrichtlinien entwickelt, die auch den Schutz von Unternehmensdaten und Urheberrechten berücksichtigten sowie im Optimalfall ein Modell für die im Bereich Datenqualität erforderlichen Kompetenzen umfassten. Hierbei werde klargestellt, dass es sich nicht um ein Produkt handele, das sich als abgeschlossenes Konstrukt verkaufen ließe, sondern um einen Prozess, der dauerhaft sowohl technische Aktualisierung, als auch personelle Begleitung erfordere. Dementsprechend lasse sich zusammenfassen, dass sowohl in den Wirtschaftswissenschaften als auch in der Industrie der bereits erwähnte Kulturwandel sowie eine gemeinsame Ontologie und der Ausbau von Kompetenzen im Bereich Datenqualität dringend vonnöten seien, damit Standards übergreifend etabliert werden können.

Über den Aufbau einer nachhaltigen Dateninfrastruktur für die Lebenswissenschaften auf supranationaler Ebene berichtete NIKLAS BLOMBERG am Beispiel von ELIXIR. An diesem Netzwerk sind über 20 europäische Staaten und eine Vielzahl von Institutionen beteiligt. Die über ELIXIR via Open Access bereitgestellten Daten würden sowohl von der Forschung als auch von Wirtschaftsunternehmen – vor allem aus der pharmazeutischen Industrie – intensiv nachgefragt. So arbeiteten Wissenschaftlerinnen und Wissenschaftler z. B. grenzüberschreitend auf dieser Datenbasis an der Genomforschung. Die Notwendigkeit und nicht zuletzt auch der ökonomische Wert dieser Infrastruktur habe bereits Auswirkungen auf die Forschungspolitik in zahlreichen Staaten gezeigt. Grundlage für die Nachfrage und Nutzbarkeit der bereitgestellten Daten seien auch bei ELIXIR die in den wissenschaftlichen Fachgemeinschaften und Forschungs-Communities erarbeiteten Standards sowie erstklassige Metadaten. Blomberg sah hierin eine Best-Practice, die prinzipiell auch auf andere wissenschaftliche Bereiche übertragbar sei. Mit Blick auf die europäische Forschung und im Zusammenhang mit der EOSC könne sich ein funktionierendes Datenqualitätsnetzwerk für die Wissenschaften zu einem relevanten globalen Wettbewerbsvorteil entwickeln. Gleichwohl sei zu beachten, dass die hohe wechselseitige Vernetzung von öffentlich geförderten wissenschaftlichen Einrichtungen und Industrie im Rahmen von ELIXIR auch dem hohen ökonomischen Mehrwert qualitativ hochwertiger lebenswissenschaftlicher Daten für die pharmazeutischen Unternehmen geschuldet sei.

Zusammenfassung und Ausblick

Im Ergebnis verdeutlichte die Konferenz, dass Datenqualität heute eine große Herausforderung angesichts der sich schnell ändernden digitalen Rahmenbedingungen ist. Anders als in der Vergangenheit geht die Thematik dabei weit über eine Diskussion anzuwendender Methoden in der WISSENSCHAFT hinaus. Ganz im Gegenteil zeigt sich, dass Datenqualität eine Herausforderung von hoher gesamtgesellschaftlicher Relevanz ist. Dies ist nicht zuletzt in den stark anwendungsbezogenen Bereichen wissenschaftlicher Forschung festzustellen. Auch die Wirtschaft selbst sucht hier nach Lösungen – und nach wissenschaftlicher Unterstützung – wie die Beispiele „Corporate Data“ und „ELIXIR“ zeigen konnten. In ihrer Zusammenfassung der Konferenzerträge identifizierte PETRA GEHRING „erstaunlich viele Wahlverwandtschaften“ und wiederkehrende Bedarfsmuster in den Beiträgen, die auf eine Beschreibungsarbeit „neuen Typs“ hinauslaufe, die sich sowohl im Dialog der Disziplinen als auch im Dialog von Wissenschaft und Gesellschaft entwickeln müsse.

Für die Messung von Datenqualität wurde „fit for purpose“ als ein erster Ansatz identifiziert, der aber keineswegs ausreichend sei, da „purpose“ nur in Relation zu Erhebungszeit und -kontext nachvollzogen werden könne und die erzielten Ergebnisse für sich genommen ohne neuerliche Überprüfbarkeit ihre Validität verlieren würden. So sei Datenqualität nie vollkommen authentisch, sondern immer ein jeweils relatives und zeitgebundenes Produkt, das unter der Sonde neuer inhaltlicher Fragestellungen, neuer Forschungsmethoden und weiterentwickelter Technologien und Infrastrukturen jeweils neu bewertet werden müsse. Die Unabschließbarkeit wissenschaftlicher Forschung werde in der Qualitätsdimension durch die rasante technologische Entwicklung im Bereich der Soft- und Hardware noch unterstrichen.

Dies zeigte sich auch in der von JAN-MARTIN WIARDA moderierten Paneldiskussion mit LARS BERNARD, MARTINA BROCKMEIER, FRAUKE KREUTER, KLAUS POHL und KATRIN STUMP. Gemeinsam wurde sowohl der wachsende Bedarf an wissenschaftlich und informationstechnisch qualifiziertem Personal als auch die steigende Notwendigkeit von Kooperationen im Wissenschaftssystem betont. Gewinner und Verlierer im digitalen Transformationsprozess der Wissenschaft seien zum gegenwärtigen Zeitpunkt nicht konkret auszumachen. Zu einer neuen Qualitätskultur gehörten auch die weniger datenintensiven Disziplinen und Forschungsinfrastrukturen, deren „analoge“ Artefakte und Expertise durch Digitalisate und digitale Methoden nicht ersetzt, sondern ergänzt würden. Zu dem proklamierten Kulturwandel in Sachen wissenschaftlicher Datenqualität gehöre auch ein neues wechselseitiges Verständnis von Forschung und Infrastrukturbereich. Das hartnäckige Klischee des Bibliothekars mit Zettelkasten und Rollwagen sei z.B. schon lange überholt, so Katrin Stump. Vielmehr gehe es heute darum, in digitalen Forschungsumgebungen Kommunikation auf Augenhöhe zu etablieren sowie Hilfestellungen und Dienstleistungen für die Wissenschaft auf einem neuen Niveau anzubieten – von der Literaturversorgung im Post-Subskriptionszeitalter bis hin zu Aufgaben des Managements und der Archivierung im Rahmen wissenschaftlicher Datenlebenszyklen.

Von der Nationalen Forschungsdateninfrastruktur (NFDI) erwarteten sich alle Experten auf dem Podium einen großen Schub für die Weiterentwicklung der wissenschaftlichen Datenqualität in Deutschland und Europa. Hier gelte es, im Schwung der nationalen Aufbauarbeit die europäische Dimension nicht aus den Augen zu verlieren und insbesondere im Rahmen der European Open Science Cloud zu einer klugen Verzahnung von Dateninfrastrukturen im globalen Maßstab beizutragen.