

Konzepte für das Forschungsdatenmanagement an der Universität Rostock – Extended Abstract

Concepts for the Management of Research Data at the University of Rostock (Extended Abstract)

Ilvio Bruder, Andreas Heuer, Sebastian Schick und Sascha Spors

Universität Rostock

Abstract: Research Data Management aims at gathering, capturing, storing, tracking, and archiving all the data being produced in scientific projects and experiments. Besides these data, all the processing steps on these data - eventually resulting in scientific publications - have to be stored as well.

Research Data Management is not only a scientific discipline in Computer Science. Universities and Research Institutes have to provide organizational structures and processes and pragmatic solutions (hardware and software resources) to implement first, simple tasks of Research Data Management.

In this paper, we sketch the organizational, pragmatic, and research aspects of Research Data Management from a local (University of Rostock) point of view. At the University of Rostock, we have wider experiences with research data management in marine biology and medical research. The research aspects are part of modern database research topics such as temporal databases, data integration, schema evolution, and provenance management.

1 Einleitung

Ziel des Forschungsdatenmanagement ist es, Daten, die bei wissenschaftlichen Projekten, Experimenten oder Beobachtungen erfasst werden, zu verfolgen und zu archivieren. Zusätzlich muss die weitere Verarbeitung der Daten bis hin zur Publikation gespeichert und nachvollziehbar gemacht werden. Viele Konferenzen und Workshops widmen sich unter anderem genau diesem Thema und Forschungsgeldgeber erwarten bei Ausschreibungen fast immer Konzepte bzgl. Nachhaltigkeit, Nachvollziehbarkeit und Offenlegung von Forschungsdaten. Daher müssen an Universitäten verstärkt organisatorische Konzepte für das Forschungsdatenmanagement entwickelt werden sowie zeitnah pragmatische Lösungen für eine erste Umsetzung des Forschungsdatenmanagements implementiert werden. In diesem Beitrag soll ein Überblick über die organisatorischen Maßnahmen und die pragmatischen Lösungen an der Universität Rostock gegeben werden.

Darüberhinaus ist Forschungsdatenmanagement ein aktuelles Forschungsthema, bei dem insbesondere moderne Datenbanktechnologien benötigt werden. In diesem Beitrag werden wir einige Forschungsansätze skizzieren, mit denen wir

uns speziell in der Datenbankforschungsgruppe der Universität Rostock befassen. Insbesondere werden Forschungsergebnisse aus den Bereichen temporale Datenbanken, Datenintegration, Schemaevolution sowie Provenance-Management erwähnt.

An der Universität Rostock haben wir speziell Erfahrungen im Bereich von Informatik- [8] und Elektrotechnik-Forschungsgebieten [18] und im Bereich meeresbiologischer Forschungsdaten [2], mittlerweile aber auch verstärkt in anderen naturwissenschaftlichen und medizinischen [19] Forschungsbereichen. Ein ganz anderer Bereich sind die weniger (mess- und sensor-)datengetriebenen geisteswissenschaftlichen Forschungsgebiete (Digital Humanities), in denen sehr dokumentenzentriert gearbeitet wird [15]. Im Folgenden werden wir uns eher mit den datengetriebenen Szenarien befassen.

2 Herausforderungen des Forschungsdatenmanagements

Es gibt im Forschungsdatenmanagement eine Reihe von Herausforderungen, die zum einen eher organisatorisch bzw. praktisch orientiert sind. Zum anderen gibt es auch interessante Herausforderungen in der Grundlagenforschung, gerade in Bezug auf Datenbanktechnologien. Die Herausforderungen in der Datenbank-Grundlagenforschung sind

- die Heterogenität der Daten,
- ein nicht oder nur unvollständig vorhandenes Schema,
- die Sicherung der Provenance (Herkunft) der Forschungsergebnisse sowie die Reproduzierbarkeit der wissenschaftlichen Auswertungen,
- die Spezifikation und Nachverfolgung von wissenschaftlichen Arbeitsabläufen (Scientific Workflows), sowohl organisatorisch als auch datentechnisch,
- die Einbettung und Speicherung von anwendungsspezifischen Funktionen und Methoden, insbesondere zur Analyse von Daten,
- temporale Aspekte zur Reproduzierbarkeit von Auswertungen über Messdaten, die als Stromdaten ständig produziert werden
- sowie die Komplexität in den Auswertungen und Veränderungen in den Auswertungsroutinen über einen längeren Zeitraum hinweg.

Eher praktische Herausforderungen, die zeitnah mit pragmatischen und bereits vorhandenen Hardware- und Software-Lösungen implementiert werden müssen, sind

- Open Science, inklusive Zugriff auf und Präsentation von Daten und Ergebnissen,
- die Umsetzbarkeit von Lösungen in der Praxis, etwa durch flexible Architekturen,
- die Nachhaltigkeit der implementierten Lösungen,
- Usability bzw. Ergonomie des Softwaresystems
- sowie Lizenz- und Rechtsfragen bei benutzten Originaldaten und Softwarewerkzeugen zur Auswertung und Darstellung der Daten.

Andere Ansätze unterstützen auch kollaboratives Arbeiten und stellen die Erfassung, Verwaltung und Nutzung von Forschungs-Metadaten in den Vordergrund (wie LabBook: [11]). Weiterhin können Forschungsdaten und (datenbankgestützte) Auswertungen auch in das zu publizierende Dokument integriert werden wie in Janiform mit den Portable DataBase Files (PDbF) [5].

Einige der oben genannten Herausforderungen werden wir im Folgenden genauer diskutieren.

3 Forschungsdaten-Lifecycle

In Forschungseinrichtungen wird man mit einer Vielzahl an Experimenten und Auswertungen aus sehr unterschiedlichen, interdisziplinären Forschungsbereichen konfrontiert. Die Daten, Methodik und verwendeten Werkzeuge sind sehr heterogen. Für die Zusammenarbeit müssen Austauschbarkeit und Interoperabilität gewährleistet werden. Hierzu gibt es verschiedene Modellierungsansätze, die entweder abstrakt und damit allgemeingültiger oder eher spezialisiert und damit komplexer sind. Des Weiteren gibt es unterschiedliche Daten und Ziele der Verarbeitung entlang des Lifecycles eines Experiments. Abbildung 1 veranschaulicht einen solchen Lifecycle von der ersten Erfassung der Daten als Primärforschungsdaten über die weitere Verarbeitung bis hin zur Publikation. Dabei ist für die Nachvollziehbarkeit der Forschungsergebnisse die gesamte Kette der Datenverarbeitung von der Erfassung bis zur Publikation aufzunehmen und zu archivieren (grüner Rahmen).

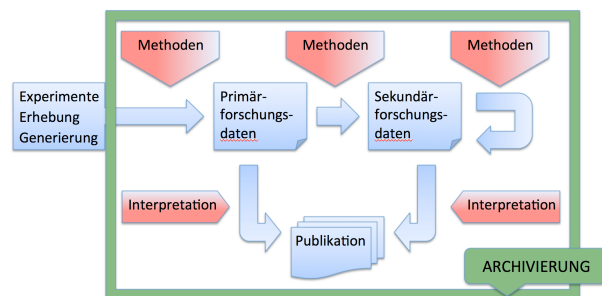


Abb. 1. Lifecycle von Forschungsdaten aus [2]

Forschungsdaten- und Publikations-Workflows sind in der Wissenschaft eng verzahnt. Dabei gibt es viele mögliche Abläufe, die beliebige Zusammenhänge zwischen Experimenten, Primärforschungsdaten, Sekundärforschungsdaten sowie Publikationen abbilden. Hier haben sich Scientific-Workflow-Systeme wie z.B. Kepler [13] etabliert. Publikationsprozesse sind Bestandteil des Gebiets der Digitalen Bibliotheken (siehe z.B. [21]).

Grundlegendes Prinzip, das im Rahmen dieser wissenschaftlichen Arbeitsabläufe realisiert werden muss, ist das FAIR-Prinzip aus [20]. FAIR beinhaltet,

dass Daten auffindbar (findable), zugreifbar (accessible), interoperabel und wiederverwendbar (reusable) sein müssen. Wir werden nun zunächst einige organisatorische Maßnahmen sowie pragmatische Lösungen beschreiben, die dieses Prinzip fördern sollen.

4 Forschungsdatenmanagement an der Universität Rostock

An der Universität Rostock wurden und werden verschiedene organisatorische Konzepte sowie grundlegende als auch praxisnahe Lösungen entwickelt. Einige sollen im Folgenden kurz vorgestellt werden.

Rostocker Modell

Das „Rostocker Modell“ der Universität Rostock [9] regelt die Zuständigkeit und generelle Verfahrensweise bei der Verstetigung und Nachhaltigkeit von Spezialarchiven, die im Rahmen von Forschungsprojekten an verschiedenen Fakultäten der Universität Rostock entstehen oder entstanden sind.

Derartige Spezialarchive zeichnen sich dadurch aus, dass sie nicht nur die Daten selbst speichern, sondern auch (Text-, Bild-, Multimedia-)Dokumente sowie archiv-spezifische Funktionen und Methoden. Solche Funktionen und Methoden gehen über das hinaus, was Standard-Software und Standard-DBMS bieten. Das bedeutet, dass man solche Archive nicht einfach durch herkömmliche Software-Maintenance warten kann, sondern auch die spezielle Funktionalität aktuell halten muss. Die Rostocker Lösung ist eine Arbeitsteilung zwischen dem Universitätsrechenzentrum (ITMZ: IT- und Medienzentrum), der Universitätsbibliothek (UB) sowie dem Institut für Informatik, die in zugrundeliegenden Projekten oftmals beteiligt sind. Darüberhinaus wird mit regionalen Softwarefirmen die Wartung unterstützt. Hierdurch ist es möglich, auch längere Zeit nach Beendigung des Projektes eine zumindest grundlegende Funktionalität des Spezialarchivs eines abgeschlossenen Software-Projektes innerhalb und außerhalb der Universität Rostock verfügbar zu halten.

Beispielsweise wurden in einer aktuellen Arbeit in 2017 Forschungsergebnisse des Projekts eNoteHistory [1] auf einen aktuellen Softwarestand gebracht. Dieses Projekt war bereits vor 10 Jahren abgeschlossen worden. Im Projekt wurden Notenhandschriften digitalisiert, gespeichert und so aufbereitet, dass ein Auffinden nach Merkmalen des Schreibers der Noten möglich war. Die Software musste dabei von einer IBM-DB2-Lösung (in 2004 noch mit den damals vorhandenen Text und Image Extendern entwickelt) auf eine aktuelle PostgreSQL-Version umgestellt werden. Dabei wurden auch viele Erfahrungen gesammelt, was in der Zukunft bei analogen Projekten konzeptionell zu überdenken ist, wenn keine kontinuierliche Evolution der Soft- und Hardware nach Projektende möglich ist.

Forschungsdatenworkflow und Publikationsworkflow an der Universität Rostock

Die Universität Rostock unterstützt aktiv Open-Access-Veröffentlichungen und hat u.a. die “Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities” [12] unterzeichnet. Außerdem wurde ein Publikationsfonds für Open-Access-Veröffentlichungen unter Beteiligung der Universitätsbibliothek und Vertretern aller Fakultäten aufgebaut. Weiterhin befinden sich an der Universitätsbibliothek Dienste im Bereich des Forschungsdatenmanagements im Aufbau. Dazu zählen Services zur Strategie und Wahl des Datenmanagements und der Datenstandards, zur Repräsentation des Wissens sowie bei der digitalen Archivierung von Daten.

Für die Publikationen, deren Langzeitarchivierung sowie deren stabilen Zitierbarkeit wird an der Universität Rostock auf entsprechende Systeme, die die eindeutige Referenzierbarkeit per Digital Object Identifiers (DOI) realisieren, zurückgegriffen. Basis ist das System “RosDok” (<http://rosdok.uni-rostock.de>), das Dokumente in einer Digitalen Bibliothek organisiert, speichert und zugreifbar hält. Erweitert wird das System, um neben den Publikationen selbst auch Daten und Referenzen für die Forschungsdaten und -methoden anzubieten. RosDok basiert auf dem MyCoRe-System [17]. Die Universität Rostock plant das Forschungsdatenmanagement zentral zu koordinieren und die vorhandenen Dienstleistungen kontinuierlich zu erweitern.

Open Science und pragmatische Lösungen für die kurzfristige Umsetzung des Forschungsdatenmanagements

Open Science ist eine Forderung an die Wissenschaft aus dem gesellschaftlichen Kontext heraus: wissenschaftliche Ergebnisse, von öffentlich-rechtlichen Einrichtungen erzielt oder aus öffentlichen Mitteln finanziert, sollten auch der Öffentlichkeit zur Verfügung stehen. Open Science wird daher gerade von staatlichen Forschungsförderern als Voraussetzung für gute wissenschaftliche Praxis angesehen.

Zur Open Science gehören eine Reihe von Begrifflichkeiten und Konzepten, um die Wissenschaft transparent und nachvollziehbar für jedermann anzubieten (siehe auch [18]):

- Open Source – freie Verfügbarkeit des Sourcecodes und der im Forschungsprojekt verwendeten Werkzeuge
- Open (Science) Data – freie Verfügbarkeit der wissenschaftlichen Daten. Dazu gehören auch Protokolle, Beschreibungen, Kalibrierdaten, etc.
- Open Access – freier Zugriff auf die veröffentlichten Ergebnisse, wie Forschungsartikel
- Open Methodology – Beschreibung aller verwendeten Methoden, die für das Forschungsergebnis genutzt und entwickelt wurden
- Open Notebook Science – freie Verfügbarkeit sämtlicher Basisdaten, Aufzeichnungen, Planungs- und Einsatzinformationen zu einem Forschungsprojekt

- Open Educational Resources – freier Zugang zu Lehr- und Lernmaterialien bzgl. der Forschung
- Open Peer Review – Transparenz und öffentlicher Zugang zum Reviewing-Prozess im wissenschaftlichen Publikations- bzw. Antragsprozess

Für die Realisierung von Open (Science) Data werden in einigen Forschungsgruppen der Universität Rostock derzeit zunächst für das interne Management der Daten Versionskontrollwerkzeuge wie SVN and Git benutzt. Veröffentlicht werden die Daten dann über GitHub. Nachteilig bei diesen Werkzeugen ist insbesondere die Versionsverfolgung bei einer sehr großen Anzahl von Forschungsdaten (und -dateien). Ein positiver Seiteneffekt bei der öffentlichen Bereitstellung ist dagegen die Notwendigkeit für eine saubere Dokumentation (die sonst oft unterbleibt), eine Bereinigung von Daten und Ergebnissen, und eine ausführliche gruppeninterne Diskussion über Originaldaten und Ergebnisse.

Gerade die angesprochenen Probleme bei der Versionsverfolgung bei Forschungsergebnissen, die über einen längeren Zeitraum, etwa auch permanent, gewonnen und ausgewertet werden führten zu einigen grundlegenden Forschungsarbeiten, die derzeit in der Datenbank-Forschungsgruppe der Universität Rostock vorangetrieben werden.

5 Forschungsfragestellungen im Forschungsdatenmanagement

Im Folgenden wollen wir einige Fragestellungen des Forschungsdatenmanagements mit Bezug zu Datenbanktechniken genauer vorstellen. Im Gegensatz zu [2] werden wir uns in diesem Artikel nicht nur auf die meeresbiologischen Szenarien eines unserer Aninstitute konzentrieren, sondern auch andere datengetriebene Szenarien und die dort auftretenden Fragestellungen berücksichtigen. Andererseits werden wir uns im Gegensatz zu [2] eher auf die Teilbereiche des Provenance Management und die Reproduzierbarkeit von Forschungsergebnissen konzentrieren und die gesamte Integrationspipeline nur überblicksartig vorstellen.

In der Systematik von Abbildung 1 werden wir uns dabei um die Auswertung von Primär- und Sekundärforschungsdaten in langfristigen Messreihen kümmern. Dabei ist die Reproduzierbarkeit dieser Auswertungsergebnisse für die Nachprüfbarkeit von Publikationsergebnissen in naturwissenschaftlicher Forschung wichtig. Das Provenance Management wird aber auch benötigt, um bei der Erforschung und Entwicklung von smarten Systemen (Assistenzsystemen) beurteilen zu können, warum die Situations- und Aktivitätserkennung in Assistenzsystemen in bestimmten Fällen fehlerhafte Ergebnisse geliefert hat. Sowohl bei langfristigen Messreihen als auch bei der Entwicklung von smarten Systemen sind daneben auch noch temporale Aspekte wichtig, da über die Zeit sowohl erfasste Daten als auch Auswertungsmethoden sich verändern können.

Provenance Management und Reproduzierbarkeit von Forschungsergebnissen

Bei der Reproduzierbarkeit von Forschungsergebnissen gibt es verschiedene Stufen. Man kann testen

- ob ein Ergebnis plausibel ist: hier ist die Fragestellung, ob eine Aussage in einer Publikation zu den gespeicherten Forschungsdaten passt; üblicherweise eine manuell durch Gutachter durchzuführende Tätigkeit, die u.a. durch Textanalysen softwaretechnisch zumindest unterstützt werden kann
- ob das Ergebnis nachvollziehbar ist: hier ist die Fragestellung, ob ein etwa tabellarisch aufbereitetes Ergebnis aus strukturierten Daten zu den gespeicherten Forschungsdaten passt; hier greifen schon weitere Analyseverfahren, die auf strukturierte Daten angewendet werden können
- oder ob das Ergebnis reproduzierbar ist: hier muss dasselbe Ergebnis aus strukturierten Daten auf Basis der gespeicherten Primärforschungsdaten mit der gleichen Methodik (etwa einer gespeicherten Analysefunktion) softwaretechnisch *ohne Medienbruch* wieder errechnet werden können.

Wenn man davon ausgeht, dass die Forschungsarbeiten auf Basis von Open Science durchgeführt werden, sollten die Forschungsergebnisse prinzipiell immer von den Primärforschungsdaten her reproduzierbar sein. Der entgegengesetzte Weg, vom Ergebnis zurück zu den Primärforschungsdaten, ist die Rückverfolgbarkeit von Forschungsergebnissen. Das zugrundeliegende Forschungsgebiet im Datenbankbereich ist das Provenance Management.

In der Provenance-Theorie unterscheidet man Where-, Why- und How-Provenance [16,4]. Die Fragestellungen dahinter sind:

- Where – woher kommen die Daten, die zu dem Ergebnis führten? Ergebnis ist die Herkunft der Daten in Form von zugrundeliegenden Datenbanken, Dateiverzeichnissen, Datensammlungen oder Repositorien.
- Why – welche Daten spielten exakt eine Rolle? Ergebnis sind die Einzeldaten (relational: die Tupel in Relationen), die in die Methodik eingeflossen sind.
- How – wie ist es mit diesen Daten zu diesem Ergebnis gekommen? Ergebnis ist eine Darstellung der angewendeten Operationen inklusive der Reihenfolge der Bearbeitungsschritte, die für die Ermittlung des Ergebnisses relevant waren.

Dabei sind Where- und Why-Anfragen auf die Forschungsdaten mit aktuellen Mitteln, wie z.B. einer Datenbankerweiterung der Open-Source-Datenbank PostgreSQL [7], machbar. Leider eignen sich die derzeit vorhandenen Provenance-Techniken nur unter sehr speziellen Randbedingungen, die üblicherweise bei Forschungsprojekten mit komplexen Auswertungsfunktionen nicht gegeben sind:

- Die Auswertungen auf den Daten werden nur mit Selektion, Projektion und Verbund vorgenommen. Nur in speziellen Fällen können Aggregatfunktionen berücksichtigt werden. Analysen auf Forschungsdaten enthalten aber üblicherweise komplexe statistische Funktionen oder gar Machine-Learning-Algorithmen.

- Die Auswertungen auf den Daten werden auf einer fixierten Datenbank vorgenommen: die Forschungsergebnisse berücksichtigen keine Updates (bei Stromdaten: ständig sich ändernde Menge an Primärforschungsdaten) auf den Forschungsdaten. Hier müssen wir die Provenance-Techniken mit temporalen Aspekten (siehe unten) verknüpfen.

Ein grundlegendes Forschungsproblem ist dabei, welcher (minimale) Ausschnitt von den Primärforschungsdaten in welcher Version (unveränderbar) gespeichert werden muss (wir nennen es: eingefroren werden muss), so dass man aus den Ergebnissen der Auswertung und den eingefrorenen Daten die Inverse der Auswertung (Anfrage an die Datenbank) automatisiert berechnen kann. Für unsere Forschungsarbeiten hierzu erweitern wir den Begriff der inversen Schemaabbildungen, der für Datenaustausch und Datenintegration etwa in [6] eingeführt wurde.

Ziele unserer Forschungsarbeiten zum Provenance Management haben wir in [8] bereits im Kontext der Entwicklung von Assistenzsystemen vorgestellt. Die zu entwickelnden Verfahren des Provenance Management werden nun aber auch auf die Rückverfolgbarkeit von Forschungsergebnissen im Rahmen des Forschungsdatenmanagements angewandt [2] und weiterentwickelt.

Temporale Aspekte im Forschungsdatenmanagement

Für den Umgang mit Forschungsdaten sind Auswertungen von historischen Daten als auch die spätere Nachvollziehbarkeit von Forschungsergebnissen ein wesentliches Qualitätsmerkmal für eine Nachhaltigkeit dieser Forschung. Temporale Informationen über die Daten und Experimente sind nötig, um eine solche Nachhaltigkeit zu erreichen.

Das Speichern temporaler Informationen zu den Forschungsdaten sind z.B. für Provenance-Anfragen, historische Vergleiche bzw. für sogenannte Time-Travel-Anfragen essentiell. Time-Travel-Anfragen sind Anfragen auf Datenbestände, die auch zu anderen Zeitpunkten als dem aktuellen gültig waren. Seit dem SQL:2011-Standard [10] sind temporale Aspekte fester Bestandteil des Datenbankstandards. Unser Prototyp zum Forschungsdatenmanagement, der im Rahmen einer Kooperation mit einem Aninstitut der Universität Rostock entwickelt wurde, wurde auf Basis von PostgreSQL implementiert, da dieses System verglichen mit anderen Open-Source-DBMS mehr Voraussetzungen für die Umsetzung von SQL:2011-Konzepten angeboten hat [14].

Da sich neben den Daten auch Funktionen über die Zeit verändern können, müssen auch diese versioniert und mit entsprechenden Zeitintervallen beschrieben werden können. Eine Versionierung von User Defined Functions (UDFs in RDBMS) ist im SQL-Standard noch nicht vorgesehen, so dass gleiche Funktionen (gleiche Namen und Parameter) mit unterschiedlicher zeitlicher Gültigkeit nicht gleichzeitig in der Datenbank registriert sein können. Hierzu bedarf es zur Zeit noch manuell definierter Tabellen zur Verwaltung der zeitlichen Gültigkeit von Funktionen bzw. zusätzliche Werkzeuge zur An- und Abmeldung von UDFs, die gerade für eine spezielle Anfrage benötigt werden.

Integration statt Migration – Integrationspipeline

Bei der Organisation des Forschungsdatenmanagements werden meist Richtlinien (gibt es bei der DFG, BMBF, Hochschulen und Forschungseinrichtungen) oder Handbücher (ein allgemeines auch von der DFG anerkanntes ist [3]) ausgearbeitet, die die Vorgehensweise vorgeben. Aus bisherigen Erfahrungen wird allerdings eine solche Vorgabe nicht unbedingt beachtet und die Durchsetzung gestaltet sich ebenfalls oftmals schwierig. Kernpunkt eines nutzerfreundlichen Forschungsdatenmanagements sollte sein

- den Wissenschaftlern einen Mehrwert mit der Forschungsdatenverwaltung zu bieten,
- ihre Vorgehensweisen und Werkzeuge nicht zu ersetzen, sondern zu integrieren, sowie
- Services anzubieten, neben Beratung auch direkte, technische Unterstützung bei der Organisation und Umsetzung einer konkreten individuellen als auch verknüpften Forschungsdatenverwaltung.

Wissenschaftler stellen sich nicht zwangsweise komplett auf eine andere Art der Datenverwaltung um. Es ist bzgl. der Compliance besser, die speziellen Vorgänge der einzelnen Wissenschaftler in eine Gesamtstrategie zum Forschungsdatenmanagement zu integrieren.

Um die Forschungsdaten aus verschiedenen wissenschaftlichen Projekten zu integrieren sind Transformations- und Evolutionsprozesse der Daten, der Schemata und der Funktionalität vonnöten. Diese Vorgänge ähneln dem ETL-Prozess in Data Warehouses. Wir haben dazu eine Integrationspipeline entwickelt, die über verschiedene Stufen halbautomatisch Schemata extrahieren und integrieren kann [2].

6 Zusammenfassung

In diesem Beitrag wurden drei Aspekte des Forschungsdatenmanagements diskutiert. Es wurden die derzeitigen und geplanten organisatorischen Rahmenbedingungen an der Universität Rostock vorgestellt. Es wurden zunächst pragmatische Lösungen aufgezeigt, um zeitnah Lösungen für die Open-Science-Anforderungen umsetzen zu können. Schließlich wurden Forschungsherausforderungen insbesondere aus dem Bereich der Datenbanktechnologie präsentiert.

Die Autorengruppe arbeitet dabei an verschiedenen Teilaspekten, etwa an den organisatorischen Rahmenbedingungen (Schick; Universitätsbibliothek Rostock, verantwortlich für Digitale Bibliotheken und Forschungsdaten), den pragmatischen Lösungen für die Open Science im Kontext eines startenden Sonderforschungsbereiches (Spors; Institut für Nachrichtentechnik) und an den genannten Forschungsfragestellungen (Bruder, Heuer; Institut für Informatik, Forschungsgruppe Datenbanken).

Literatur

1. Ilvio Bruder, Temenushka Ignatova, and Lars Milewski. Knowledge-Based Scribe Recognition in Historical Music Archives. In Rachel Heery and Liz Lyon, editors, *Research and Advanced Technology for Digital Libraries, 8th European Conference, ECDL'04*, volume 3232 of *Lecture Notes in Computer Science*, pages 304–316. Springer, 2004.
2. Ilvio Bruder, Meike Klettke, Mark Lukas Möller, Frank Meyer, Andreas Heuer, Susanne Jürgensmann, and Susanne Feistel. Daten wie Sand am Meer - Datenerhebung, -strukturierung, -management und Data Provenance für die Ostseeforschung. *Datenbank-Spektrum*, 17(2):183–196, 2017.
3. Stephan Büttner, Hans-Christoph Hobohm, and Lars Müller. *Handbuch Forschungsdatenmanagement*. BOCK+HERCHEN, 2011.
4. James Cheney, Laura Chiticariu, and Wang Chiew Tan. Provenance in databases: Why, how, and where. *Foundations and Trends in Databases*, 1(4):379–474, 2009.
5. Jens Dittrich and Patrick Bender. Janiform intra-document analytics for reproducible research. *PVLDB*, 8(12):1972–1975, 2015.
6. Ronald Fagin, Phokion G. Kolaitis, Lucian Popa, and Wang Chiew Tan. Schema mapping evolution through composition and inversion. In Zohra Bellahsene, Angela Bonifati, and Erhard Rahm, editors, *Schema Matching and Mapping*, Data-Centric Systems and Applications, pages 191–222. Springer, 2011.
7. Boris Glavic and Gustavo Alonso. The PERM provenance management system in action. In *Proc. SIGMOD'09*, 2009.
8. Andreas Heuer. METIS in PARADISE: Provenance Management bei der Auswertung von Sensordatenmengen für die Entwicklung von Assistenzsystemen. In Norbert Ritter, Andreas Henrich, Wolfgang Lehner, Andreas Thor, Steffen Friedrich, and Wolfram Wingerath, editors, *Datenbanksysteme für Business, Technologie und Web (BTW 2015) - Workshopband, 2.-3. März 2015, Hamburg, Germany*, volume 242 of *LNI*, pages 131–136. GI, 2015.
9. Andreas Heuer, Holger Meyer, and Ilvio Bruder. Nachhaltigkeit von digitalen Dokumenten – das Rostocker Modell. *Steinbeis Transfermagazin*, 04/2014, 2014.
10. ISO/IEC 9075-2:2011. Information technology - Database languages - SQL-Part 2: Foundation (SQL/Foundation). Technical report, ISO/IEC JTC 1/SC 32, 2011.
11. Eser Kandogan, Mary Roth, Peter M. Schwarz, Joshua Hui, Ignacio G. Terrizzano, Christina Christodoulakis, and Renée J. Miller. Labbook: Metadata-driven social collaborative data analysis. In *Proc. International Conference on Big Data*, pages 431–440. IEEE, 2015.
12. Max-Planck-Gesellschaft. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003. erreichbar unter <https://openaccess.mpg.de/>.
13. Timothy McPhilips, Shawn Bowers, and Bertram Ludäscher. Collection-oriented scientific workflows for integrating and analyzing biological data. In *Proceedings of the DILS Workshop*, 2006.
14. Frank Meyer. Temporale Aspekte und Provenance-Anfragen im Umfeld des Forschungsdatenmanagements. Master-Arbeit, Universität Rostock, 2016.
15. Holger Meyer, Alf-Christian Schering, and Andreas Heuer. The Hydra.PowerGraph System - Building Digital Archives with Directed and Typed Hypergraphs. *Datenbank-Spektrum*, 17(2):113–129, 2017.
16. Luc Moreau and Paul T. Groth. *Provenance: An Introduction to PROV*. Morgan & Claypool, 2013.

17. Wiebke Oeltjen. Virtuelle Bibliotheken flexibel gestalten. In *eLibrary – den Wandel gestalten, Proceedings of the WissKom'10*, pages 259–266. Schriften des Forschungszentrums Jülich, 2010.
18. Sascha Spors, Matthias Geier, and Hagen Wierstorf. Towards Open Science in Acoustics: Foundations and Best Practices. In *Tagungsband der DAGA'17*, pages 218–221. DEGA, 2017.
19. Georgi Straube, Ilvio Bruder, Dortje Löper, and Andreas Heuer. Data Integration in a Clinical Environment Using the Global-as-Local-View-Extension Technique. In Yanchun Zhang, Guiqing Yao, Jing He, Lei Wang, Neil R. Smalheiser, and Xiao-Xia Yin, editors, *Health Information Science - Third International Conference, HIS 2014, Shenzhen, China, April 22-23, 2014. Proceedings*, volume 8423 of *Lecture Notes in Computer Science*, pages 148–159. Springer, 2014.
20. Mark D. Wilkinson and et.al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 2016.
21. Ian Witten, David Bainbridge, and David Nichols. *How to Build a Digital Library*. Morgan Kaufmann, 2nd edition, 2010.