

Software und Services for Science (S3) – 2nd Conference on Non-Textual Information

80 Fachleute treffen sich zum Thema wissenschaftliche Software in Hannover

Sandra Niemeyer

„Texte, ob in gedruckter oder in elektronischer Form, sind heute nicht mehr die einzige Quelle für Wissen und Information. In Forschung und Lehre nimmt die Bedeutung von nicht-textuellen Materialien wie audiovisuellen Medien, Forschungsdaten und Software stetig zu“, begrüßte Barbara Hartung vom Niedersächsischen Ministerium für Wissenschaft und Kultur (MWK) und Vorsitzende des Stiftungsrates der Technischen Informationsbibliothek (TIB) die 80 Teilnehmenden der 2nd Conference on Non-Textual Information. Zwei Tage sollte das Leibnizhaus in Hannover Ort für Diskussionen und

sam mit ihren Partnern ZB MED – Informationszentrum Lebenswissenschaften und ZBW – Leibniz-Informationszentrum Wirtschaft ausgerichtet wurde, stand unter dem Titel „Software and Services for Science“ das Thema wissenschaftliche Software.

„Auf dieser Konferenz werden wir uns unter anderem mit Nachhaltigkeit und Referenzierung von wissenschaftlicher Software sowie Trends in der Programmierpraxis, aber auch mit rechtlichen Aspekten und dem Teilen von Software beschäftigen“, fasste Irina Sens, Kommissarische Leitung der TIB, die vielfältigen Themen zusammen. „Nutzen Sie die beiden Tage für interessante Diskussionen und intensiven Austausch, ob in den Pausen oder am Abend beim Get-together im Alten Rathaus in Hannover“, forderte sie die Anwesenden auf. Neben den Teilnehmerinnen und Teilnehmern begrüßte Sens zu Beginn der Konferenz noch einen besonderen Gast: Sören Auer, der seit Juli 2017 neuer Direktor der TIB ist. Am zweiten Konferenztag hielt Auer, der zum Zeitpunkt der Konferenz noch an der Rheinischen Friedrich-Wilhelms-Universität Bonn und am Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS tätig war, einen Vortrag zu Big Data. „Wir freuen uns darauf, dass wir an der TIB ab Juli mit Herrn Professor Auer an neuen, innovativen Forschungsthemen wie Data Science, Digital Libraries und Open Knowledge arbeiten werden. Es gibt viele Ideen für Kooperationen, beispielsweise mit dem For-

schungszentrum L3S“, sagte Sens in ihrer Begrüßungsrede.

„Die vergangenen Jahre zeigen ganz deutlich die Entwicklung hin zu nicht-textuellen Materialien wie Videos oder Forschungsdaten in der Wissenschaft“, sagte Wolfgang Nejdl, Direktor des Forschungszentrums L3S, das die TIB bei der Programmplanung der Konferenz unterstützt hatte. So helfe beispielsweise DataCite Forschenden seit Jahren mit der Vergabe von Digital Object Identifiern (DOI), ihre Forschungsdaten zitierfähig und dauerhaft zugänglich zu machen, erklärte er. „Eine der zu lösenden Herausforderungen für Bibliotheken ist die Frage, wie in der Forschung mit wissenschaftlicher Software umzugehen ist“, so Nejdl. In den Konferenzvorträgen ging es dann auch darum, Antworten auf Fragen rund um wissenschaftliche Software zu finden und mögliche Lösungsansätze zum Umgang mit Software in der Forschung vorzustellen – aber auch darum, neue Herausforderungen aufzuzeigen.

Nachhaltige Infrastruktur für Software

Edzer Pebesma (Westfälische Wilhelms-Universität, Münster) eröffnete die englischsprachige Konferenz mit seiner Keynote „Incentives and rewards in scientific software communities“ (<https://doi.org/10.5446/31027>). Darin beschäftigte er sich am Beispiel der Programmiersprache R, einer freien Programmiersprache für statistische Berechnungen, mit Anreizen und Belohnungssystemen für Wissenschaftle-



Freuen sich auf viele interessante Vorträge: Wolfgang Nejdl, Irina Sens, Barbara Hartung und Sören Auer (von links)

Gespräche zu den mit diesem Wandel einhergehenden Herausforderungen für Wissenschaftlerinnen und Wissenschaftler sowie Fachleute aus Infrastruktureinrichtungen wie Bibliotheken sein, so Hartungs Wunsch. Im Fokus der Konferenz am 10. und 11. Mai 2017, die von der TIB gemein-

rinnen und Wissenschaftler, die Software entwickeln und bereitstellen. Die Basispakete von R können um zusätzliche Pakete zur Lösung spezieller statistischer Problemstellungen erweitert werden, es existieren aktuell mehr als 10.000 solcher Pakete von 8.000 verschiedenen Autorinnen und Autoren. Bei jeder Änderung an der Programmiersprache R müssten die Autorinnen und Autoren ihre Pakete, die in dem R-Repository CRAN (Comprehensive R Archive Network) mit Metadaten und Abhängigkeiten frei zur Verfügung stehen, jedoch aktualisieren, so dass garantiert ist, dass die jeweiligen Pakete auf einem aktuellen Stand sind. Damit gibt R Wissenschaftlerinnen und Wissenschaftlern die Möglichkeit, für ihre Forschungen von anderen bereits programmierte Software nachzunutzen und bereits existierende R-Pakete zu verwenden. „Die R-Community ist ein gutes Vorbild für eine nachhaltige Infrastruktur einer Software“, so Pebesma abschließend.

Forschungssoftware aus wissenschaftlicher und rechtlicher Sicht

In dem Vortrag „What is good scientific practice for research software?“ (<https://doi.org/10.5446/31028>) von Konrad U. Förstner (Julius-Maximilians-Universität Würzburg; Vorsitzender der Ad-hoc-Arbeitsgruppe Wissenschaftliche Software der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen) ging es um die Symbiose von Wissenschaft und Technologie: Software sei einerseits ein Werkzeug für die Wissenschaft, andererseits könne Software aber auch ein Forschungsergebnis sein. Stärke und Wachstum von Forschung seien heute essenziell mit Software verbunden. Deshalb sei es wichtig, Qualität, Zugang und Zitierbarkeit wissenschaftlicher Software zu sichern. Förstners Forderung: „Gute wissenschaftliche Praxis muss auch für wissenschaftliche Software gelten.“

Nikolaus Forgó (Leibniz Universität Hannover) beleuchtete die rechtliche Perspektive der Software-Thematik unter dem Titel „Legal requirements for software sharing and collaborations“ (<https://doi.org/10.5446/31029>). Sein Schwerpunkt lag auf den relevanten Paragraphen des deutschen Urheberrechtsgesetzes (UrhG). Er betonte, dass Software häufig kollaborativ entwickelt werde und der Aspekt der Urheberschaft bereits vor Beginn einer solchen Kooperation thematisiert werden müsse, um spätere rechtliche Auseinandersetzungen zu vermeiden. Forgó plädierte in diesem Zusammenhang auch für eine bessere Kommunikation und Zusammenarbeit zwischen Juristen und Entwicklern.

Nachhaltige Infrastrukturen für wissenschaftliche Software

Timo Borst (ZBW – Leibniz-Informationszentrum Wirtschaft, Kiel) stellte in seinem Vortrag „Managing research software from the perspective of a scientific infrastructure provider“ die Sicht einer Infrastruktureinrichtung auf wissenschaftliche Software dar. Wissenschaftliche Software sollte Teil von Open Science sein, der Umgang mit Forschungssoftware werde zukünftig ein zentrales Thema für Infrastruktureinrichtungen wie die ZBW. Die Frage sei nur, welche Aspekte das Management von Forschungssoftware umfassen werde – dazu könnten beispielsweise die Verbreitung, die Anerkennung oder die Abbildung verschiedener Versionen von wissenschaftlicher Software gehören.

In dem Vortrag „Solid scenarios for sustainable software“ (<https://doi.org/10.5446/31030>) betonte Patrick J. C. Aerts (Data Archiving and Networked Services (DANS); Netherlands eScience Center), dass Software und Forschungsdaten den gleichen Stellenwert haben sollten, da es sich bei beiden um verschiedene Formen wissenschaftlichen Outputs handele. Aus seiner Sicht spielten



beim Umgang mit wissenschaftlicher Software verschiedene Aspekte eine Rolle: So müsse gefragt werden, welche Software bewahrenswert sei und wie Softwarecode gepflegt werden könnte. Er forderte für die Zukunft klare Richtlinien zum Umgang mit und zur Entwicklung von Softwarecode. Zur Förderung von Nachhaltigkeit von Software sprach Aerts sich für die Anwendung der FAIR-Prinzipien auf wissenschaftliche Software aus – sie sollte auffindbar (findable), zugänglich (accessible), interoperabel (interoperable) und nachnutzbar (reusable) sein.

*Keynote-Speaker
Edzer Pebesma
von der Universität
Münster*

Nachhaltiger Zugriff auf wissenschaftliche Software

Daniel S. Katz (University of Illinois Urbana-Champaign, USA) beschäftigte sich in seinem Vortrag „Software citation: a cornerstone of software-enabled research“ (<https://av.tib.eu/media/31031>) mit der Frage, wie wissenschaftliche Software zitiert werden sollte. Im Rahmen einer FORCE11-Arbeitsgruppe



entstanden Regeln zum Zitieren von Software (<https://doi.org/10.7717/peerj-cs.86>). Ziel dieser Zitiergrundsätze sei es unter anderem, diszipliniübergreifend die Anerkennung von Software als Forschungsergebnis

enthalten, die heute in Webarchiven bewahrt werden, dauerhaft zugänglich gemacht werden können. Ein Weg ist die sogenannte Wayback Machine, mit der verschiedene Versionen von Internetseiten abgerufen werden können. Wünschenswert wäre es jedoch, nicht die URLs mit Zeitstempel zu haben, sondern die Objekte selbst – zum Beispiel wissenschaftliche Software. Das „Tempas TimePortal“, ein im Rahmen des Fachinformationsdienstes Mathematik entwickelter Web-Service, bietet Nutzerinnen und Nutzern die Möglichkeit, den Stand einer Software zu einem bestimmten Zeitpunkt – zum Beispiel dem Publikationsdatum einer wissenschaftlichen Veröffentlichung – in der Vergangenheit darzustellen.

Der erste Konferenztag endete schließlich nach vielen informativen Vorträgen beim Get-together im Alten Rathaus in Hannover. Dort hatten die Teilnehmerinnen und Teilnehmer die Gelegenheit zum Netzwerken und ausreichend Zeit, um sich in historischer Umgebung über die Vorträge des Tages auszutauschen.

Impact und Nachhaltigkeit durch zugängliche, wiederverwendbare und freie Software

„BigDataEurope – The collaborative creation of an open software platform for researchers addressing Europe's societal challenges“ (<https://doi.org/10.5446/31034>) lautete der Titel des Vortrags von Sören Auer (Rheinische Friedrich-Wilhelms-Universität Bonn; Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS, Sankt Augustin). Darin beschäftigte sich Auer zu Beginn des zweiten Konferenztages mit dem Management und der Analyse von Big Data, von großen und komplexen Datenmengen. Durch Big Data eröffnen sich in Bereichen wie Gesundheits-, Energie- und Verkehrswesen oder beim Klimawandel zahlreiche positive Möglichkeiten – sie können dazu beitragen, gesellschaftliche Herausforderungen zu meis-

tern. Das Problem: Die Integration von Big Data in den genannten Bereichen ist oft schwierig. Hier setzt die kollaborative Plattform „BigDataEurope“ (<https://www.big-data-europe.eu/>) an: Die unter Leitung des Fraunhofer IAIS entstandene Plattform soll europäische Unternehmen und Einrichtungen dieser Communities beim Einsatz von und der Arbeit mit Big Data unterstützen, ihnen den Wert von Big Data für ihre Zwecke aufzeigen, den Zugang zu Big-Data-Anwendungen sowie deren Integration in bestehende Arbeitsprozesse erleichtern und so die Verbreitung der Big-Data-Nutzung in verschiedenen Fachgebieten fördern.

Neil Chue Hong (Software Sustainability Institute, Großbritannien) stellte in seinem Vortrag „Software sustainability – guidelines for the selfish scientist“ (<https://doi.org/10.5446/31022>) fest, dass Forschende sich häufig keine Gedanken beim Umgang mit Software machten und sich nicht als Programmierer sähen. Er forderte ein Umdenken hin zu einer stärkeren Anerkennung für das Teilen von Software und einer Verbesserung der Programmierfähigkeiten in der Wissenschaft. Denn: „Better software, better research“ – bessere Software bedeutet bessere Forschung.

Thomas Koprucki (Weierstraß-Institut für Angewandte Analysis und Stochastik – WIAS, Berlin) setzte sich in seinem Vortrag „How to tidy up the jungle of mathematical models? A prerequisite for sustainable research software“ (<https://doi.org/10.5446/31023>) mit mathematischer Modellierung und Simulation (MMS) auseinander, die heute in vielen Fachbereichen grundlegend für das wissenschaftliche Arbeiten sind. Mathematische Modellierung und Simulation seien genauso Teil der Forschungsergebnisse wie Software und Forschungsdaten, so dass auch hier Bedarf an einer Infrastruktur besteht, um die Nachhaltigkeit von MMS zu verbessern.

Im Mittelpunkt des Vortrags „Jupyter



Sören Auer, seit Juli 2017 Direktor der TIB, bei seinem Vortrag über Big Data auf der Konferenz in Hannover

und damit verbundene Softwarezitationen – möglichst mit Digital Object Identifiern (DOI) – zu verbessern. Zu diskutieren seien in diesem Zusammenhang viele offene Fragen, beispielsweise wie mit verschiedenen Versionen von Software umgegangen werden sollte.

Martin Fenner (DataCite, Hannover) zeigte in seinem Vortrag „Workflows für assigning and tracking DOIs für scientific software“ (<https://doi.org/10.5446/31032>), welche Aufgaben im Zusammenhang mit wissenschaftlicher Software noch zu lösen sind: Neben der grundsätzlichen Frage, wie wissenschaftliche Software definiert werden sollte, ging es dabei unter anderem um fehlende Metadaten und fehlende Archive sowie um die Versionsproblematik bei Software. Er plädierte dafür, wissenschaftliche Software mit einem Digital Object Identifier zu versehen, weil diese Software und die mit ihr zusammenhängenden digitalen Objekte (wie Veröffentlichungen oder Autorenidentifikationen wie ORCID) besser recherchier- und auffindbar seien.

In dem Vortrag „Software as a first-class citizen in web archives“ (<https://doi.org/10.5446/31033>) von Helge Holzmann (Forschungszentrum L3S, Hannover) ging es darum, wie die unzähligen Arten von Infor-

and IPython facilitating open access and reproducible research“ (<https://doi.org/10.5446/31024>) von Benjamin Ragan-Kelley (Simula Research Laboratory; Jupyter, Norwegen) stand die Open-Source-Webanwendung „Jupyter Notebook“, über die Dokumente interaktiv erstellt und geteilt werden können. Diese Dokumente können Code enthalten, aber auch Text, Visualisierungen oder Berechnungen. Das Besondere neben den interaktiven Aspekten an Jupyter ist, dass die Notebook-Dokumente Repräsentationen aller Inhalte speichern, darunter beispielsweise auch Eingaben von Berechnungen oder erklärende Texte, so dass dadurch eine Reproduzierbarkeit der wissenschaftlichen Vorgehensweise gegeben ist.

„Blockchain for science and knowledge creation: An intro and overview“ (<https://doi.org/10.5446/31025>) lautete der Vortrag von Sönke Bartling (Alexander von Humboldt Institute für Internet und Gesellschaft (HIIG), Berlin), in dem es darum ging, wie die sogenannte Blockchain-Technologie als dezentrales und transparentes Datenregister für den Bereich Open Science genutzt werden kann. Neben der primären Aufgabe der Datenerhebung könnten alle anderen Schritte im Forschungszyklus (zum Beispiel die Datenanalyse und deren eindeutige Identifizierung) in einem Blockchain-System erfolgen. So wären beispielsweise die Daten, die verarbeitet werden, dank eines Zeitstempels eindeutig zuordbar. Die Blockchain-Umgebung ermöglicht weiterhin die Bereitstellung von Computerprogrammen, sogenannten Smart Contracts, die definierte Anwendungen ausführen können – so ermöglichen sie es, durch den dezentralen Zugriff auf Datenquellen auch sensible, personenschutzrechtlich relevante Teilmengen von Daten auszuwerten, ohne zum Beispiel einen Personenbezug herzustellen.

James Littlejohn (Edinburgh Napier University, Großbritannien) beschäftigte sich in seinem Vortrag „Dsensor.org peer to peer science“ (<https://doi.org/10.5446/31026>) mit den drei Säulen von Blockchain: Kryptographie, Wirtschaftlichkeit und Verhalten. Er demonstrierte, wie die Blockchain-Technologie eingesetzt werden kann, um die wissenschaftliche Forschung offen und transparent zu gestalten („to keep science honest“). Am Beispiel einer Live-Demonstration für pharmazeutische Produkte verknüpfte er innerhalb einer Blockchain-Umgebung ein Software-Programm zur Behandlung von Cholesterin-Überschuss mit der dazugehörigen Datenbank. Er zeigte, dass die Blockchain-Technologie Wissenschaftlerinnen und Wissenschaftler dabei unterstützen kann, wissenschaftliche Forschungsfragen nachvollziehbar gestalten, wodurch die Problematik der mangelnden Reproduzierbarkeit vieler wissenschaftlicher Fragestellungen mithilfe von Blockchain angegangen werden kann.

Wissenschaftliche Software – es ist noch viel zu tun

Die Konferenz bot Wissenschaftlerinnen und Wissenschaftlern sowie Fachleuten aus Infrastruktureinrichtungen ein Forum, sich über Anforderungen, Erwartungen und Bedürfnisse, die sie in der Praxis an wissenschaftliche Software stellen, auszutauschen. Die Vorträge zu verschiedenen Aspekten wissenschaftlicher Software zeigten deutlich, dass in diesem Bereich noch viel zu tun ist: von der Schaffung von Regeln für den Umgang mit wissenschaftlicher Software über Infrastrukturen zum Bewahren von Software, urheberrechtliche Fragen bei der Softwareentwicklung und die Anerkennung von Softwarecode als wissenschaftlichem Output bis hin zur Nutzung der Blockchain-Technologie in



der Wissenschaft. „Nicht-textuellen Materialien wie Software kommt in der Wissenschaft eine wichtige Rolle zu, das Thema ist noch lange nicht umfassend behandelt“, sagte Irina Sens bei ihrer Verabschiedung am zweiten Konferenztag. „Einige Fragestellungen könnten auf einer dritten Conference on Non-Textual Information diskutiert werden“, kündigte sie eine mögliche Fortführung der Konferenzreihe an.

Die Vorträge der Konferenz (<https://av.tib.eu/series/310/2nd+conference+on+non+textual+information+software+and+services+for+science+s3+may+10+11+2017+in+hannover>) sind im AV-Portal der TIB (<https://av.tib.eu/>) verfügbar. |

Mehr Informationen:

www.nontextualinformation2017.de

Dr. Sandra Niemeyer

Technische Informationsbibliothek (TIB)
Pressereferentin
Kommunikation und Marketing
Welfengarten 1 B
30167 Hannover
sandra.niemeyer@tib.eu