



Forschungsdatenmanagement – Herausforderungen für Rechenzentren

Philipp Wieder, GWWDG

14.03.2018, ZKI 2018, Konstanz

Kontext



Georg-August-Universität Göttingen

- 200 Institute und 13 Fakultäten
- Anzahl Mitarbeiter: 14.000
- Wissen. Beschäftigte: 3.000, Professoren: 400
- Studierende: 34.000

Universitätsmedizin (UMG)

- 7.000 Beschäftigte

Göttingen Campus Partner

- 5 Max Planck Institute, Leibniz Institut DPZ, DLR, Akademie der Wissenschaften, ...

Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

- Gemeinsames Rechen- und Kompetenzzentrum der Universität Göttingen und der Max Planck Gesellschaft
- IT-Dienste für über 80.000 Nutzer

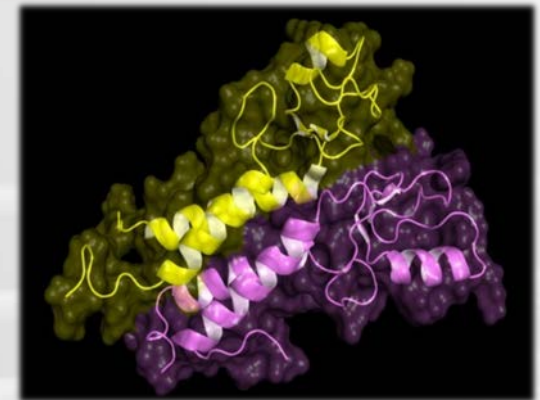


Standortbestimmung

Daten im Spannungsfeld



- „Digitalisierung“ rückt auch das Thema Daten in den Fokus
 - Gesellschaft
 - Politik
 - Fördereinrichtungen
- Forschung hängt zunehmend von der Integration daten- und rechenbezogener Methoden und Infrastrukturen ab
 - „Scientific Computing“ „Data Science“, „Big Data“, „Data Analytics“, ...
- Universitäten und Forschungseinrichtungen stehen vor massiven Herausforderungen

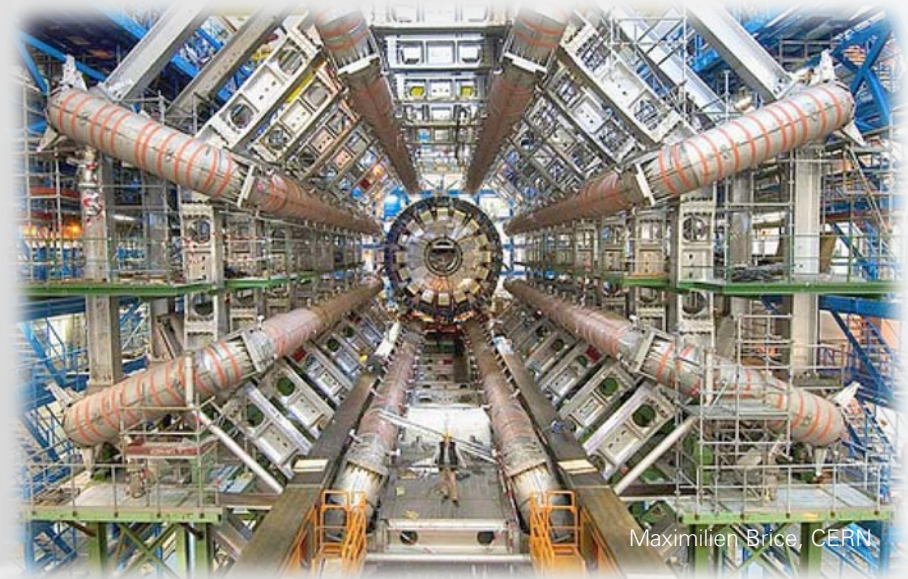


(Forschungs-)Daten sind Assets von gesellschaftlicher Bedeutung

Beispiel I

Large Hadron Collider – CERN

- 600 Mio. Kollisionen pro Sekunde
- 1 Petabyte Rohdaten werden pro Tag verarbeitet
- 30 Petabyte Daten werden pro Jahr gespeichert (entspr. 7,5 Mio. DVDs)
- Weitere Beispiele
 - Astronomie
 - Meteorologie



Beispiel II

Sozial-ökologische Systeme – FOR2432



Herausforderungen für RZ

Themenbereiche



- Einordnung und Abgrenzung
- Beratung, Training & Ausbildung
- Infrastruktur & Dienste

Systematik

- Status
- Herausforderungen
- Beispiele für Lösungen

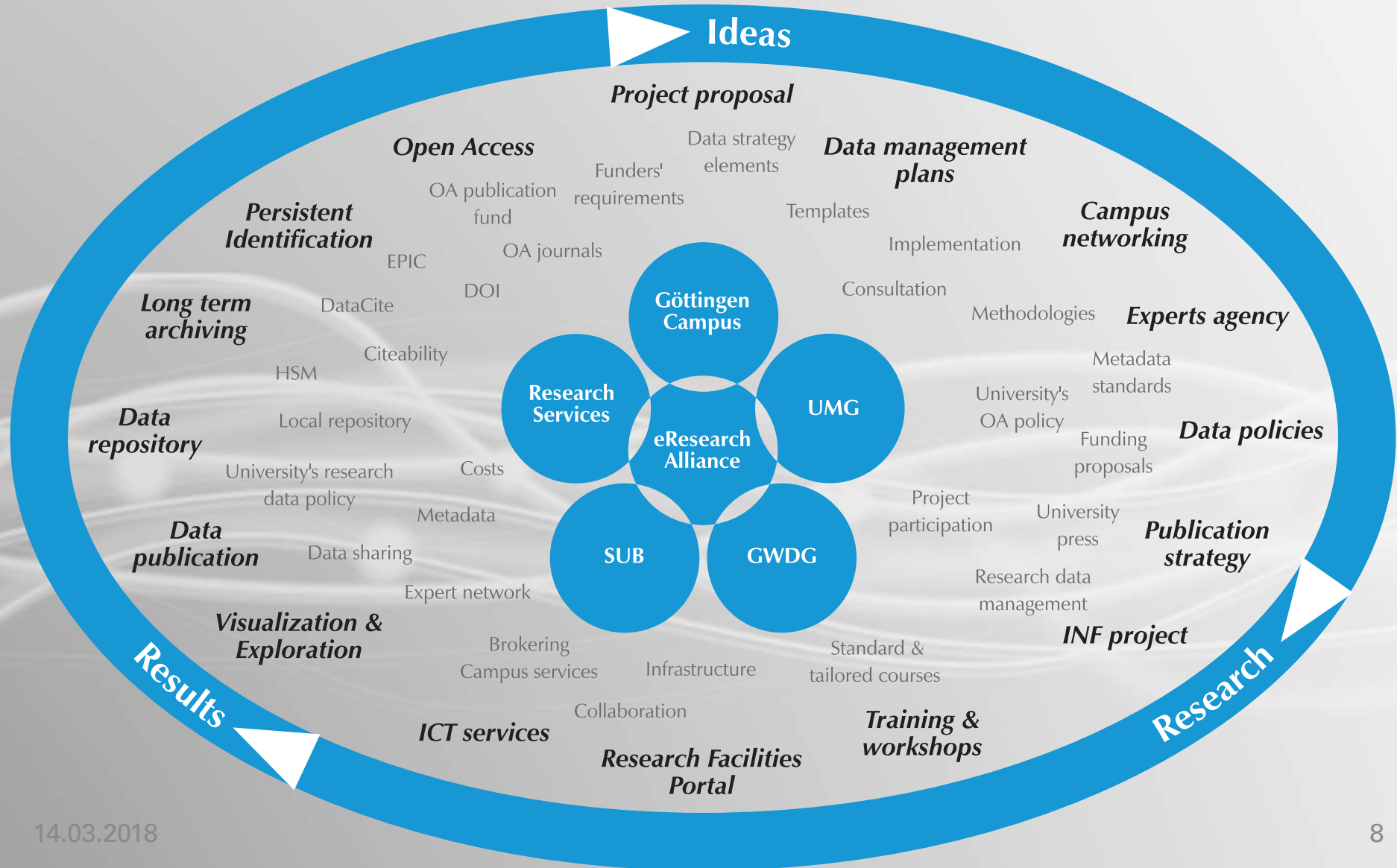


Einordnung & Abgrenzung

Research Data Life-Cycle A



Einordnung & Abgrenzung Research Data Life-Cycle B



Einordnung & Abgrenzung

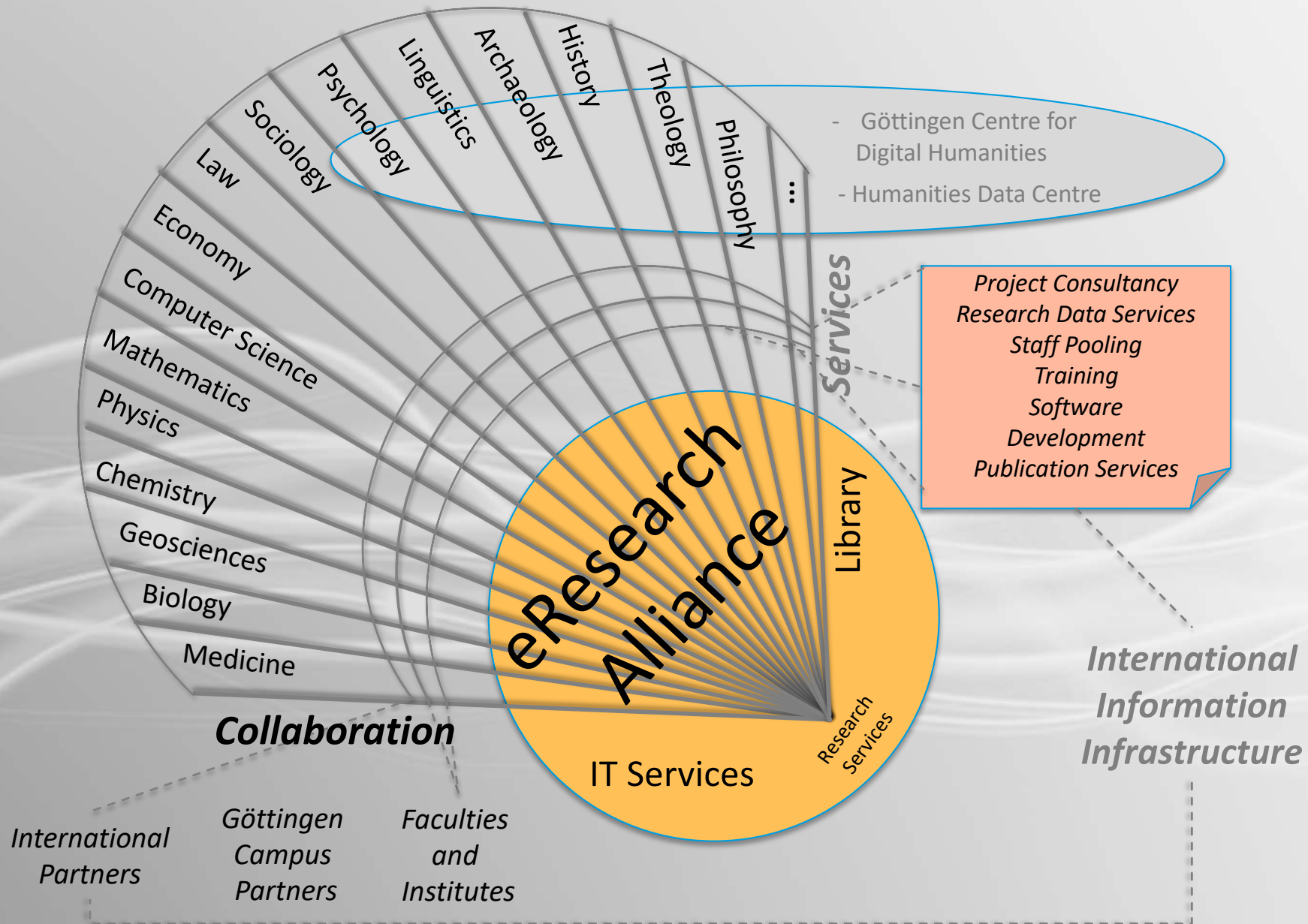
Herausforderungen



- **Definition der eigenen Aufgaben**
 - Spannungsfeld zwischen Basisdiensten & komplexen Anwendungen für Forschungsdatenmanagement
- **Konkurrenz & Partnerschaften**
 - Lokal, national, international
- **Outsourcing**
 - Regularien und Kosten
- **Technologie Watch**
 - Hochdynamischer Markt
 - Wachsende Komplexität

Einordnung & Abgrenzung

Göttingen eResearch Alliance



„400TB Bilddaten aus Projekt X müssen in diesem Jahr irgendwie gespeichert werden.“

„Ich habe hier eine DB, die 2010 entwickelt wurde, und die noch mind. 10 Jahre laufen muss.“

„Für ein Journal muss ich meine Daten referenzierbar machen.“

„Wir benötigen Datenspeicher mit S3 Schnittstelle.“

„Ich möchte meine Anwendung inklusive der virtuellen Maschine archivieren.“

- **Beratungsangebot wird zunehmend breiter**
 - Speicherung → Analyse → Nachnutzung
- **Wissensmanagement bekommt wachsende Bedeutung**
 - DCIM, ITSM, CRM, DMP, ...
 - Brokering
- **Zunehmende Diversifizierung der Qualifizierungen**
 - Data Engineer vs. Data Scientist vs. ...
- **Zunehmende Konkurrenz**
 - Genereller Fachkräftemangel
 - RZ in Konkurrenz zu Fachwissenschaften

Beratung, Training & Ausbildung

Beispiele für Lösungen



Göttingen / eResearch Alliance

Why Research Data Management?

- Information Event -
for Researchers regarding eResearch at the UMG

Research, Data, Management:

Data management basics and services for your daily research routine at the UMG

9 December 2016, at 10 - 11 am
In lecture hall 81 at the UMG

This information event addresses all researchers at the UMC.

During your daily research routine, digital research data occur in different amounts and forms (e.g. measuring data, pictures, videos, DNA-sequences etc.). What possibilities do exist to store and archive research data in a secure and sustainable manner to process the data further?

Meet the contact persons and get to know useful services supporting you during all phases of the research process.

Jens Dierkes
Göttingen eResearch Alliance
Dagmar Härter
SUB Bibliothek/Medizin
Marie Hedderich
Referat Forschung
Harald Kasch
Department of Medical Informatics
Andreas Lehmann
Med. Biometry and Statistical Bioinformatics
Sandra Sellke
UMC BioBank
Doris Wetscherek
Ethikkommission der UMG
Philipp Wieder
GWDC



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

In publica commoda

DIE UNIVERSITÄT FAKULTÄTEN FORSCHUNG STUDIUM EINRICHTUNGEN ALUMNI SERVICE INTERNATIONAL

» STU... » STUDIEN... » STUDIENFÄCHER... » ANGEWANDTE DATA SCIENCE (B.SC.) – GEPLANT ZU...

Angewandte Data Science (B.Sc.) – geplant zum Wintersemester 2018/2019

Name des Studienfachs: Angewandte Data Science (B.Sc.)
Abschluss: Bachelor of Science (B.Sc.)
Regelstudienzeit: 6 Semester
Studienbeginn: Wintersemester
Teilzeitstudium möglich

Was ist Angewandte Data Science?
Das Gebiet "Data Science" ist an der Schnittstelle der Mathematik, der Informatik, und der Statistik angesiedelt. Data Science beschäftigt sich mit der Analyse und dem Erkenntnisgewinn aus Daten sowie den benötigten Techniken zum Verarbeiten von großen und teilweise unstrukturierten Datenmengen. Im Bachelorstudiengang "Angewandte Data Science" werden aufbauend auf den Grundlagen der Informatik und Mathematik vertiefende Kenntnisse der Datenanalyse vermittelt. Weiterhin lernen Studierende in einem Anwendungsfach praktischen Einsatz der erlernten Methoden zur Datenanalyse kennen. Zu den vermittelten Kenntnissen der Datenanalyse gehören Aspekte des maschinellen Lernens, der Statistik, der Mustererkennung und der für eine effiziente Analyse benötigten Infrastrukturen. Als Anwendungsfächer können Studierende Wirtschaft, Biologie, digitale Geisteswissenschaften, medizinische Informatik, und Züchtungsinformatik wählen.

INFORMATIONEN ZUM FACH

Homepage des Programms
Ordnungen
Fakultät für Mathematik und Informatik

STUDIENAUFBAU

Aufbau
Stundenplan für Erstsemester

Data Science Summer School

July 10 – 21 2017
Göttingen, Germany

Topics

- Data Lifecycle and Data Management
- Data Science Methods (e.g., Statistics, Machine Learning, Text Analysis)
- Visualization & Storytelling
- Infrastructures and Platforms
- Ethical, Legal, and Social Aspects
- Application Examples

Keynotes from Academia & Economy + Theory + Practice + Hands-On + ECTS
Deadline for Application: March 15 2017
More Information and Application at <http://bit.ly/datascienceUGOE17>

Part of International Short Term Programs Göttingen

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

Göttingen / eResearch Alliance

This map charts the most popular cosmic locations as measured by the number of SkyView image requests since June 2007. Brighter colors indicate more views. Credit: NASA/SkyView/Thomas McClintock

Infrastruktur und Dienste

Portfolio Management



NEUIGKEITEN

DIENSTE ×

FORSCHUNG & LEHRE

ÜBER UNS

SUPPORT

ANMELDEN

EN DE



Speicherdienste

- Backup
- Archivierung
- GWWDG CrashPlan PROe
- Fileservice
- GWWDG Cloud Share
- GWWDG ownCloud
- Cryptshare

E-Mail & Kollaboration

- E-Mail-Service (MS Exchange 2010)
- Spam- und Virenerkennung
- Mailinglisten
- MS SharePoint
- Managed Services
- Projektmanagement-Service
- Etherpad
- ShareLaTeX
- Rocket.Chat
- GitLab

Server-Dienste

- Virtuelle Server
- Hosting/Housing von Servern
- Webhosting
- GWWDG Cloud Server
- FTP Server
- Puppetserver

Netzwerkdienste

- IP-Adress-Management-System
- Kabel- und Trassenmanagement-System
- Systemüberwachung
- Einrichtung von eduroam
- Anschluss an das Active Directory
- Benutzerverwaltung mit OpenLDAP
- Client-Management
- Client-Management für macOS und iOS

Anwendungsdienste

- Persistent Identifier (PID)
- High Performance Computing
- Bibliotheksdienst Aleph
- Datenbankdienst Oracle
- Datenbankdienst MySQL
- Bewerbungs-/Registrierungssysteme
- Plagiatsvermeidung
- Online-Umfragen
- Bioinformatik-Programme
- Statistik-Programme
- Jupyter

IT-Sicherheit

- Sicherheitstests eines Netzwerkes zur Dat...
- Public-Key-Infrastruktur (PKI)
- Authentifizierungs/Autorisierungs-Infrastruk...
- Virenschutz (Sophos-Update-Service)

Allgemeine Dienste

- Identity Management
- Kurse
- Software und Lizenzverwaltung
- Videoconferencing
- Leihrechnerpool
- Druck- & Scandienste

IT-Beratung

- Forschungsdaten-Management
- IT-Sicherheit
- Hardwarekauf
- Apple-Beratungszentrum
- Aufbau von Verzeichnisdiensten (AD, LDAP)
- Planung von Datenübertragungsnetzen

DIENSTE

Infrastruktur und Dienste

Fokus Datenmanagement

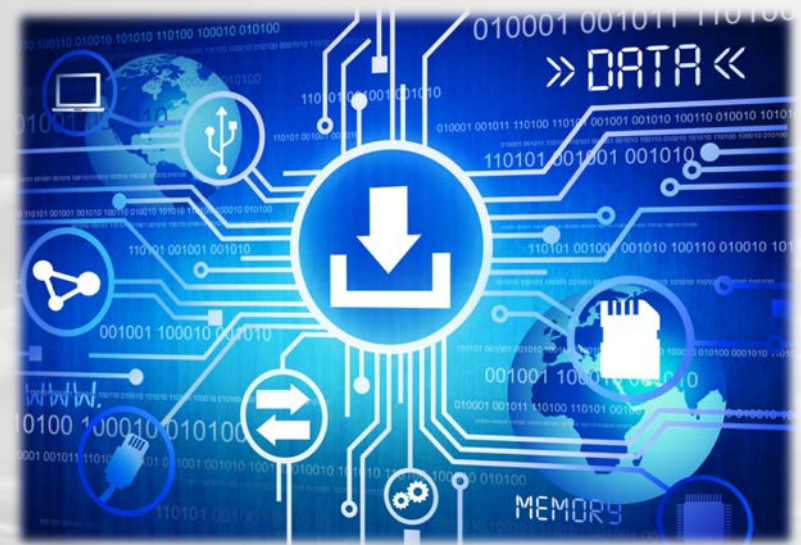


- **Spezielle Anforderungen je nach Anwendungsfall**
- **In der Regel aber campusweite Lösungen**
 - Daten-Repository
 - Datenmanagementplanung
 - Langzeitarchiv
- **Aktuelle Bestrebungen integrativ bzw. prozessbezogen**
 - Stichwort Reproducible Research
- **Open Access goes Open Science**

Infrastruktur und Dienste

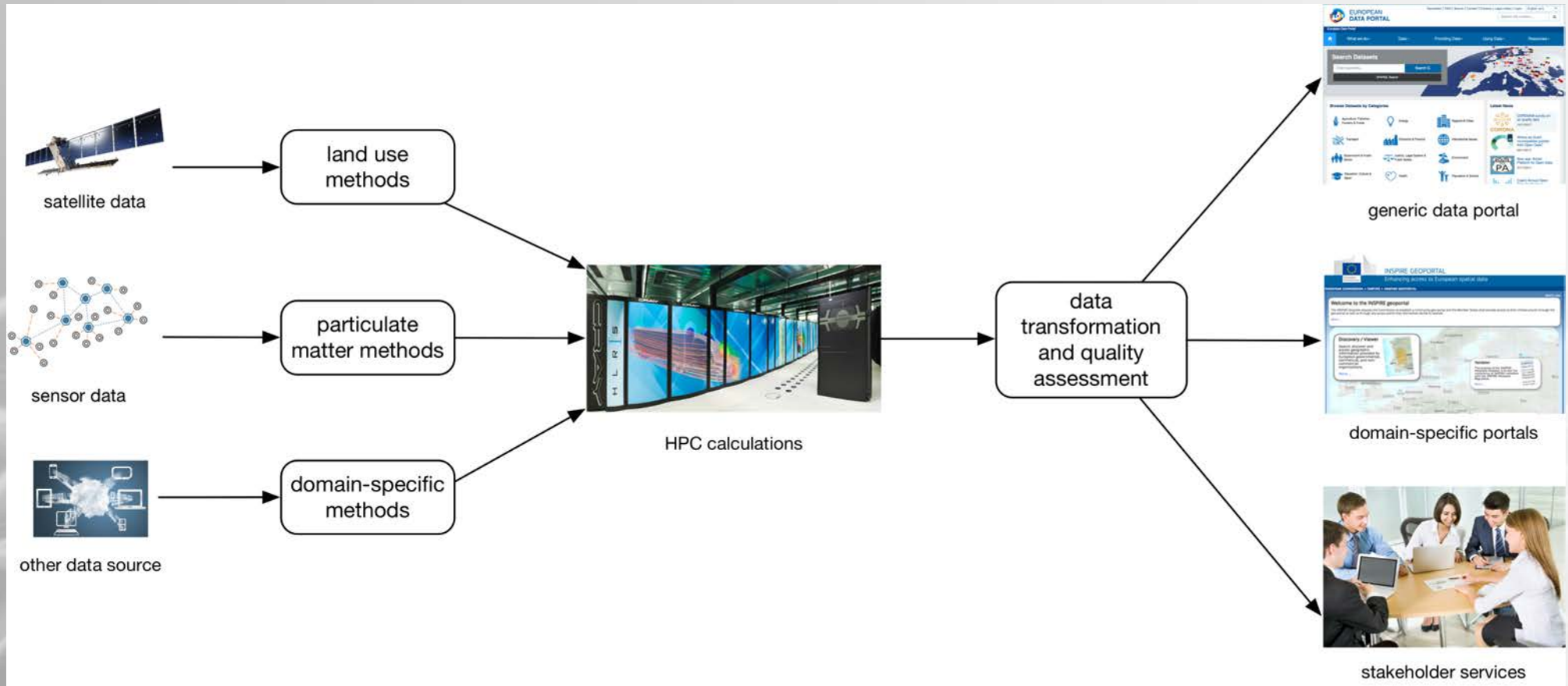
Herausforderungen

- **Wirtschaftlichkeitsbetrachtung**
 - Gegenwärtig PByte O(100) kEuros
 - Lizenzgebühren und –gebaren
 - Budgetaufwuchs
- **Umgang mit Daten**
 - Gesetze & Policies
 - Aufbewahrungsfristen
- **Long tail of science**
- **Einrichtung und Betrieb von Pipelines für das Forschungsdatenmanagement**



Infrastruktur und Dienste

Beispiele für Lösungen



Was bringt die Zukunft?

Ausblick

- **Einen massiven Zuwachs an Datenquellen**
 - Neue Disziplinen, Anwendungen, Methoden
 - Mehr Speicherbedarf, aber auch bessere Übertragungswege
- **Höhere Integrationsdichte**
 - Einzelne Dienste vs. Pipelines
 - Übergang zwischen Diensten
- **Informationsgewinnung**
 - Speicherung vs. Analyse
 - Neue Anforderungen an Infrastrukturen



Quellen

- **Digital Curation Center**
 - <http://www.dcc.ac.uk/>
- **Göttingen eResearch Alliance**
 - <http://www.eresearch.uni-goettingen.de/>
- **nestor**
 - http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html
- **OpenAIRE**
 - <https://www.openaire.eu/>
- **Research Data Alliance**
 - <https://www.rd-alliance.org/>
- **RFII/NFDI**
 - <http://www.rfii.de/de/category/dokumente/>

Fragen und Antworten



Bildnachweise

- Titelfolie: © j-mel | Fotolia
- Folie 3: © pressmaster | Fotolia, © petarg | Fotolia, © sdecoret | Fotolia (v. oben nach unten)
- Folie 4: © CERN, © CERN/Maximilien Brice, (v. oben nach unten)
- Folie 5: unbekannt, © Philipp Wieder, <http://www.mdpi.com/1424-8220/15/12/29765xml>, unbekannt, © FOR 2432, © FOR 2432 (im Uhrzeigersinn beginnend oben rechts)
- Folie 6: © Antrey | Fotolia
- Folie 13: © Daniel Kurzawe
- Folie 16: © Rawpixel.com | Fotolia
- Folie 18: © Rawpixel.com | Fotolia
- Folie 20: © Rawpixel.com | Fotolia