

Rat für Sozial- und
Wirtschaftsdaten (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

Nr. 168

Aktuelle Herausforderungen für die
wissenschaftliche Informationsinfrastruktur

Sabine Brünger-Weilandt

Januar 2011

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienen empirischen Forschungsarbeiten, sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Aktuelle Herausforderungen für die wissenschaftliche Informationsinfrastruktur

Sabine Brünger-Weilandt

Dieses Working Paper dokumentiert die Präsentationsfolien eines Vortrags auf der 5. Konferenz für Sozial- und Wirtschaftsdaten (KSWD) am 13. Januar 2011

Abstract

Über Infrastruktur redet man nicht ... sie ist schlicht da und „funktioniert“. Sichtbar wird sie erst auf Grund der Tatsache, dass sie NICHT da ist oder nur eingeschränkt. Das gilt in gleicher Weise für Mobilität, Energieversorgung, Müllentsorgung aber eben auch für Informations-Infrastrukturen.

Versorgung ist eine Kernaufgabe der wissenschaftlichen Informationsinfrastruktur – nämlich die Versorgung von Wissenschaft und Forschung mit (digitaler) Information und damit zusammenhängenden Dienstleistungen. Diese Kernaufgabe richtet sich auf die Unterstützung der Wissenschaften – und des einzelnen Wissenschaftlers – in allen Phasen des Forschungsprozesses. Die großen gesellschaftlichen Themen der Zukunft stellen Wissenschaft, Wirtschaft und Politik vor völlig neue Herausforderungen. Dies bedeutet für alle Beteiligten starke Veränderungen hinsichtlich Organisationen, Strukturen, Innovationsprozessen und Wertschöpfungsketten. Forschungsprozesse sind ebenfalls Wertschöpfungsketten. Hier werden innovative und leistungsstarke Informationsinfrastrukturen als Fundament für die Weiterentwicklung von Wissenschaft und Forschung benötigt.

Aktuelle Herausforderungen für die wissenschaftliche Informationsinfrastruktur

**Rat für Sozial- und Wirtschaftsdaten
5. KSWD, 13. Januar 2011**

Sabine Brünger-Weilandt



Herausforderungen für die wissenschaftliche Informationsinfrastruktur

Informationsinfrastruktur gestern und heute: Auftrag und Ziel

Zwei Anwendungsszenarien: Realität und Vision

Sidestep: Daten

Aktuelle Herausforderungen und Konsequenzen

Politische Initiativen

Ausblick

Über Infrastruktur redet man nicht ...



„Versorgung“ ist ein zentrales Element auch von Informationsinfrastruktur

SOFIS SOLIS



Foto: Nicola, „Natur und Technik“, CC-Lizenz (BY 2.0)
<http://creativecommons.org/licenses/by/2.0/de/deed.de>
Quelle: www.piqs.de

Früher war sowieso alles besser ...

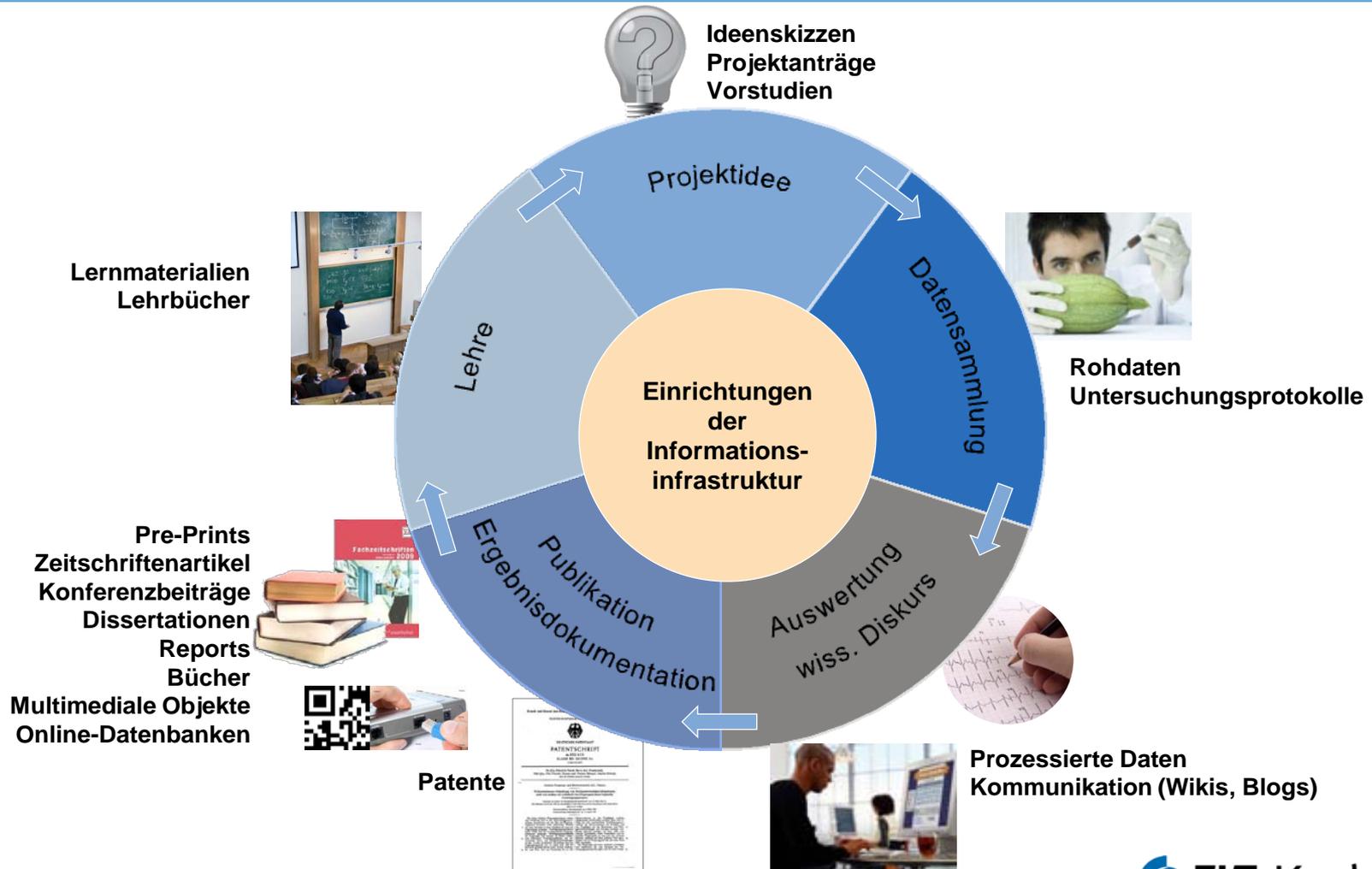
Die Informationsinfrastruktur
war überschaubar und
geregelt:

Archive, Bibliotheken,
Sammlungen; Quellen,
Bücher und Zeitschriften



Foto: Palmbeach, „Allein“, CC-Lizenz (BY 2.0)
<http://creativecommons.org/licenses/by/2.0/de/od/>
Quelle: www.pigs.de

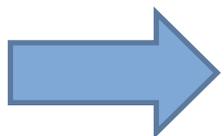
Die Informationsinfrastruktur unterstützt den **gesamten** wissenschaftlichen Arbeitsprozess



Die Informationsinfrastruktur ist Bestandteil der Forschungsinfrastruktur

Sie unterstützt die Forschung, indem sie

- die Sammlung und Aufbereitung ebenso wie die Erstellung und Erschließung von Information betreibt
- die – nachhaltige! – Retrieval- und Analysefähigkeit von relevanter Information sicherstellt
- das Management von Information aller Art (Daten, Objekte, Medien) betreibt
- den dauerhaften Zugriff darauf sicherstellt
- Sicherheit und Vertraulichkeit gewährleistet
- unabhängig von Disziplinen und Methoden ausgerichtet ist



Nutzen für die Wissenschaft

Ötzi, the Iceman

Pollenanalyse

Gut erhaltene Pollen der Hopfenbuche lassen darauf schließen, dass Ötzi im Frühling oder Frühsommer ums Leben kam.

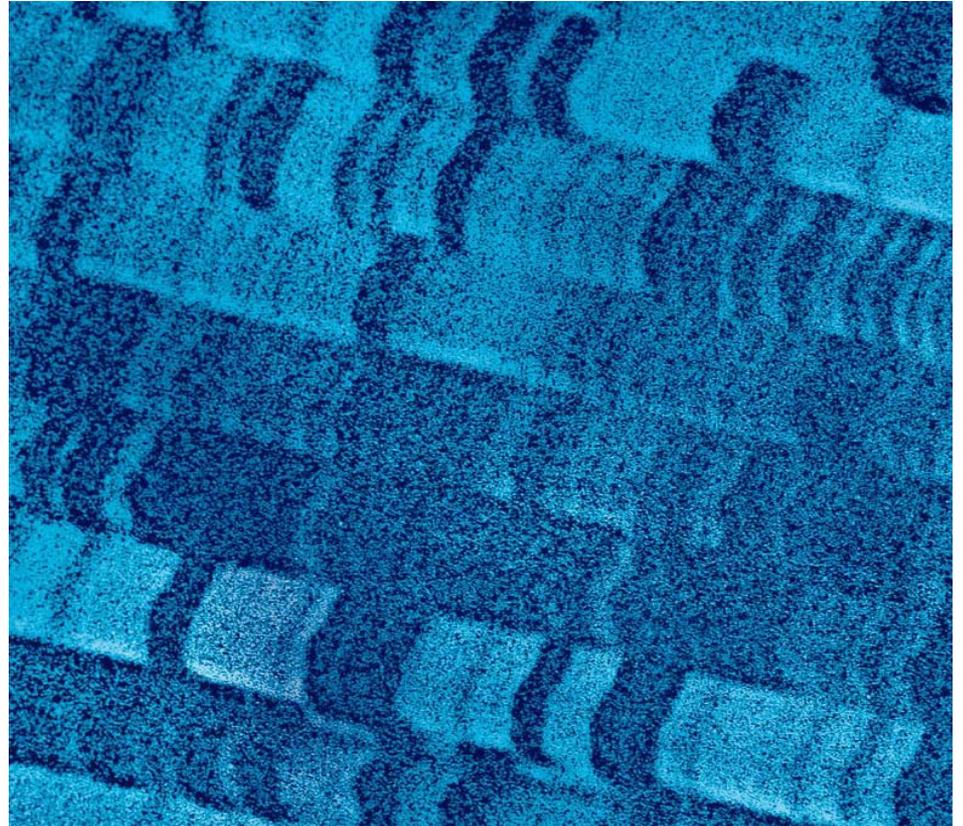


Chemische Analyse

- ❑ Kupfer und Arsen in den Haaren lassen vermuten, dass Ötzi mit Kupferschmelze zu tun hatte.
- ❑ Isotope bestimmter chemischer Elemente in den Zähnen weisen ein für das Eisacktal (Südtirol) typisches Muster auf.

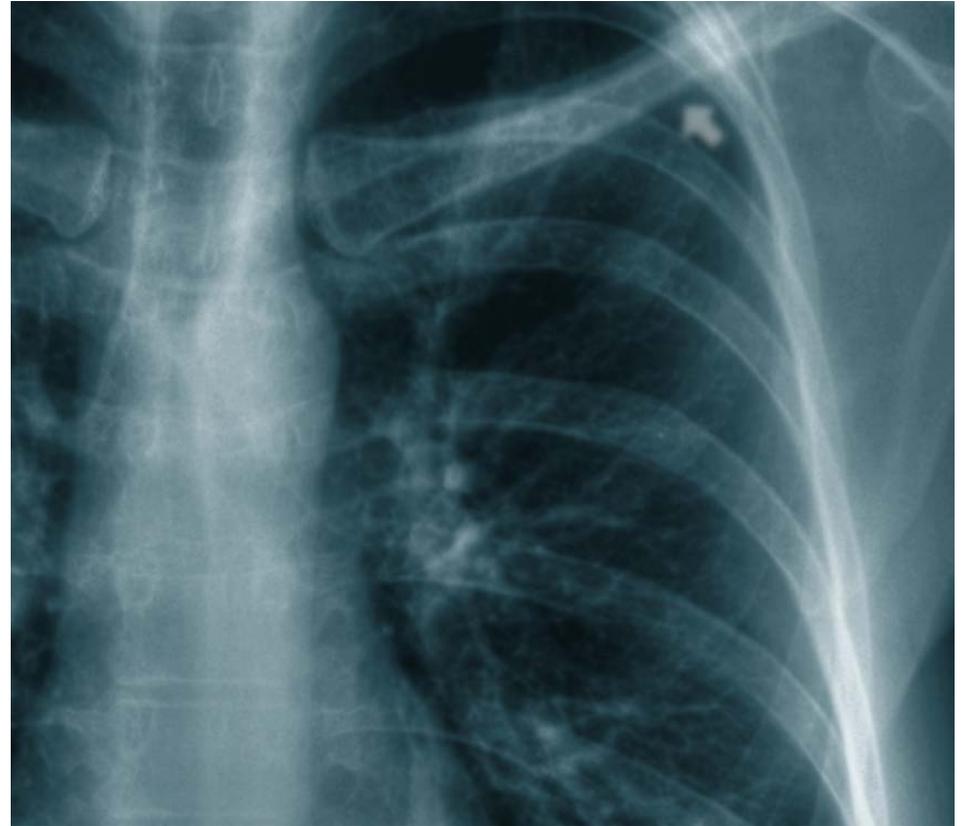
DNA-Analyse

- ❑ Ötzi gehörte zur K-Gruppe, einer sehr alten unter den 9 europäischen Gruppen.
- ❑ 2 Mutationen lassen auf Unfruchtbarkeit schließen.



Computertomographie

Ötzi starb durch eine Arterienverletzung im Schulterbereich.



Stereolithographie

- ❑ Modell von Ötzis Schädel:
mit Lasertechnik aus
Acrylharz gefertigt
- ❑ ursprünglich als Technik für
die Erstellung von Proto-
typen in der Automobil-
industrie entwickelt
- ❑ mittlerweile Einsatz in
vielen Disziplinen – von der
plastischen Chirurgie bis hin
zur Archäologie



Advancing Science – Forschung nach dem Google-Prinzip?

„Sergey Brin lässt sein Genom entschlüsseln – und erfährt, dass er sehr wahrscheinlich eines Tages an Parkinson erkranken wird.“

Süddeutsche Zeitung Magazin, Heft 45/2010



welcome

ancestry

health

how it works

store



Let your DNA help you plan for the important things in life.

Take charge of your health and wellness today.



[Order Now »](#)

Upgrade your health records with your carrier status.

Learn if you carry inheritable disease markers. [» Learn more](#)



Carrier Status →

Live well at any age.

Make better health decisions by learning your genetic risks. [» Learn more](#)



Disease Risk →

Get the treatment that's right for you.

Know in advance how you'll respond to certain medications. [» Learn more](#)



Drug Response →

welcome

ancestry

health

how it works

store



Your Health

- Discover disease risk factors
- Screen for carrier status
- Know your predicted response to drugs



Your Ancestry

- Trace your ancestral lineage
- Find and connect with family members
- Uncover your heritage



You'll also get exclusive access to participate in groundbreaking genetic research.

Find a disease or trait that we cover:

Select a Disease or Trait

Popular Topics:

- Type 2 Diabetes
- Restless Legs Syndrome
- Phenylketonuria
- Prionosis
- Breast Cancer
- Colorectal Cancer
- Prostate Cancer
- Celiac Disease
- Crohn's Disease
- Hemochromatosis
- Parkinson's Disease
- Sexuality
- Placental Efficacy

[Browse all 182 health and trait topics >](#)

News and Press

23andMe Announces Additional Investment in Series C Financing **January 7, 2011**

23andMe Receives Funding from the National Institutes of Health to Evaluate Web-Based Research on the Genetics of Drug Response **December 16, 2010**

23andMe Announces Immediate Availability of Upgraded Genotyping Array **November 24, 2010**

[View more >](#)

Additional Learning and Resources

Research Initiatives

- **• Parkinson's Disease**
- Huntington's
- Sarcoma

Physician Resources

[Read our open letter to the medical community >](#)

Our Science

- Scientific Advisory Board
- [Read our open letter to the scientific community >](#)

Insurance, Privacy and Genetic Discrimination

[Learn how the Genetic Non-Discrimination Act protects your genetic privacy >](#)

welcome

ancestry

health

how it works

store

Parkinson's Disease - Sample Report

[View all sample reports >](#)

Established Research report on 1 reported marker.

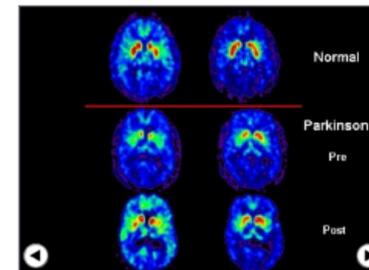
Example Data | How It Works | Technical Report

Keep in mind that estimates of Parkinson's risk due to the G2019S mutation in the LRRK2 gene vary greatly. While it is well established that the mutation's effect is very strong, there is no consensus about its exact magnitude.

About Parkinson's Disease

Parkinson's disease is a disorder of the brain's motor system caused by a loss of dopamine-producing brain cells. Approximately one and a half million Americans have the disease, and about 50,000 are diagnosed each year. The main symptoms are trembling in the hands, arms, legs, jaw, and face; stiffness of the limbs and trunk; slow movement; and impaired balance and coordination. Symptoms of Parkinson's disease usually come on gradually and affect people over the age of 50, although there are rare forms that progress more quickly and strike at a younger age. Though very little is known about the genetics of Parkinson's, mutations in a gene known as LRRK2 have been found to greatly increase a person's likelihood of developing the condition.

[Learn more about the biology of Parkinson's Disease >](#)



1 of 2. Decreased dopamine activity in the brains of people with Parkinson's disease can be seen on a PET scan.



[visit the store >](#)

[try a demo >](#)

Example Genetic Data

Information for Greg Mendel (Dad) assuming European ethnicity and an age range of

Greg Mendel (Dad) does not have the LRRK2 G2019S mutation. The G2019S mutation is not the only factor contributing to Parkinson's. People without this mutation can still develop Parkinson's Disease.



Greg Mendel (Dad)

1.6 out of 100

people of European ethnicity who share Greg Mendel (Dad)'s genotype will get Parkinson's Disease between the ages of 30 and 79.



Average

1.6 out of 100

people of European ethnicity will get Parkinson's Disease between the ages of 30 and 79.

What does the Odds Calculator show me?

Use the ethnicity and age range selectors above to see the estimated likelihood of Parkinson's Disease due to genetics for someone with Greg Mendel (Dad)'s genotype. The 23andMe Odds Calculator assumes that a person is free of the condition at the lower age in the range. You can use the same selector above to see the estimated likelihood of Parkinson's Disease for the genotypes of other people in your account.

The 23andMe Odds Calculator only takes into account effects of markers with known associations that are also on our genotyping chip. Keep in mind that aside from genetics, environment and lifestyle may also contribute to one's Parkinson's disease.

The Fourth Paradigm

Thousand years ago: Science was **empirical**

- describing natural phenomena

Last few hundred years: **theoretical** branch

- using models, generalizations

Last few decades: a **computational** branch

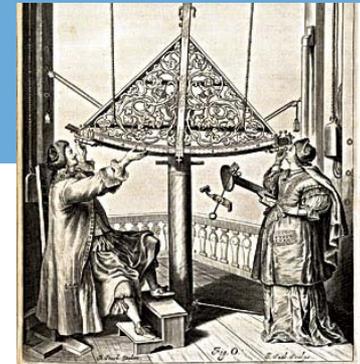
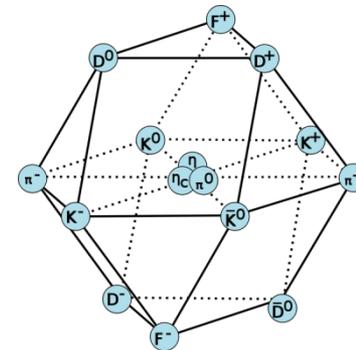
- simulating complex phenomena

Today: **data exploration** (eScience)

unify theory, experiment, and simulation

- Data captured by instruments or generated by simulation
- Processed by software
- Information/knowledge stored in computer
- Scientist analyzes database/files using data management and statistics

nach: Jim Gray on eScience (Talk), January 2007



$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}$$



Advancing Science – „A vision for 2030...“

*The creation of **scientific e-infrastructure** is a means, not an end. It is a means to new science, new solutions and new progress in society.*

*...our vision points in the direction of an infrastructure that supports **seamless access, use, reuse, and trust of data**.*

It suggests a future in which the data infrastructure becomes invisible, and the data themselves have become infrastructure – a valuable asset, on which science, technology, the economy and society can advance.

Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Commission, October 2010

Herausforderungen für die wissenschaftliche Informationsinfrastruktur

Informationsinfrastruktur gestern und heute: Auftrag und Ziel

Zwei Anwendungsszenarien: Realität und Vision

Sidestep: Daten

Herausforderungen und Konsequenzen

Politische Initiativen

Ausblick

Data curation

„Data curation is the active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities across the sciences, social sciences, and the humanities.

*Data curation activities enable data discovery and retrieval, maintain **data quality**, add value, and provide for re-use over time. This new field includes representation, archiving, authentication, management, preservation, retrieval, and use.“*

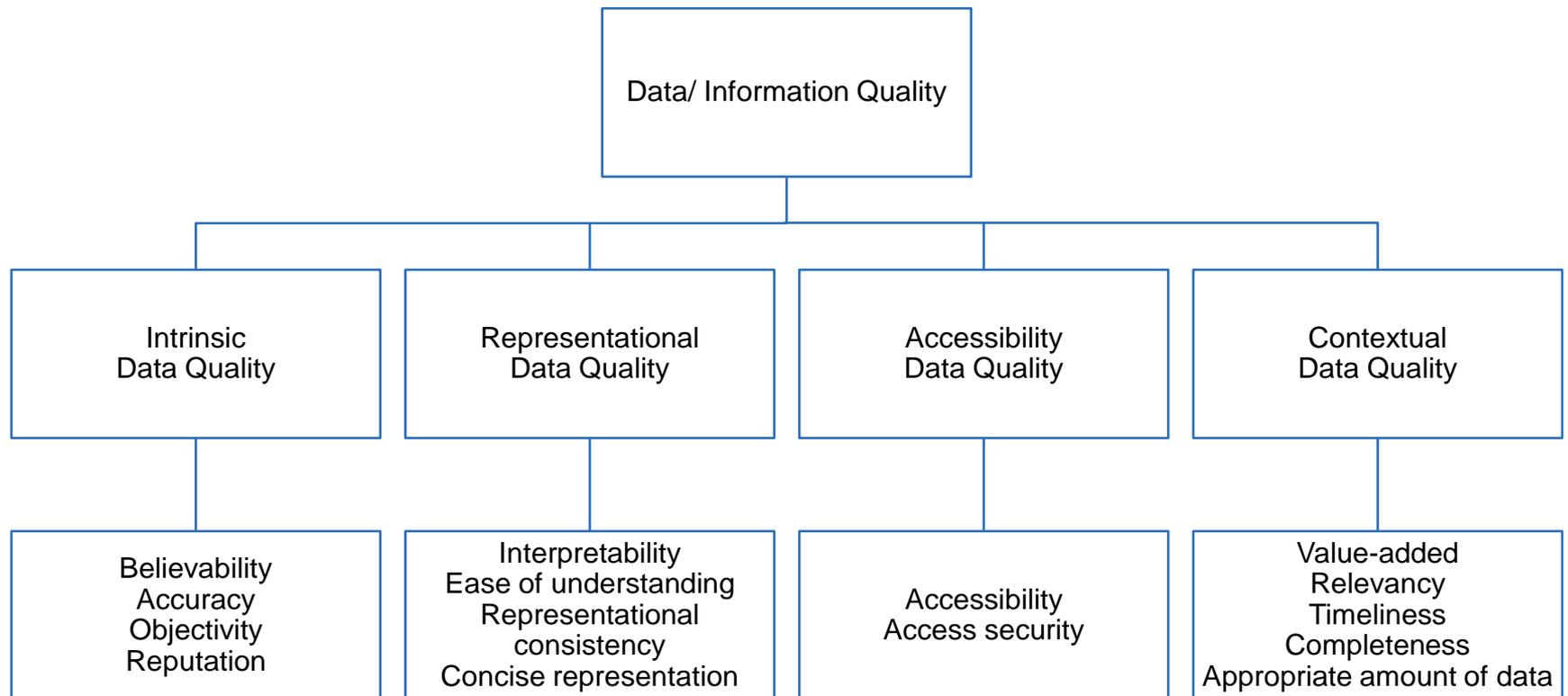
GSLIS-UIUC (Graduate School of Library and Information Science – University of Illinois at Urbana-Champaign), January 2011

Data Quality: 179 Definitionskriterien

Ability to be Joined With	Competitive Edge	Detailed Source	Form of Presentation	No lost Information	Reputation	Understandable
Ability to Download	Completeness	Dispersed	Format	Normality	Resolution of Graphics	Uniqueness
Ability to Identify Errors	Comprehensiveness	Distinguishable Updates	Friendliness	Novelty	Responsibility	Unorganized
Ability to Upload	Compatibility	Files			Retrievability	Up-to-Date
Acceptability	Concise	Dynamic	Generality	Objectivity	Revealing	Usable
Access by Competition	Conciseness			Optimality	Reviewability	Usefulness
Accessibility	Confidentiality	Ease of Access	Habit	Oderliness	Rigidity	User-Friendly
Accuracy	Conformity	Ease of Comparison	Historical Compatibility	Origin	Robustness	
Adaptability	Consistency	Ease of Correlation				Valid
Adequate Detail	Content	Ease of Data Exchange	Importance	Parsimony	Scope of Info	Value
Adequate Volume	Context	Ease of Maintenance	Inconsistencies	Partitionability	Secrecy	Variability
Aestheticism	Continuity	Ease of Retrieval	Integration	Past Experience	Security	Variety
Age	Convenience	Ease of Understanding	Integrity	Pedigree	Self-Correcting	Verifiable
Aggregatability	Correctness	Ease of Update	Interactive	Personalized	Semantic Interpretation	Volatility
Alterability	Corruption	Easy to Change	Interesting	Pertinent	Semantics	
Amount of Data	Cost	Easy to Question		Portability	Size	Well-Documented
Auditable	Cost of Accuracy	Efficiency	Level of Abstraction	Preciseness	Source	Well-Presented
Authority	Cost of Collection	Endurance	Level of Standardization	Precision	Specificity	
Availability	Creativity	Enlightening	Localized	Proprietary Nature	Stability	
	Critical	Ergonomic	Logically Connected	Purpose	Storage	
Believability	Current	Error-Free			Synchronization	
Breadth of Data	Customizability	Expandability	Manageability	Quantity		
Brevity		Expense	Manipulable		Time-Independence	
	Data Hierarchy	Extendibility	Measurable	Rationality	Timeliness	
Certified Data	Data Improves Efficiency	Extensibility	Medium	Redundancy	Traceable	
Clarity	Data Overload	Extent	Meets Requirements	Regularity of Format	Translatable	
Clarity of Origin	Definability		Minimality	Relevance	Transportability	
Clear Data Responsibility	Dependability	Finalization	Modularity	Reliability		
Compactness	Depth of Data	Flawlessness		Repetitive	Unambiguity	
Compatibility	Detail	Flexibility	Narrowly Defined	Reproducibility	Unbiased	

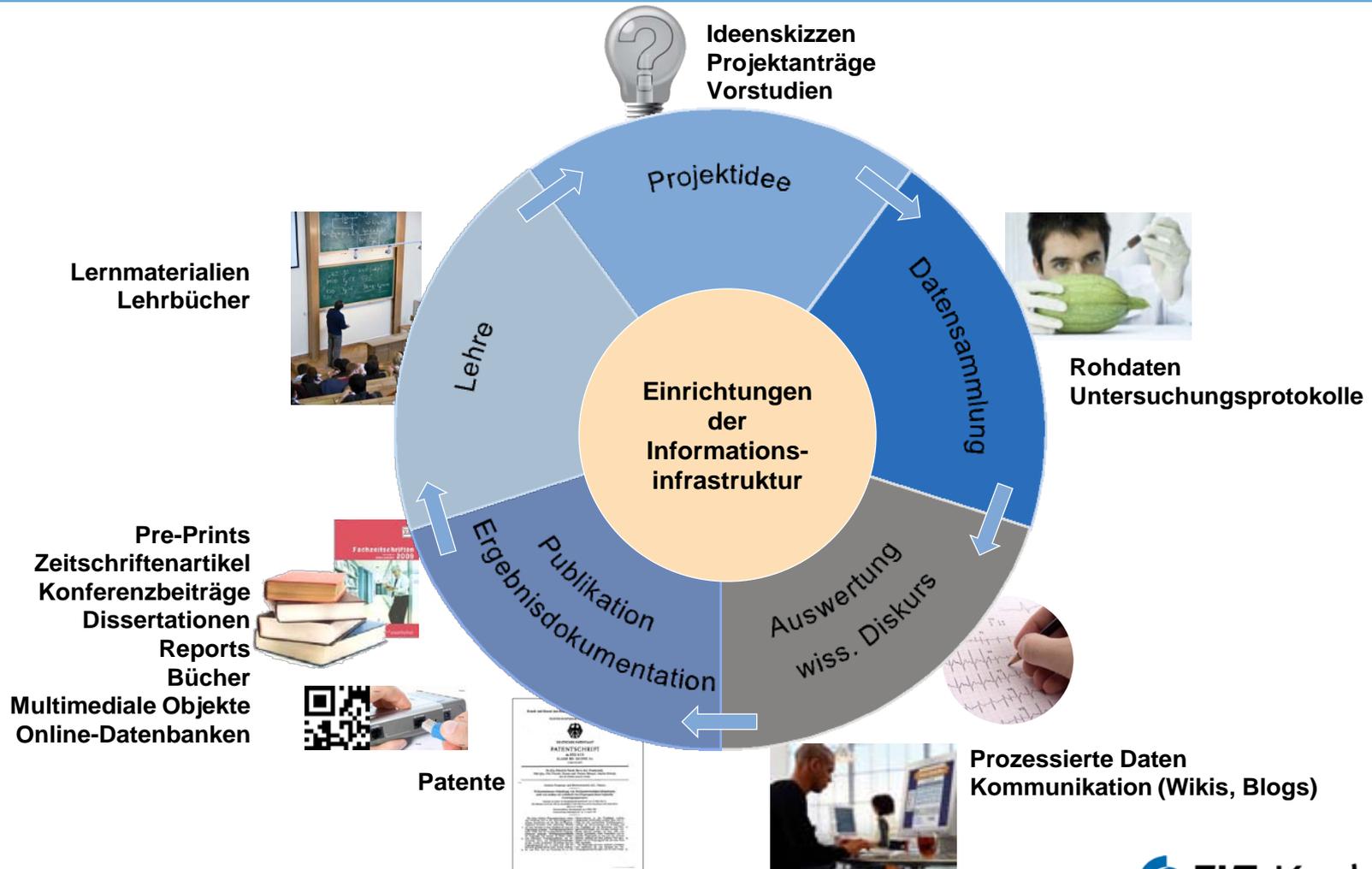
Beyond Accuracy: What Data Quality Means to Data Consumers, Richard Y. Wang and Diane M. Strong, Journal of Management Information Systems, Spring 1996, Vol. 12, No. 4, pp 5-34

Definition von Datenqualität: Vier maßgebliche Kategorien



nach: Beyond Accuracy: What Data Quality Means to Data Consumers, Richard Y. Wang and Diane M. Strong, Journal of Management Information Systems, Spring 1996, Vol. 12, No. 4

Die Qualitätskriterien gelten für **alle** Stufen des wissenschaftlichen Arbeitsprozesses



Die aktuellen Herausforderungen

Die großen gesellschaftlichen Themen der Zukunft stellen Wissenschaft, Wirtschaft und Politik vor völlig neue Herausforderungen.

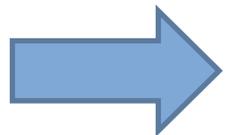
Starke Veränderungen hinsichtlich Organisationen, Strukturen, Innovationsprozessen und Wertschöpfungsketten

Innovative und leistungsstarke Informationsinfrastrukturen werden als Fundament für die Weiterentwicklung von Wissenschaft und Forschung benötigt.

... bedeuten für die Informationsinfrastruktur: Große Erweiterung des Aufgabenspektrums

Infrastrukturen müssen bereitgestellt *und integriert* werden für:

- Open Access- und Open Data-Publikationen
- Data Curation
 - Zertifizierte Datenzentren oder -agenturen
- Hosting von digitalem Content
 - Lizenzierte Verlagspublikationen
 - Primärdaten
- Langzeitverfügbarkeit und sichere Nachweisbarkeit
- Virtuelle Forschungsumgebungen



Neue Aufgaben, zusätzlich zu den klassischen

Der Begriff der „Versorgung“ bekommt neue Dimensionen

- ❑ Vom Content zum Context: Informationserschließung, -vernetzung und Semantic Enrichment; Verknüpfung von unterschiedlichen Informationstypen und -objekten
 - Abstract, Volltext, Foto, Video, Grafik, DOI
- ❑ Standardisierung und Interoperabilität
 - auch: standardisierte Schnittstellen und Workflowintegration
- ❑ Knowledge Discovery und User Interaction
 - Analysen von Mikro- ebenso wie von Makroinformation
- ❑ Privacy und Security
 - Datenschutz, Zugriffsschutz und -Sicherheit
- ❑ Qualität

Herausforderungen für die wissenschaftliche Informationsinfrastruktur

Informationsinfrastruktur gestern und heute: Auftrag und Ziel

Zwei Anwendungsszenarien: Realität und Vision

Sidestep: Daten

Aktuelle Herausforderungen und Konsequenzen

Politische Initiativen

Ausblick

Internationale Initiativen

OECD: „Principles and Guidelines for Access to Research Data from Public Funding“, 2007

NSF (National Science Foundation, USA): „Sustainable Digital Data Preservation and Access Partners“, DataNet-Programm 2008-2013

ESFRI (European Strategy Forum on Research Infrastructures): Roadmaps 2006, 2008, 2010

ERA (European Research Era): „Preparing Europe for a New Renaissance“, 2009

e-IRG-HLG (High Level Expert Group on Scientific Data): „Riding the Wave. How Europe can gain from the rising tide of scientific data“, 2010

Aktuelle nationale Initiativen – gestaltet durch die unmittelbar Beteiligten

Allianz-Schwerpunktinitiative Digitale Information (seit 2008)

Ziel: „Wissenschaftlerinnen und Wissenschaftler mit der bestmöglichen Informationsinfrastruktur auszustatten, die sie für ihre Forschung brauchen“

Kommission Zukunft der Informationsinfrastruktur (KII)

Ziel: Entwicklung eines Gesamtkonzepts für die Informationsinfrastruktur in Deutschland im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK)

Auftrag der GWK (Auszug aus dem Protokoll der Ausschusssitzung vom 28./29.09.2009)

„Der Ausschuss nimmt das "Rahmenkonzept für die Fachinformationsinfrastruktur in Deutschland" zur Kenntnis. Er bittet die Leibniz-Gemeinschaft, bis spätestens Anfang 2011 auf dieser Grundlage und unter Einbeziehung von Vertretern der Nutzer ein Gesamtkonzept für die Informationsinfrastruktur in Deutschland zu entwickeln.

Das Konzept sollte Empfehlungen enthalten, die den Rahmen der Finanzierungsmöglichkeiten berücksichtigen und auch mögliche Einsparungen durch Synergien umfassen.“

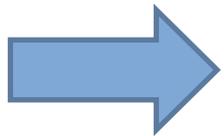
„Er bittet den Wissenschaftsrat, unter Berücksichtigung der dort bereits laufenden Arbeiten und unter Einbeziehung des von der WGL zu erarbeitenden Konzepts Empfehlungen zur Weiterentwicklung des Gesamtsystems der Fachinformationsinfrastruktur bis zum Jahr 2020 (einschließlich der Frage von Synergieeffekten) zu erarbeiten und in diesem Zusammenhang auch zum Konzept der WGL Stellung zu nehmen.“

KII – Die Ziele des Gesamtkonzepts

Optimierte „Landschaft“ der Informationsinfrastruktur

Definition der notwendigen Rahmenbedingungen (Strukturen, Prozesse, Koordination)

Aufzeigen von Synergien, Arbeitsteilung, Kooperationen, Ressourcenfreisetzung bzw. -bedarf



Handlungsempfehlung an die GWK (Vorlage: Mai 2011)

KII – Die Themen

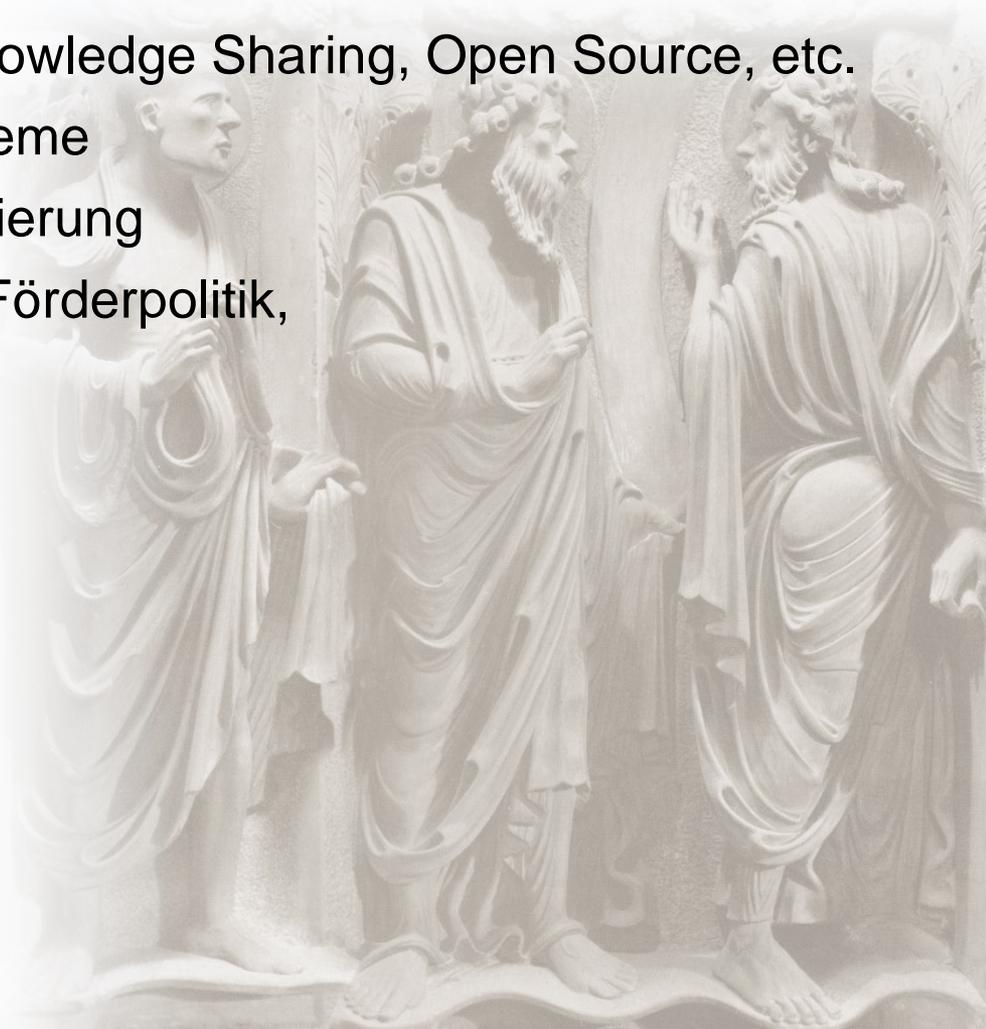
1. Lizenzierung
2. Hosting / Langzeitarchivierung
3. Nichttextuelle Materialien
4. Retrodigitalisierung / kulturelles Erbe
5. Virtuelle Forschungsumgebungen
6. Open Access
7. Forschungsdaten
8. Informationskompetenz / Ausbildung

KII – Die Dimension

- ❑ 6 von 8 Themen werden parallel durch die AGs der KII und die Allianz-AGs behandelt. Die themengleichen AGs treffen sich möglichst zum gleichen Termin
- ❑ Ca. 50 Treffen der KII-Gremien (Plenum, SC, AGs) finden zwischen Dez. 2009 bis Jan. 2011 statt
- ❑ Ca. 130 Personen sind in den gesamten Prozess involviert
- ❑ 54 nationale Institutionen wirken in der Kommission mit
- ❑ 6 der 10 Allianz-Wissenschaftsorganisationen sind in den Prozess integriert
- ❑ Einbezogen sind die im Börsenverein zusammengeschlossenen wissenschaftlichen Verlage sowie Nutzer-Vertreter aus der forschenden Industrie (Expertenforum)
- ❑ Kontakt zu internationalen Verlagen (STM) ist hergestellt

Informationsinfrastruktur im Spannungsfeld von Wirtschaft, Wissenschaft und Politik

- ❑ Elsevier, Microsoft, Google, Facebook & Co.
- ❑ Open Access, Open Data, Knowledge Sharing, Open Source, etc.
- ❑ Rechte, Lizenzen, Anreizsysteme
- ❑ Policies, Governance, Finanzierung
- ❑ Nationale und internationale Förderpolitik, Funding bodies, etc.
- ❑ Neutralität
- ❑ Nachhaltigkeit, Kontinuität



Vielen Dank!





Diese Unterlagen sind ausschließlich zu Präsentationszwecken bestimmt.

Das Copyright liegt bei FIZ Karlsruhe.

Die Weitergabe und Verwendung ganz oder in Teilen bedarf der ausdrücklichen Zustimmung durch FIZ Karlsruhe GmbH.

© FIZ Karlsruhe 2011