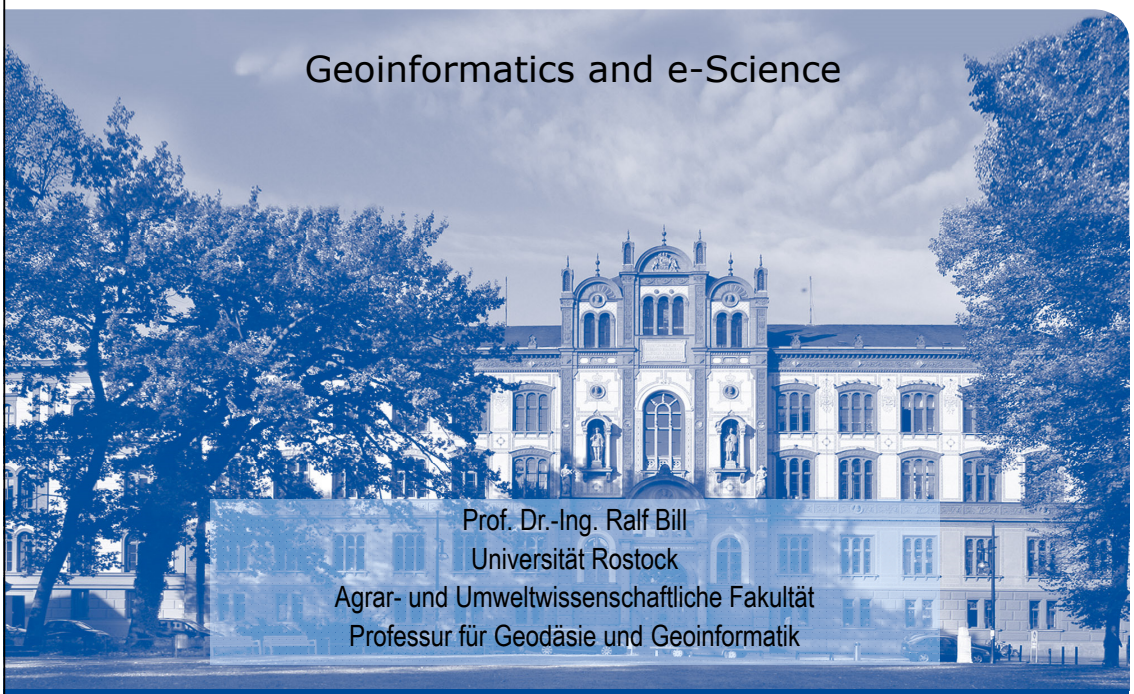
Universität Rostock Traditio et Innovatio



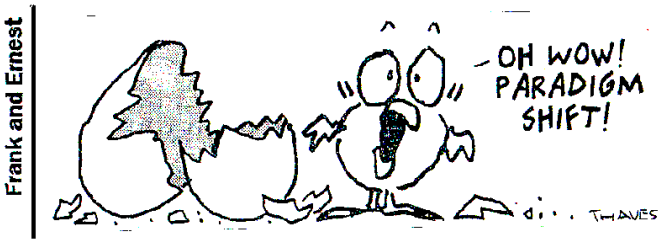
Geoinformatics and e-Science

Prof. Dr.-Ing. Ralf Bill
Universität Rostock
Agrar- und Umweltwissenschaftliche Fakultät
Professur für Geodäsie und Geoinformatik

Universität Rostock Traditio et Innovatio

Content

- Science paradigms and e-Science
- Terms and components of e-Science
- Geoinformatics and e-Science: Examples
- Conclusions



https://openparachute.files.wordpress.com/2008/06/paradigm-shift-cartoon.gif

Science paradigms and e-Science

Science paradigms through the ages

1. Science purely empirically and **observation**ally oriented
2. Science based on **theory** and **model** development
3. Science **simulating** complex phenomena using **information technology** capabilities
4. Science based on the **exploration** of abundantly available or collected **data** ("**data-driven science**")

=> precondition for such a paradigmatic understood "data-driven science" is a **systematic management**, **in the sense of infrastructure, regulations** and - what is often forgotten - **additional human resources.**

Source: HEY/TANSLEY/ TOLLE (2009), GRAY (2005, 2009)

e-Science: science of the future

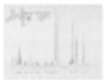
- e-Science: **Network-based science** or "**digitally enhanced science**" or **Data Driven Science** => combines theory, experiment, simulation with data
- The coinage of modern knowledge and information society with its manifold possibilities of **effective communication** and the **easy access to very large amounts of information** and **powerful computing technology** is a new challenge for science. The chance of achieving a qualitative and quantitative improvement of the scientific results with the new methods have increased significantly; in parallel, but also the difficulty of control of the **distributed, dynamic system components**.
- The focus of the upcoming work is the **integration of community-specific applications with a generic middleware services layer**. This requires the technical development and the organizational structure of a network and middleware infrastructure, with grid resources that can be (for example, computing power, data, information, application programs) offered on demand.

Source: Adapted from BMBF-Research Management 1/2005

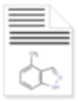
Problem – Linear value-added chain in science

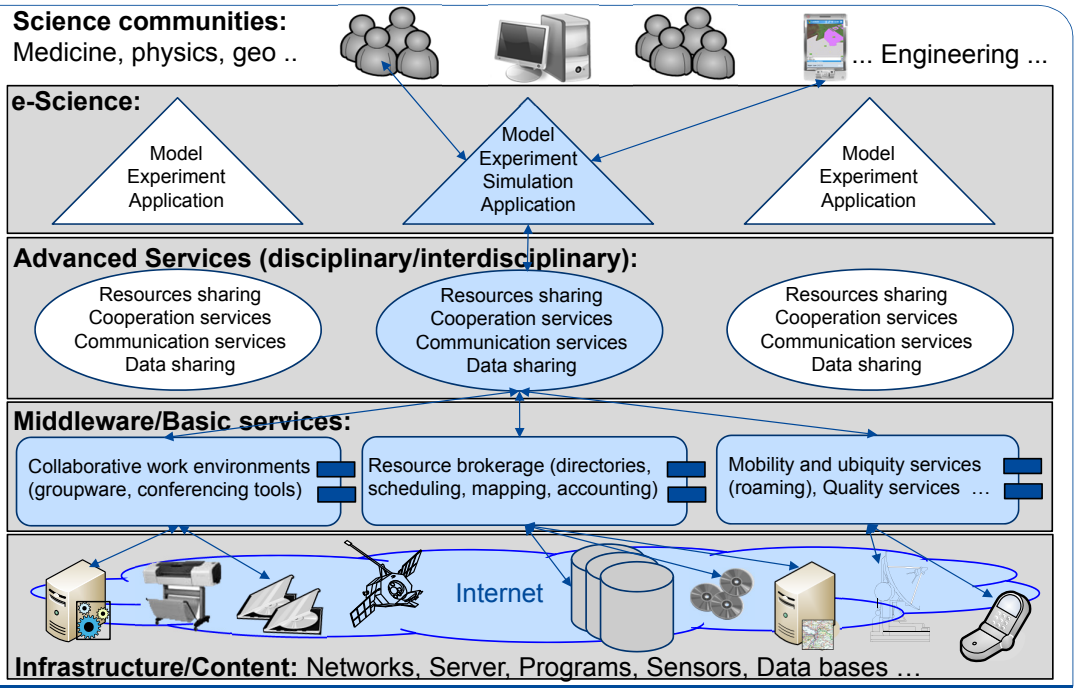
Data .. will be analysed, visualised, interpreted and result in .. get lost.

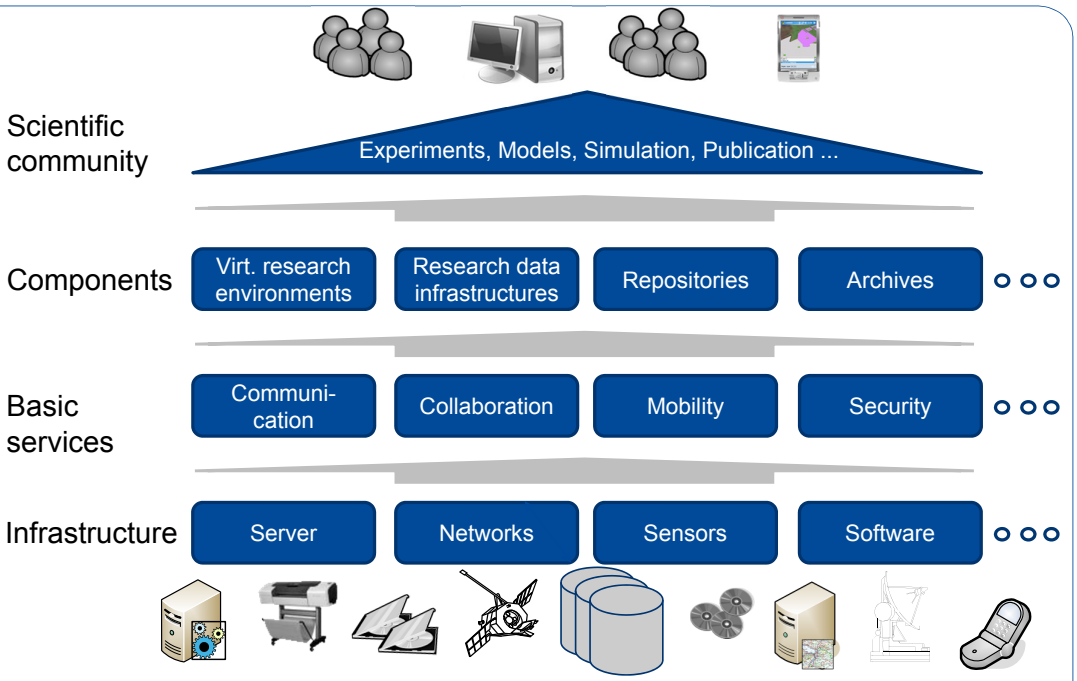
Information .. will be published and result in .. is reproducible.



Knowledge .. .. is accessible.

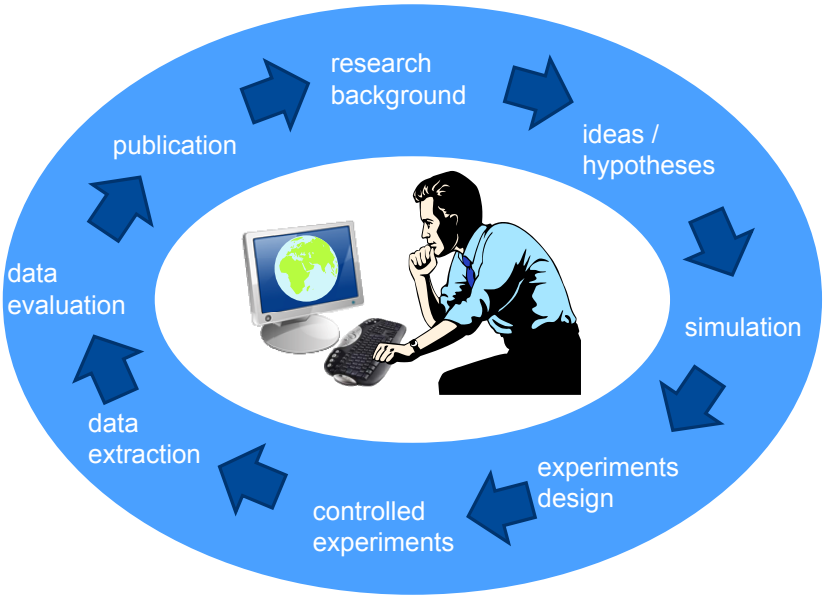



Science communities:
Medicine, physics, geo ..
... Engineering ...
e-Science:
Model
Experiment
Application
Model
Experiment
Simulation
Application
Model
Experiment
Application
Advanced Services (disciplinary/interdisciplinary):
Resources sharing
Cooperation services
Communication services
Data sharing
Resources sharing
Cooperation services
Communication services
Data sharing
Resources sharing
Cooperation services
Communication services
Data sharing
Middleware/Basic services:
Collaborative work environments (groupware, conferencing tools)
Resource brokerage (directories, scheduling, mapping, accounting)
Mobility and ubiquity services (roaming), Quality services …
Internet
Infrastructure/Content: Networks, Server, Programs, Sensors, Data bases …


Scientific community
Experiments, Models, Simulation, Publication ...
Components
Virt. research environments
Research data infrastructures
Repositories
Archives
Basic services
Communi-cation
Collaboration
Mobility
Security
Infrastructure
Server
Networks
Sensors
Software

e-Science

- Support the complete lifecycle of science



e-Science and additional competencies

- "Through e-science the insight grows that the value of the research is particularly in the data and therefore science must expand the scope of work on the **primary object, the research data**." (own translation of Büttner & Rümpel, 2011)
- ⇨ Challenges for the scientist with respect to a general, cross-curricular as well as a specific, subject-related use of modern digital information infrastructure

Media competence	Communi-cation competence	Information competence	Research competence
- Data - Documents - Multimedia - Models - Simulation - ...	- Information Services - Standards - Web technologies - Distributed Systems - ...	- Information modeling - Information retrieval - Data mining - Big data - ...	- Search mechanisms - Search strategies - Research data infrastructures - ...

Discipline specific expertise

Terms and components of e-Science

RIsources
The Research Infrastructure Portal
funded by DFG

(Primary) Research data

- Research data are **data that are generated, collected, used or assembled in the research process**. Based on these **scientific hypotheses, models or theories are formed**.
- Research data in the broadest sense are: **primary data, secondary analyzes, visualizations, models, analysis tools, collections of objects or products**.
- Primary research data provide **a valuable repository of information** that is collected with **high financial costs**. Depending on the subject area and method they are replicable or not based on repeatable observations, or measurements.
- In any case, the **data collected should be publicly accessible and freely available** after completion of the research. This is the essential precondition for:
  1. **data can be used again in the context of new issues** as well as ensure that
  2. in case of doubt as to the publication, the **data can be used for verification of published results**.

Source: DFG (2008), Jan Brase / Janna Neumann (2013), and others

(Primary) Research data

- Research data also **provide a lasting value per se** and are often the basis for research projects outside the original development context.
- For the storage and availability of primary research data **technical and organizational requirements** must be met, to be developed from the individual disciplines. With regard to offers for archiving and publication of research data it is only important that the data is in digital form or can be brought into this.
- The data generated in the science have to be acquired with **library methods** and **permanently made available for re-use** by future generations of researchers.

Different data views and states of research data

raw data

formatted data

evaluated data

public data

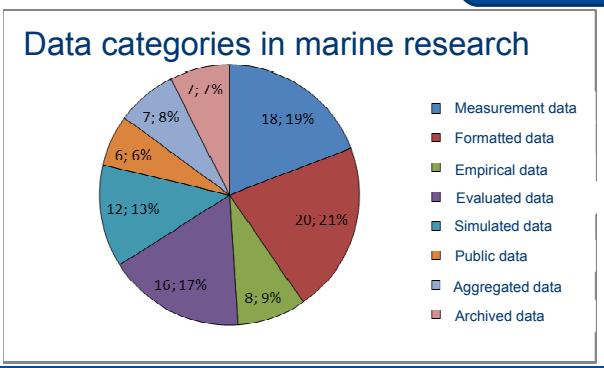
measurement data

empirical data

simulated data

aggregated data

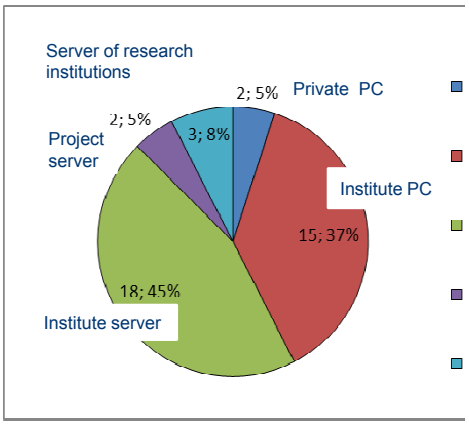
archived data



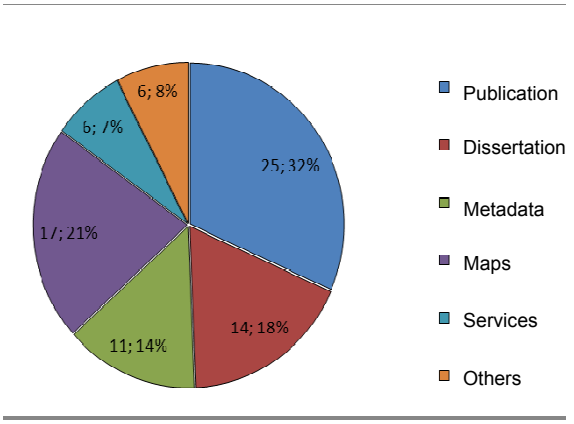
?

Research data: Locations and products

- Data storage in marine research



- Products in marine research

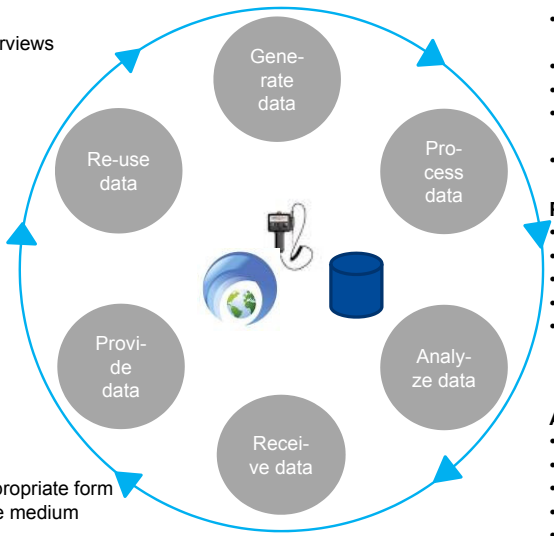


Research data lifecycle

Re-use data
- Pursue research
- New research questions
- Undertake research overviews
- Subscribe findings
- Train and learn

Provide data
- Distribute data
- Share data
- Control access
- Establish copyrights
- Advertise data

Receive data
- Convert data into an appropriate form
- Transfer data on suitable medium
- Backup and store data
- Create and document metadata
- Archive data



Generate data
- Design research
- Plan data management (Formats, storage, etc.)
- Plan data sharing
- Local existing data
- Collect data (experiment, observe, simulate)
- Capture metadata

Process data
- Digitize data, transform
- Check data, validate, clean
- Anonymize data where necessary
- Describe data
- Organize and store data

Analyze data
- Interpret data
- Derived data
- Produce research results
- Write publications
- Prepare data for preservation

Source: after http://data-archive.ac.uk/create-manage/life-cycle

Research Infrastructures

- Research infrastructures are extensive tools, resources or service facilities for research in all scientific fields, which are characterized by at least national importance for the respective scientific field and through a long life (typically over 10 years).
  - Major equipment or instruments used for research purposes
  - **Knowledge resources** of scientific research such as collections, archives, **structured information or systems for data processing**
  - **ICT infrastructures such as Grid, computing equipment, software and communication systems**
  - any other unique facilities being used for scientific research

⇒ **Research data infrastructures / scientific data infrastructures**

- "Within the totality of the research infrastructure research data infrastructure refers to infrastructure institutions, **providing research data for secondary analysis by the researchers for free or low cost**. The data provided are mostly aligned with scientific issues. "
  - **Interoperability of tools and research data** as well as their **long-term availability** and re-use
  - located at a single site or be distributed (centralized, distributed, or virtual)

Data storage and data dissemination

- A **data repository** (Latin repository = 'warehouse') is a managed place to store ordered documents that are publicly and widely available.
- Repository refers to the storage of data and documents using modern information and communication technologies and making them available on the Internet.

⇨ **Making data available**

- Examples of Research data repositories (www.re3data.org)
  - World Data Centres WDC-RSAT, WDC-Climate, Geophysics FIS ...

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

- In a **data archive** (Latin 'Archivum'), however, only selected information is managed.

⇨ Capacity
  - Long-term (eg., >> 10 years)
  - Archive formats (PDF/A, ...) instead of primary formats
  - Retrieval: if required

⇨ **Long-term storage**

- **Research Data Archive**
  - Selection criteria must be comprehensible
  - Archive formats, open standards (formats, interfaces)
  - Permanent evaluation option? (Primary data OK?)

Virtual research environments (VRE)

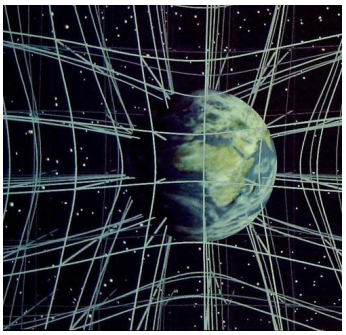
- Virtual research environments provide all the **necessary tools, data, information and services available** so that the researcher is detached from resources and access problems (memory, CPU time, log-in etc.).
- The scientist of the future **uses, independend of location and time, a virtual environment** and finds programs, research data and secondary sources (such as **publications, databases and services**), which he needs for his current research work.
- He engages with his **colleagues all over the world**, immediately shares new information, analyzes it, adds its new findings and places them on the fly to his colleagues for discussion and further use.

Source: Definition AG "Virtual Research Environments" of the Covenant initiative "Digital Information" (http://www.allianzinitiative.de/de/handlungsfelder/virtualresearchenvironments/).
Based on: Neuroth; Aschenbrenner; Lohmeier; „e-Humanities –eine virtuelle Forschungsumgebung für die Geistes-, Kultur-und Sozialwissenschaften"; BIBLIOTHEK Forschung und Praxis. Band 31, Heft 3, Seiten 272–279, DOI: 10.1515/BFUP.2007.272, /December/2007

DFG Internetportal RIsources http://risources.dfg.de/

- Digital archives and repositories (26 hits in Earth Sciences):
  - Aktuelle Wetterwerte deutscher Stationen (WETTER)
  - Animal Tracking Data (Movebank)
  - Biodiversitäts-Exploratorien (BioDiv-Exploratories)
  - Collaborative Climate Community Data and Processing Grid (C3Grid)
  - …
  - **Data Publisher for Earth & Environmental Science (PANGAEA)**
  - Deutscher Wetterdienst - Klimadaten (KlimaD)
  - ...
  - **GLUES Geodateninfrastruktur (GLUES GDI)**
  - ...
  - Historische hydrographische Daten des BSH (ICDC)
  - ...
  - **Virtuelles Kulturlandschaftslaboratorium (VKLandLab)**
  - ...
  - World Data Center for Climate (WDCC)
  - World Data Center for Remote Sensing of the Atmosphere (WDC-RSAT)

RIsources The Research Infrastructure Portal funded by DFG



http://ffden-2.phys.uaf.edu/webproj/211_fall_2014/Chris_Bon/chris_bon/spacetime.html

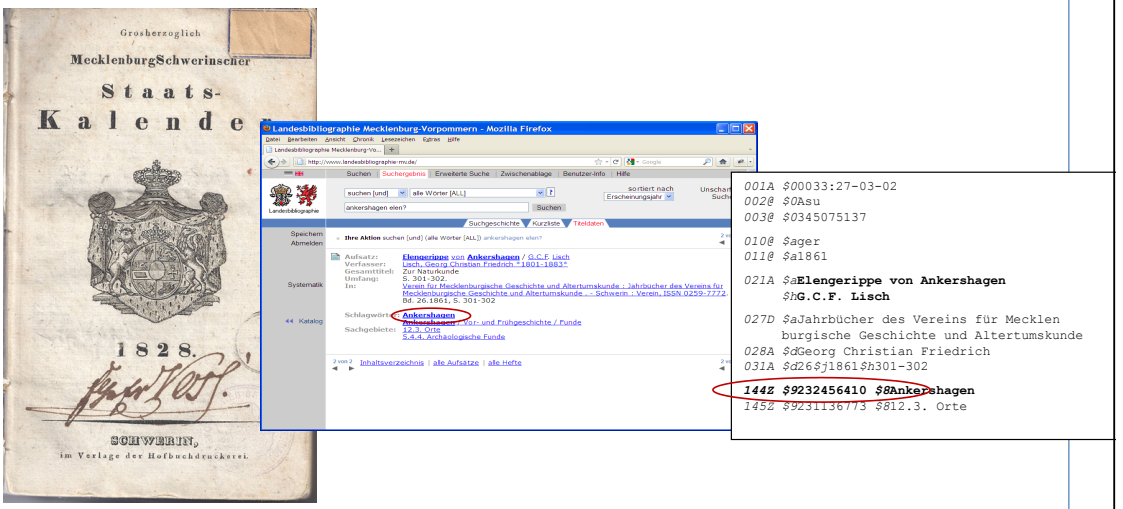
Geoinformatics and e-Science

„Whatever occurs, occurs in space and time.“ (HILL, 2006).

1. Geographical names

Geographical names = indirect spatial references

- Published knowledge in libraries
- Keyword: Geographical Names



Indirect spatial reference - Georeferenced site names

⇨ **Geographical name directories resp. gazetteers are in a classical sense simply lists of toponyms in a given region, that should be made identifiable using additional information.**

- Solution approach:
  - Spatial reference has to be made available → Georeferencing
  - Coordinates as ***formal*** instead of geographical names as ***informal representations***

- **DFG Project "Virtual Map Forum 2.0":**

⇨ Extract place names and locations contained in the plane survey sheets in different time sections (1868-1945) for the area of the German Empire

⇨ Georeferencing these place names

⇨ Link with other existing directories by establishing a service based register of historical place names

Source: Walter/Bill (2015)

Sources for geographical names

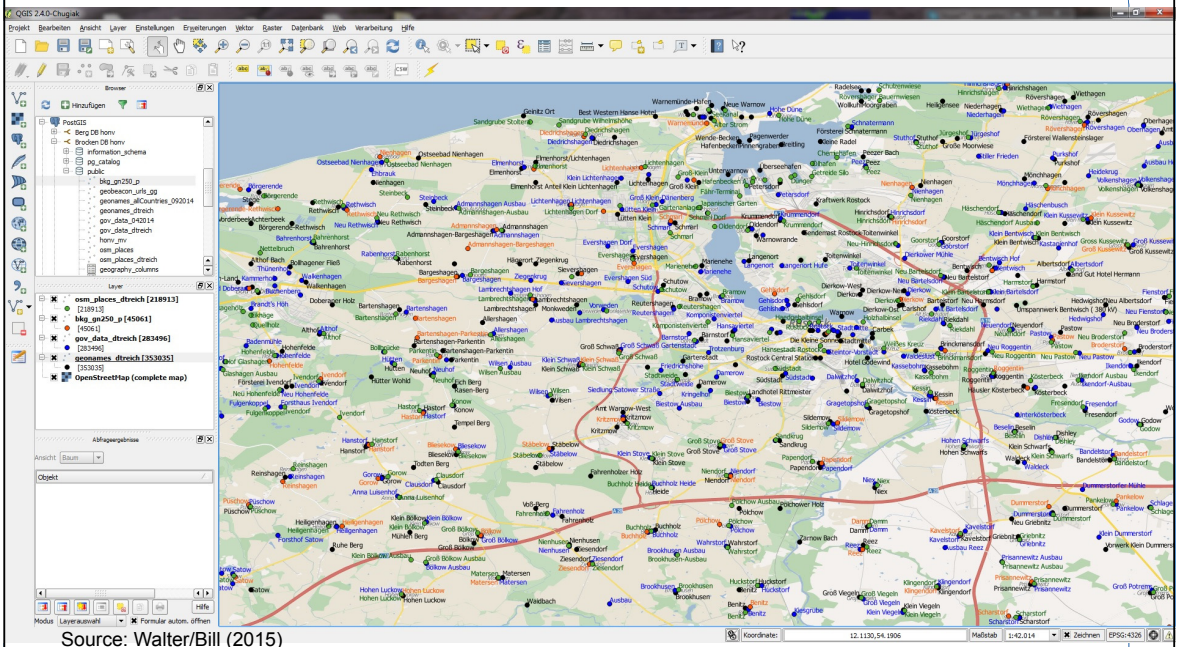
- Fusion of various data sources
  - GeoNames: http://www.geonames.org/
  - Genealogical site directory: http://gov.genealogy.net/search/index
  - OpenstreetMap: http://www.openstreetmap.org/
  - BKG: http://www.geodatenzentrum.de/geodaten/
  - GeoBeacon inkl. HONV-MV (VKLandLab-Project) http://139.30.132.26/beacon/search.html

- Number of site objects per source in the borders of the former german Empire (1871-1933)
  - Feature class filter e.g. OSM: city, town, village, suburb, hamlet, locality.

Geonames:	345.331
GOV:	283.496
OSM:	218.913
BKG:	45.061
GeoBeacon:	ca. 91.000
thereby: HONV-MV:	7.135

Source: Walter/Bill (2015)

Sample GN/GOV/OSM/BKG



Source: Walter/Bill (2015)

Geographical features spatially aggregated

- PostGIS data base, spatial aggregation (1.000m perimeter)
- Collecting additional site names one source after the other
- Contributions of site names per source after spatial aggregation
  - Geonames: 345.331
  - GOV: 19.587
  - OSM: 11.695
  - GeoBeacon/HONV-MV: ca. 1.200
  - BKG: 47

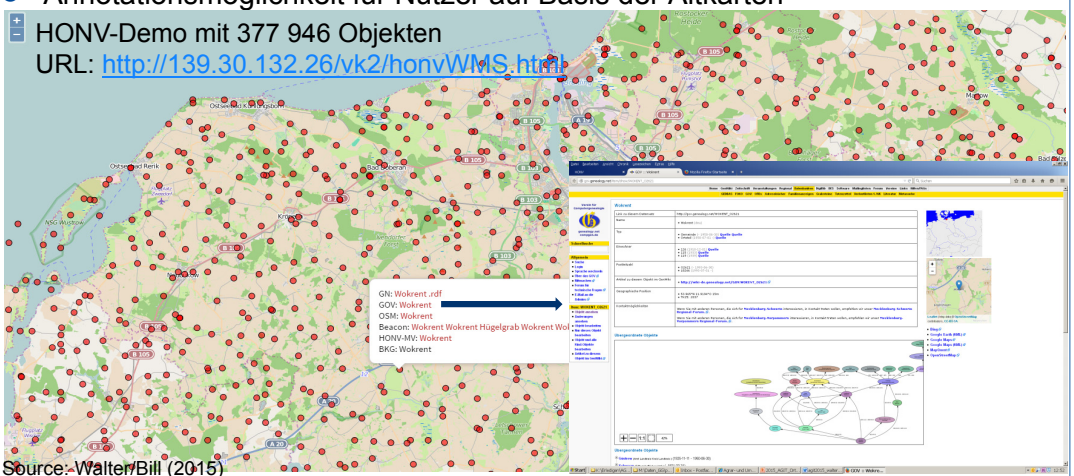
⇨ 377.946 site features for the former German Empire (currently available)

URL Demo Webclient : http://139.30.132.26/vk2/honvWMS.html

Source: Walter/Bill (2015)

Demoversion Webclient

- Ausschnitt des gesammelten Ortsnamensbestandes
- Sammlung von URLs bzw. Weblinks mit Verweis auf den jeweiligen Datensatz in der Webpräsenz des Datenanbieters
- Bereitstellung über WebMapService (GetFeatureInfo)
- Annotationsmöglichkeit für Nutzer auf Basis der Altkarten



Source: Walter/Bill (2015)

2. Virtual research environment for cultural landscape research

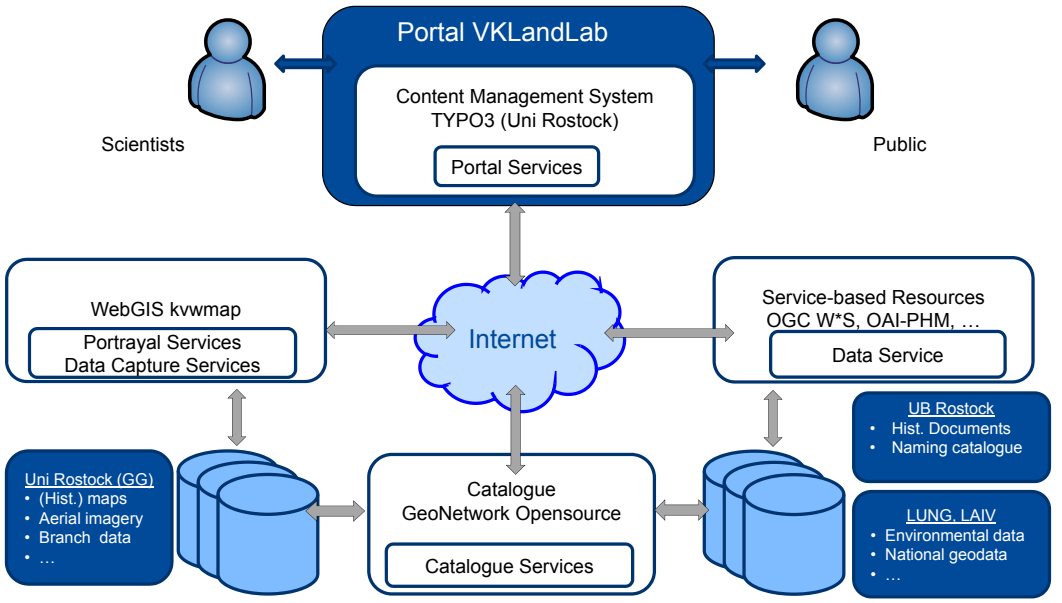
VKLandLab – Project goals

- **Technology**
  - Establishing a VRE for interdisciplinary cultural landscape research.
  - Integration of different well known components of data driven infrastructures such as InternetGIS, data bases, primary data repositories, authentification structures.
  - Enhancements with respect to collaborative elements such as wikis, blogs, project management, content management, data tagging etc.

- **Investigation area in space and time**
  - Focussing on the area Mecklenburg in the last 230 years.
  - Offering georeferenced old maps from 1786, younger aerial and satellite imagery of the last 40 years and recent geoinformation

- **Benefits for the interdisciplinary scientific community**
  - Allowing spatio-temporal research questions on various levels of scale in space (regional 1:200.000 to local 1:25.000) and time (240 years in 3 time slices, the last 40 years again in 3 time slices and recent data)
  - Supporting interdisciplinary collaboration of scientists in Rostock and elsewhere focussing on research work and not on technology

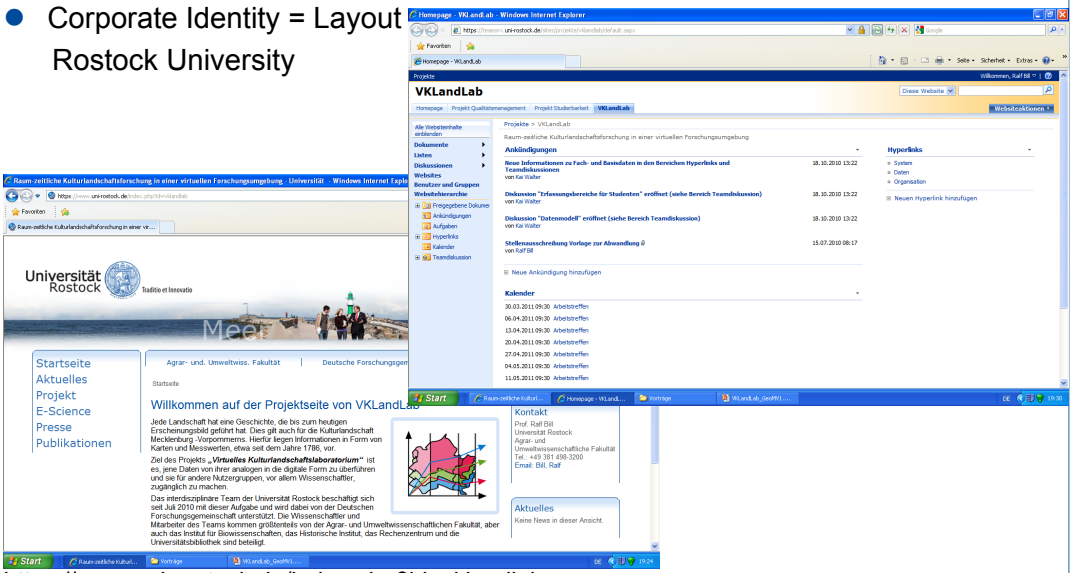
VKLandLab-Reference architecture (OGC-like)



Source: Bill [Ed.](2012)

Portal

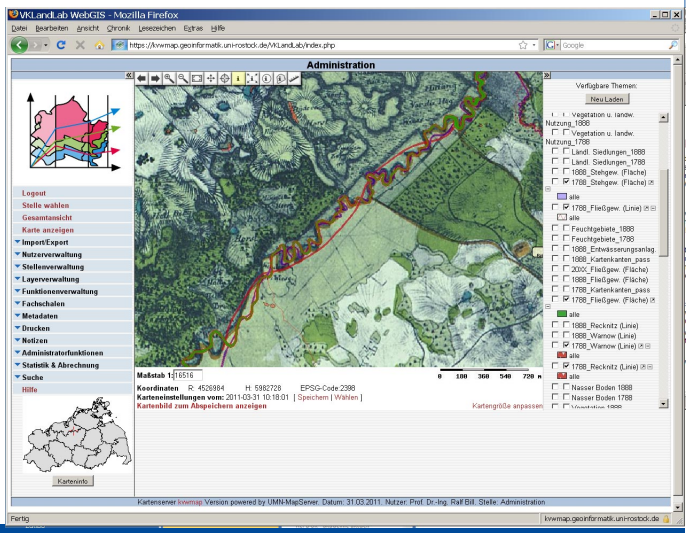
- Content Management System TYPO3
- Corporate Identity = Layout Rostock University
- Collaboration environment
- Sharepoint, Wiki



https://www.uni-rostock.de/index.php?id=vklandlab

Internet GIS kvwmap

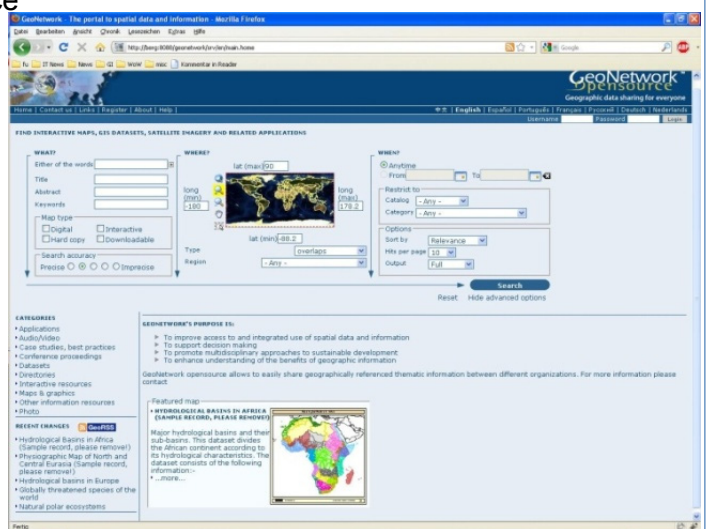
- OpenSource WebGIS-framework kvwmap to capture, store, analyse and present spatial information
- Data and portrayal services
- On top of UMN-MapServer
- PostgreSQL data base



Source: Bill [Ed.](2012)

Catalogue/Meta data

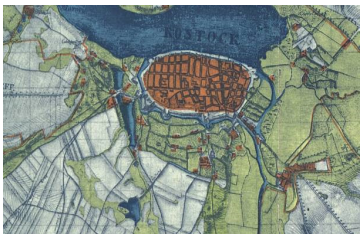
- Catalogue service: Search and find spatial data based query parameters such as data topic, -origin and -matter.
- Centralized meta information system according to ISO 19115/19139
- GeoNetwork OpenSource



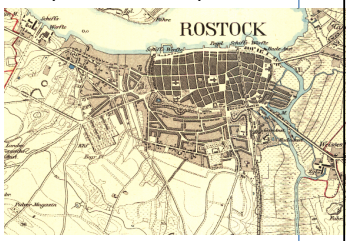
Source: Bill [Ed.](2012)

Spatial data infrastructure

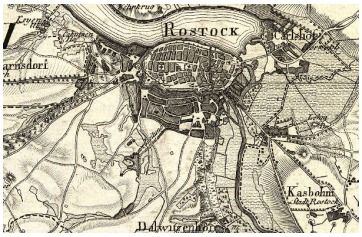
- Carl Friedrich von Wiebeking (1786-1788):
  - 1:24.000
  - 48 sheets
  - ~2 GB



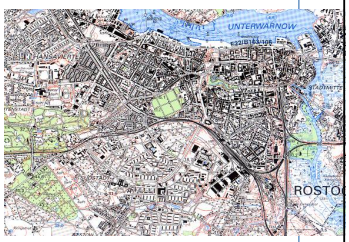
- Plane survey sheet (1877-1889)
  - 1:25.000
  - ~168 sheets
  - ~ 18GB



- Friedrich Wilhelm Karl von Schmettau (1788):
  - 1: 50.000
  - 16 sheets
  - ~2 GB

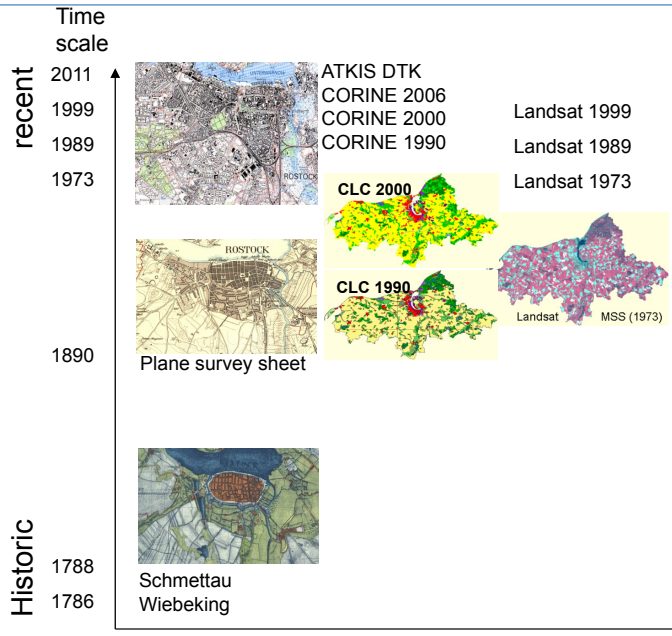


- ATKIS DTK 10 (2000) u.a.



Source: Bill [Ed.](2012)

Geodata sources

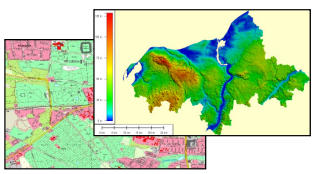


Map portals LaiV/LUNG Geobasic- & Environmental data:

- ATKIS DTK, DOP, DGM
- ATKIS DLM moorland, greenland, fallow land, acre land, water, forest, …
- Geology, soil overview, ground water, geotopes
- administrative borders

Uni Rostock Branch data:

- kettle holes, forest, water, vegetation, agriculture, wet areas/swamps, settlement areas, parks a. garden, alleys a. streets,…
- Hist. administrative borders



Source: Bill [Ed.](2012)

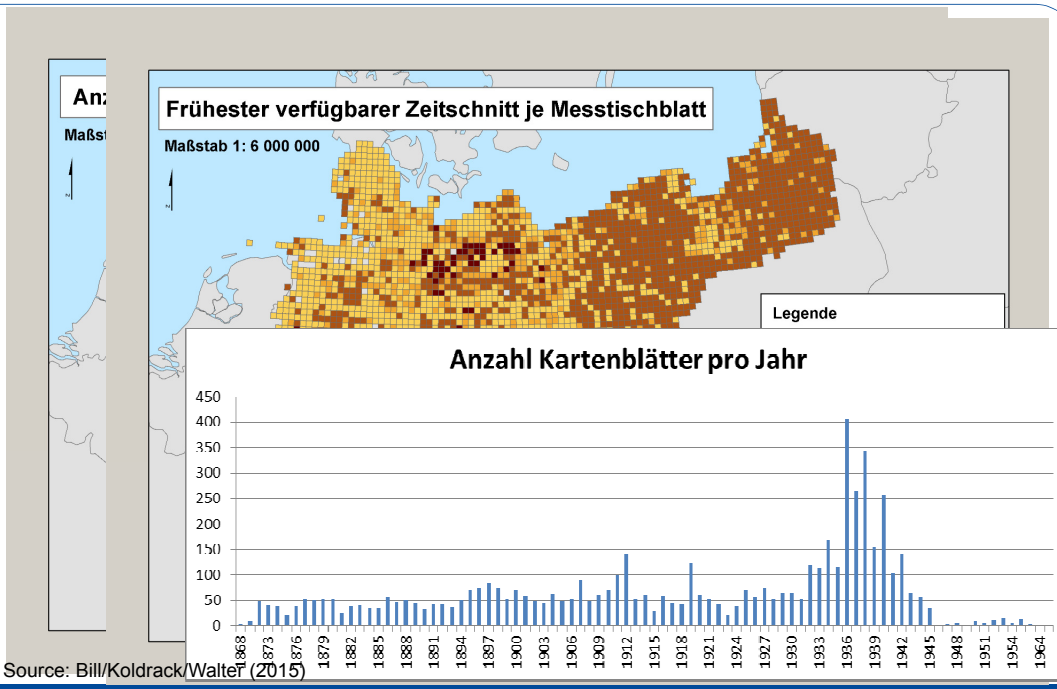
3. Virtual topographic map forum for the German Empire

Kartenforum „Deutsche Fotothek“

- Prussian land surveying - „Messtischblätter (MTB)“ plane survey sheets
  - Complete full area coverage with cartographic map sheets in the range of the former German Empire (Deutsches Reich 1871-1933)
  - More than 6.000 MTB in the scale 1 : 25.000, from 1868 til 1965
  - 674 MTB in the scale 1 : 100.000

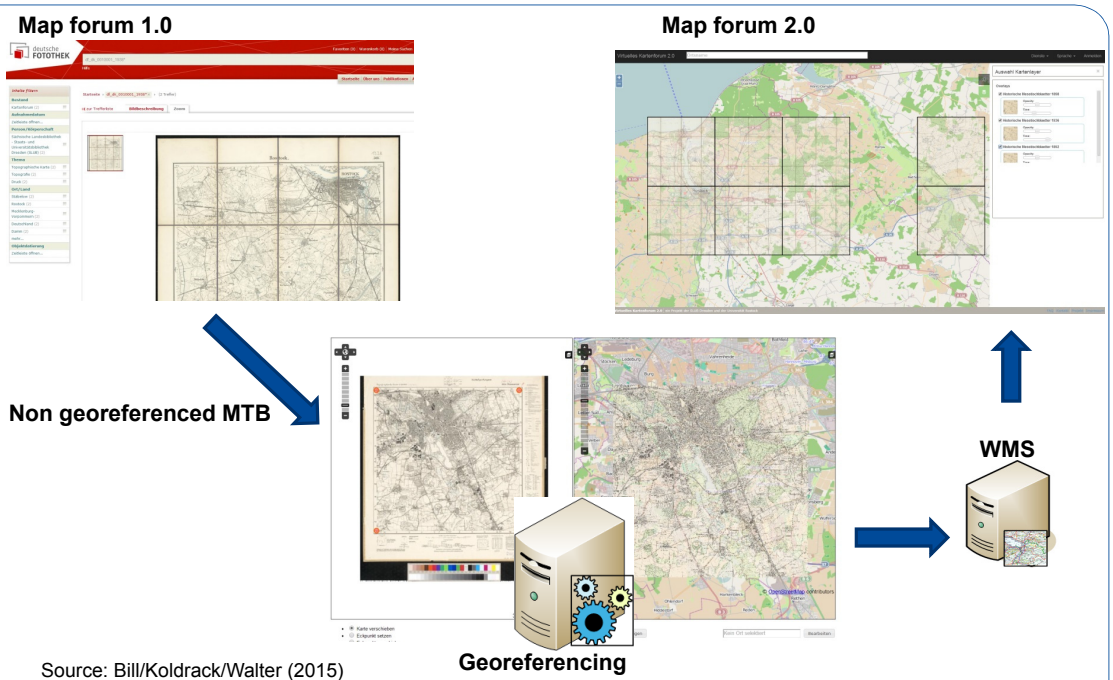
Source: SLUB Dresden

Area coverage and time slots



Source: Bill/Koldrack/Walter (2015)

Kartenforum 1.0 => Virtuelles Kartenforum 2.0



Source: Bill/Koldrack/Walter (2015)

Georeferencng Geocoding

- **„Georeferenzierung** kann als räumliches Metakonzept betrachtet werden, womit räumliche Referenzinformation einem Datensatz mitgegeben wird. Hierzu gehören die **Wahl des geodätischen Bezugssystems** und die **Festlegung der Passpunkte**, die zur Überführung verwendet werden sollen. Den eigentlichen Überführungsschritt leistet dann die Geokodierung."
  - ⇨ **1.1** EPSG:4314 = Deutsches Hauptdreiecksnetz (Geographische Koordinaten mit Datum Potsdam) und Ellipsoid von Bessel
  - ⇨ **1.2** Messung der vier Gitterpunkte am Kartenrand

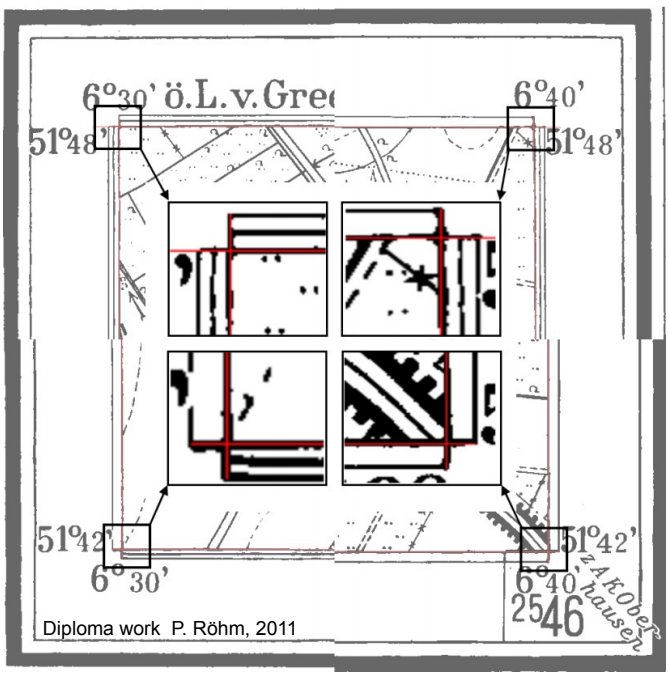
1

- **„Geokodierung** behandelt den tatsächlichen **Transformationsschritt**, der notwendig ist, um Daten verschiedenartiger Georeferenzierung in ein gewünschtes Referenzsystem umzurechnen. Bei Rasterdaten schließt dies z.B. das **Resampling** der Bildelemente mit ein. ..."
  - ⇨ **2.1** Ebene überbestimmte Transformationen mit 4 bis 6 Parametern
  - ⇨ **2.2** Resampling mittels „nearest neighbour".

2

Quelle: Bill/Zehner (2001): Lexikon der Geoinformatik bzw. GeoInformatikService http://www.geoinformatik.uni-rostock.de/

Georeferencing of MTBs and topographic maps



- POIs = Corners of the map (X,Y known)
- One person for 6.000 MTB: 10 Minutes per MTB → 75 working days

⇨ Crowdsourcing by non-experts

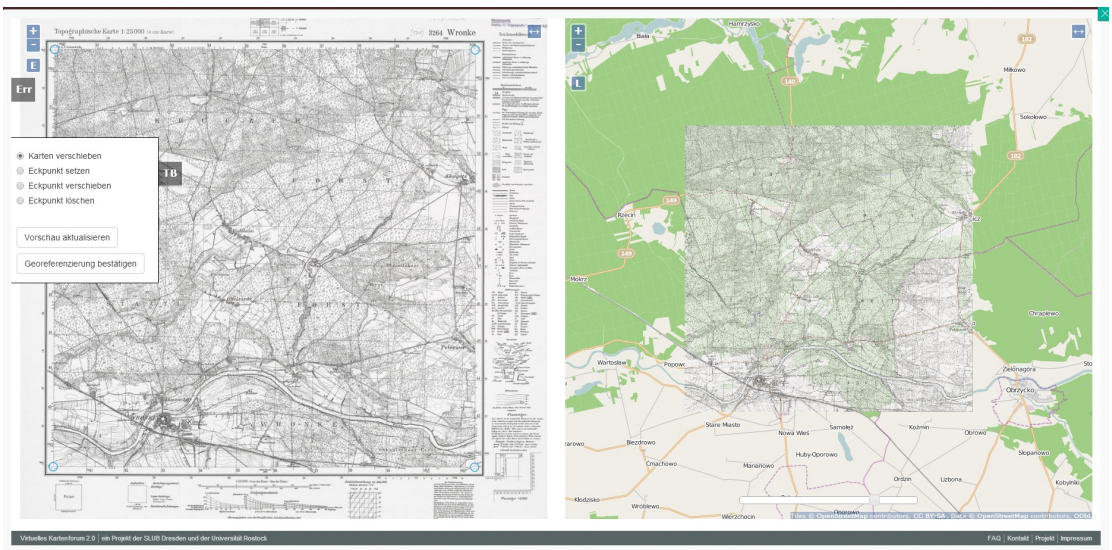
⇨ Automatic image processing

Georeferencing historical maps

- Heterogeneous map layout
- Further problems

Source: Bill/Koldrack/Walter (2015)

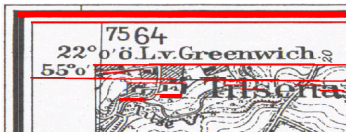
Georeferencing by Crowdsourcing



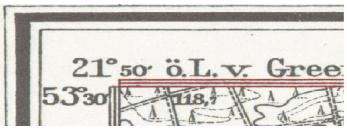
VKForum 2.0 Client for georeferencing
→ http://kartenforum.slub-dresden.de/vkviewer/

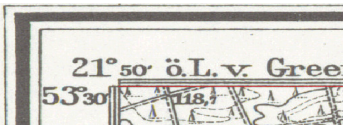
Georeferencing by image processing

- Determining the image content
- Hough-Transformation/morphologic operators

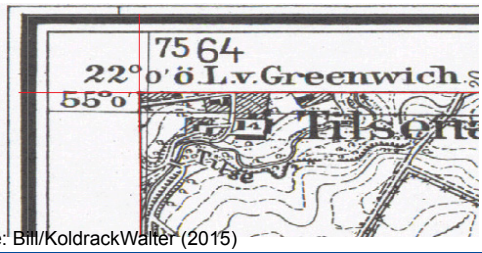


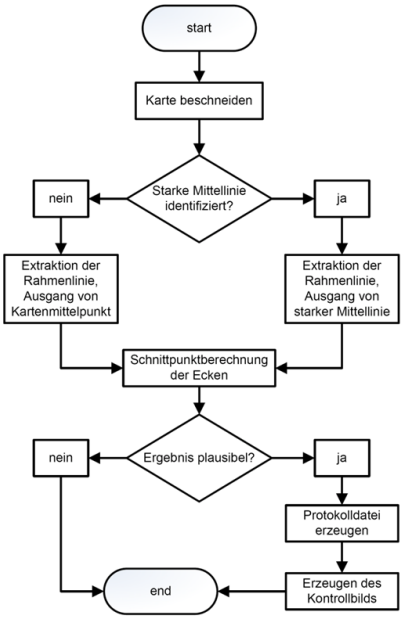
- Extraction of the inner frame lines





- Intersection of the framelines

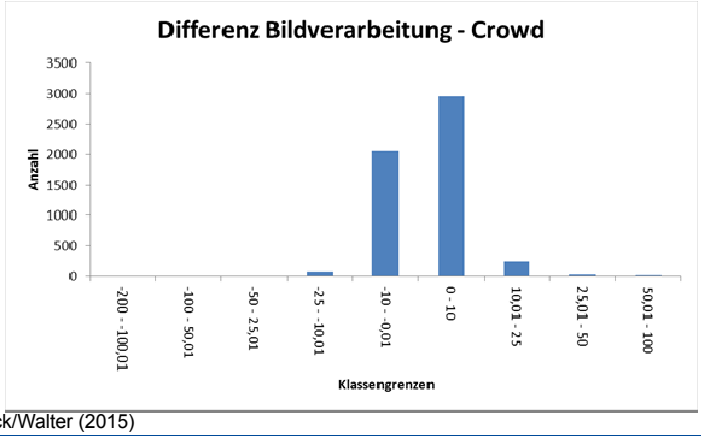




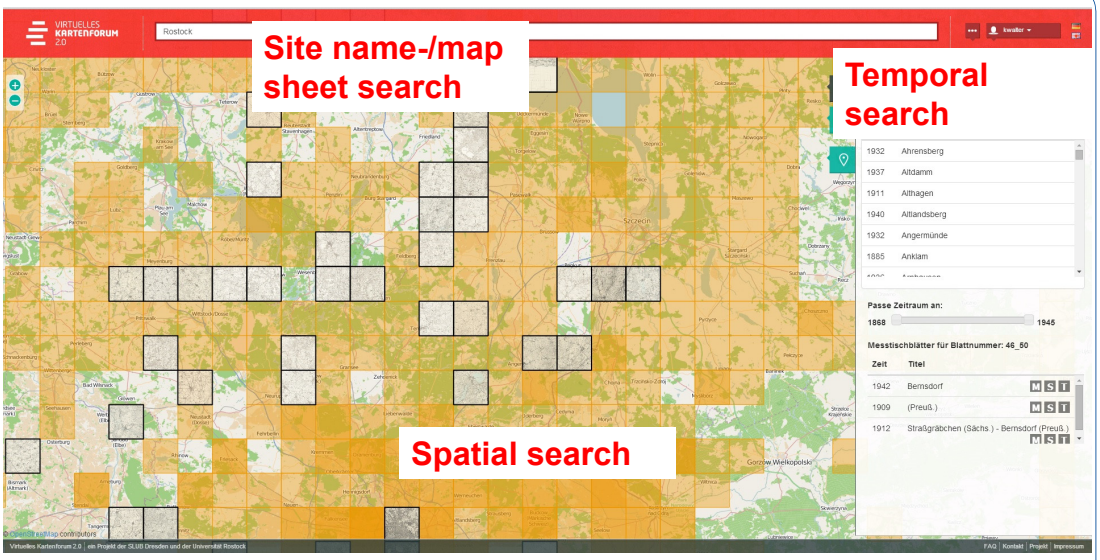
Source: Bill/KoldrackWalter (2015)

Comparing Crowd versus image processing

- In total 5.395 map sheets have been measured by both methods.
- In 40% of the georeferencing the computer achieves a better coordinate accuracy, in 60% the human crowd.
- Differences between both georeferencing approaches lies in 93% of all cases under +- 10m.



Source: Bill/Koldrack/Walter (2015)



→ http://kartenforum.slub-dresden.de/vkviewer/

4. GLUES SDI for Scientific Environmental Data

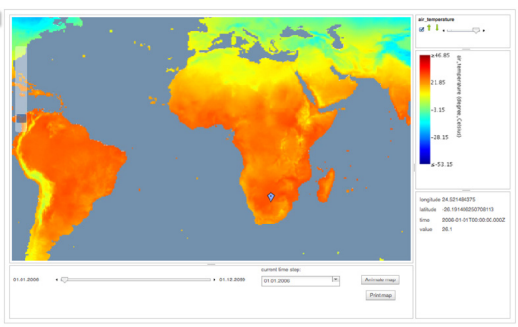
- Global Assessment of Land Use Dynamics on Greenhouse Gas Emissions and Ecosystem Services

- Objectives:
  - Publication and sharing of model data, analysis results and basic scenarios of the involved research groups.
  - Seamless integration of existing data sources through the GDI, for example, for the calculation of scientific models or comparative analysis.
  - Stakeholders of different areas are supported by spatial search and analysis tools to find research results and to understand and to use them for their own planning and management activities.

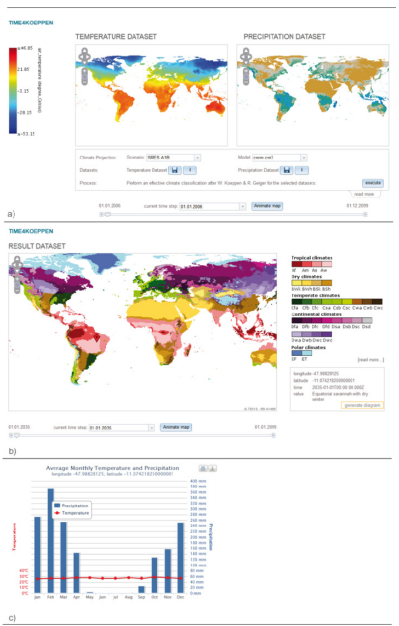
Source: Mäs/Henzen/Müller/Bernard (2014)

GLUES – spatio-temporal processing functionality

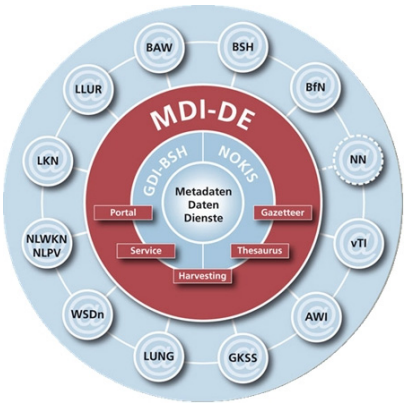
- Time series analysis



- Interactive classification

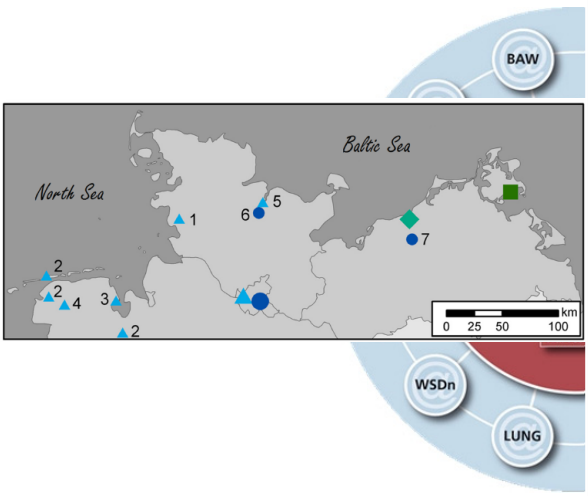


Source: Mäs/Henzen/Müller/Bernard (2014)



4. (Marine) spatial data infrastructures

Marine Data Infrastructure (MDI-DE)



Legend

SP1: Coastal engineering and coastal water protection

▲ *Principal applicant*
Federal Waterways Engineering and Research Institute

▲ *Project participants*
1 Authority for coastal protection, national parks and marine protection in Schleswig-Holstein
2 Authority for water management, coast protection and nature conservation in Lower Saxony
3 National Park Office of the Lower Saxony Wadden Sea
4 Federal Administration of Waterways and Navigation – Directorate Northwest
5 Federal Administration of Waterways and Navigation – Directorate North

SP2: Protection of the marine environment

● *Principal applicant*
Federal Maritime and Hydrographic Agency

● *Project participants*
6 State office for agriculture, environment and rural areas
7 State office for environment, conservation and geology

SP3: Maritime conservation

■ Federal Agency for Nature Conservation

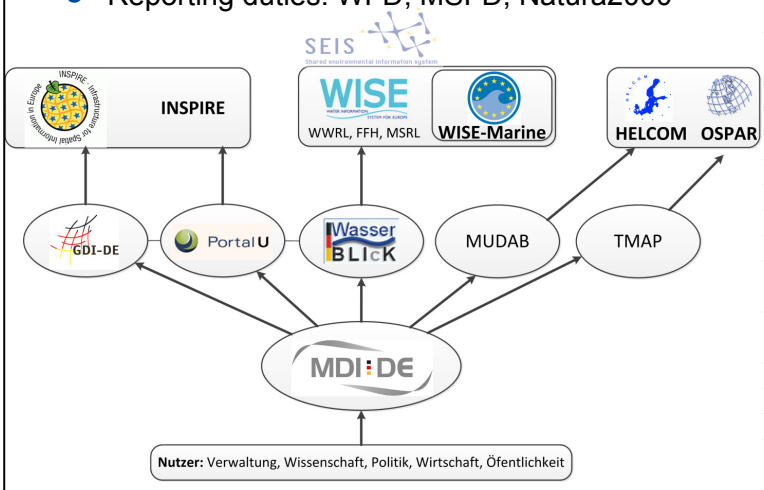
SP4: Scientific accompanying research

◆ Professorship for Geodesy and Geoinformatics, University of Rostock

Source: Die Küste (2014): Volume 82

MDI-DE
Information sharing

- INSPIRE: Interoperability of spatial data and services, meta data, data access, monitoring
- Reporting duties: WFD, MSFD, Natura2000

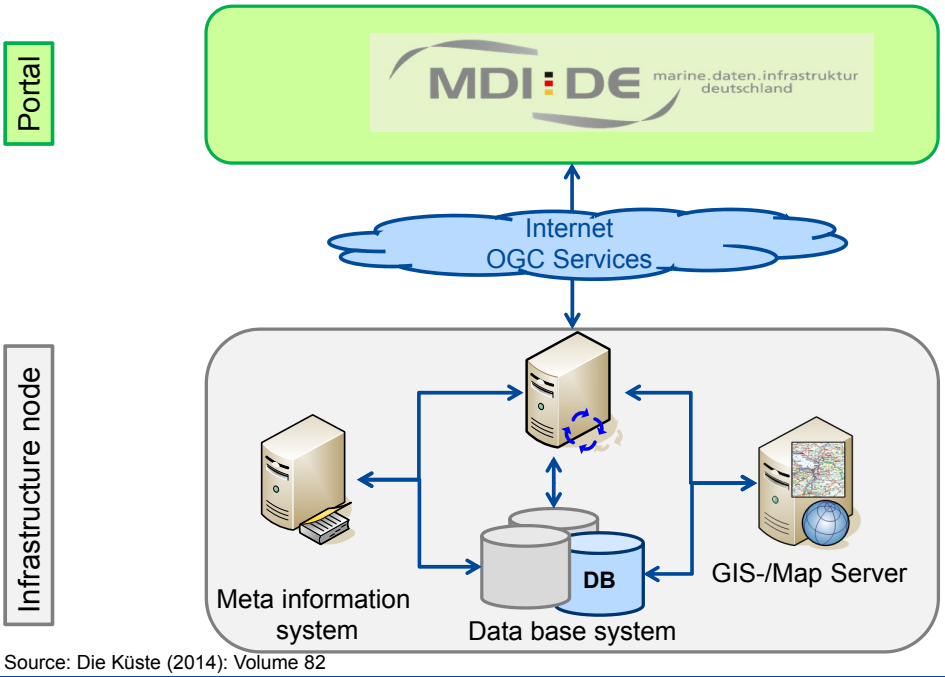


Data flow within information networks and reporting

INSPIRE: Infrastructure for Spatial Information in the European Community
SEIS: Shared Environmental Information System
WISE: Water Information System for Europe
WRRL: Wasserrahmenrichtlinie
SoE: State of Environment
FFH: Fauna-Flora-Habitat-Richtlinie
HWRL: Hochwasserrichtlinie
MSRL: Meeresstrategie-Rahmenrichtlinie
OSPAR: **Os**lo-**Par**is-Übereinkommen zum Schutz der Meeresumwelt des Nordostatlantiks
HELCOM: Helsinki Kommission zum Schutz der Meeresumwelt des Ostseeraums
ICES: International Council for the Exploration of the Sea
GDI-DE: Geodateninfrastruktur Deutschland
PortalU: Umweltportal Deutschland
WasserBLIcK: **B**und–Länder–**I**nformations- und **K**ommunikationsplattform
MUDAB: Meeresumwelt Datenbank
TMAP: Trilateral Monitoring and Assessment Program

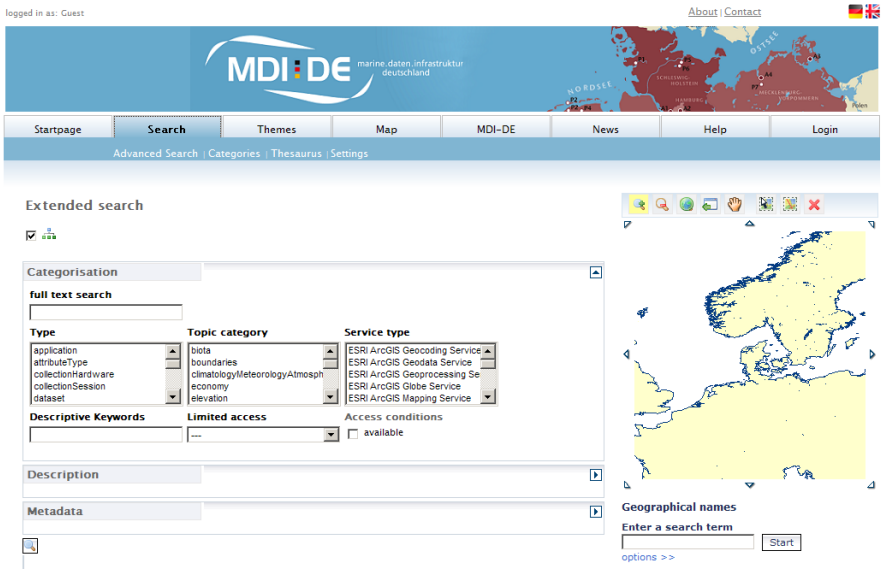
Source: Die Küste (2014): Volume 82

MDI-DE
Major components



Source: Die Küste (2014): Volume 82

Meta information system – Harvesting various ISN



www.mdi-de.org

MDI-DE - Infrastructure node

- MDI-DE is based on the linkage of spatially distributed infrastructure nodes
- Infrastructure node (ISN) within MDI-DE describes hard- and software of a local server architecture, being able to manage spatial data and metadata and to disseminate this by standardised services
- According to the „Publish-Find-Bind-Principle“ the individual components are interacting with each other using OGC conformal services

Data base:
- Oracle
- MicrosoftSQL Server
- PostgreSQL/ PostGIS
- ...

GIS-/Map-Server:
- GeoServer
- ArcGIS-Server
- UMN MapServer
- ...

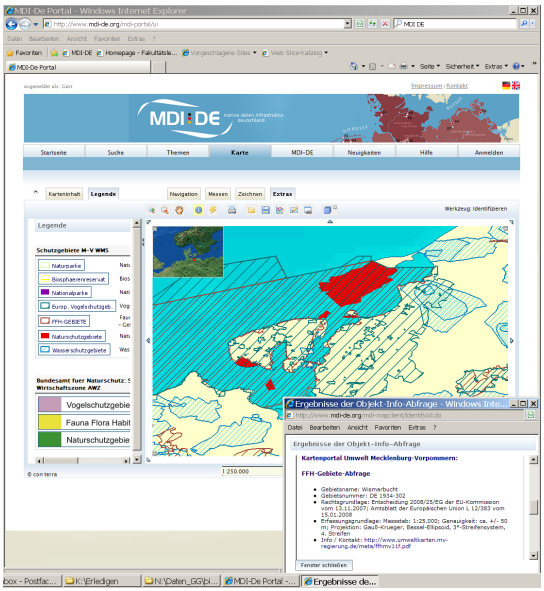
Metadata:
- NOKIS
- GeoNetwork
- InGrid/PortalU
- terracatalog
- ...

Client:
- Cadenza
- geoExt
- BfN-Viewer
- sdi.suite
- OpenLayers
- ...

Data storage → Data allocation → Data enquiry → Data usage

Source: Räder, M., Lübker, T., Prange, S., Binder, K., Schacht, C., Zühr, D., Kohlus, J. (2014): Infrastrukturknoten für Dienste - die räumlich verteilte Komponente der MDI-DE. Die Küste Volume 82

- Inquiry: Protected sites Wismarer Bucht
- Result comes from: Kartenportal M-V



Conclusion

http://www.gcu.ac.uk/library/SMILE/Communicating_information/conclusion_contents.html

Conclusion

- e-Science
  - New opportunities and new challenges for scientists
  - Research data infrastructures needed
- Geoinformatics
  - Valuable contribution to e-Science
  - Spatial reference is representing an ordering criterion
    - Direct and indirect georeferencing
  - Linking various resources
  - Web-based OGC reference architecture as basis for research data infrastructures
  - Spatial functionality offers analytic tools