

SoBigData Erfahrungen und Nutzen

Dr. Thomas Risse
Goethe Universität Frankfurt
Universitätsbibliothek J. C. Senckenberg

SoBigData e-infrastructure

Ein Ökosystem nationaler Forschungsinfrastrukturen

- SoBigData.it (CNR-ISTI, Italy)
- GateCloud (USFD, UK)
- IVAS Platform (FhG-IGD, Germany)
- Alexandria German Web Archive (LUH-L3S, Germany)

Mission

- Gemeinsame Nutzung und Kombination von Ressourcen
- Erleichterung von virtuellen Experimenten und der Reproduzierbarkeit von Ergebnissen
- Verbesserung der Zusammenarbeit

In Zahlen

- Laufzeit: 48 Monate (seit September 2015)
- Budget: 5 Millionen EUR

Big Data Ecosystem

- Open Data
- Restricted Data
- Virtual Collections

Social Mining

- Text & Social Media Mining
- Social Network Analysis
- Human Mobility Analytics
- Web Analytics
- Visual Analytics
- Social Data

Ethical and Legal Framework



Virtual Access
E-infrastructure



Transnational Access
Open calls
Exploratory projects

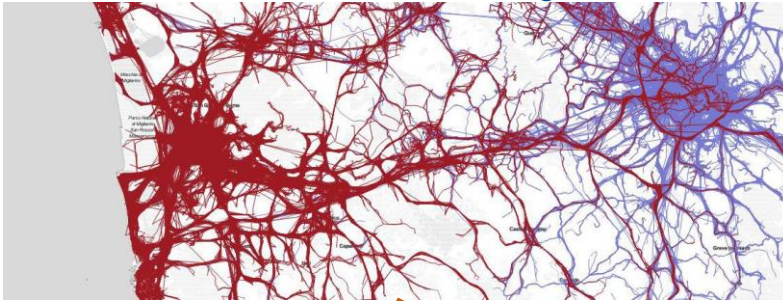


Networking
Training
Dissemination
Innovation Accelerator

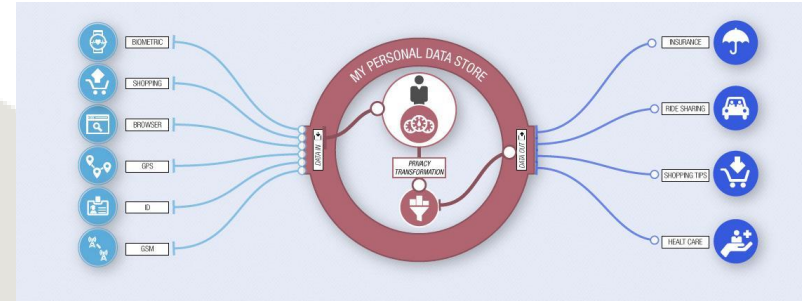


Anwendungsfelder von SoBigData

Mobilitätsforschung



Studien zum Gemeinwohl und Wirtschaft



Social Mining & Big Data Ecosystem
SoBigData
 RESEARCH INFRASTRUCTURE

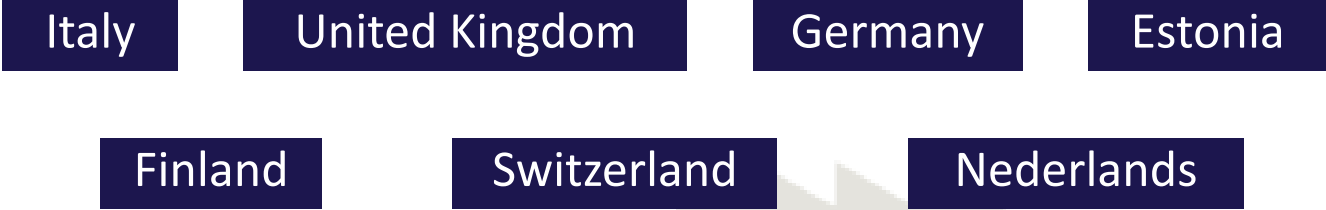


Öffentliche polarisierte Themen und Debatten



Migrationsstudien

Das Konsortium



Existing national RI's to be integrated



Entstehung des Projektes (1/2)

Anfang 2014

- Erste Idee entstand aus einem bmbf Antrag/Begutachtung zu Digital Humanities
- Kerngruppe der Projektidee:
 - LUH-L3S: Datenanalyse und „German Web Archive“ Datensatz
 - Uni Sheffield (USFD): Datenanalyse und Erfahrung mit e-Research Infrastructures
 - Kings College London (KCL): Anwender und Erfahrung mit e-Research Infrastructures
- Call: INFRAIA-1-2014-2015
 - Integrating and opening research infrastructures of European interest
 - Mathematics and ICT - Starting Communities
 - Diskussion ob Text Analyse in den Bereich Mathematik passt

Mai/Juni 2014

- EU informiert KCL über einen weiteren geplanten Antrag/Konsortium (SoBigData) koordiniert von CNR
- Erste Gespräche mit zwischen KCL/USFD/L3S und CNR
- Entscheidung einen gemeinsamen Antrag zu stellen

Entstehung des Projektes (2/2)

SoBigData Konsortium

- 8 Partner
- Folgeaktivität aus einem fehlgeschlagenen H2020 FET Flagship Antrag
- Toskana förderte den Aufbau einer lokalen Forschungsinfrastruktur SoBigData.it
- CNR entwickelt Basistechnologie für Forschungsinfrastrukturen und verwendet sie in verschiedenen e-Infrastructure Projekten

Juni – September 2014

- Erstellung des Antrags
 - Überwiegend Ideen aus dem Flagship Antrag ergänzt um Textanalysis von L3S/USFD
 - Kernteam: CNR und USFD
- Herausforderungen
 - Urlaubszeit im August
 - Viele „unübliche“ Begriffe und Angaben: Transnational Access, Virtual Access, Access Cost Calculation, ...
 - > Informationen vom Koordinator

Februar 2015: Feedback: SoBigData steht auf der Warteliste

Sommer 2015: Förderungszusage

Oktober 2015: Kick-Off Meeting

Projekttablauf / Lessons Learned

- Die versprochenen Technologie und Inhalten sollten vorhanden sein
- Integration von Diensten findet bereits im ersten Jahr statt
- Durchführung von viele Workshops, Summer Schools, etc.
 - Mehr Aufwand in die Organisation als ursprünglich geplant
 - Ursprüngliche Planung überambitioniert
- Herausforderungen
 - Zieldefinition: Was ist eine „e-Research Infrastructure“?
 - Lose Sammlung von Diensten vs. hochintegrierte Plattform
 - Wie arbeitet eine Wissenschaftler mit dem System?
 - Breites Spektrum an Benutzererfahrung
 - Das Endergebnis muss sehr früh allen Partnern klar sein
 - Virtual Access vs. Transnational Access
 - Zielgruppendefinition
 - Inhalt der Ausschreibung in Abhängigkeit zum Entwicklungsstand
 - Besucher finden (Mittel für Gäste sind beschränkt)
 - Betriebsmodell
 - Hoffnung auf Folgeanträge

Vor- und Nachteile der Teilnahme für Forschungseinrichtungen

Als reiner Technologielieferant

- Nur anwendungsorientierte Forschung, keine Grundlagenforschung
- Direkter Kontakt mit verschiedenen Nutzern → Neue Forschungsfragen für Erweiterungen
- Nutzung eigener Entwicklung im größeren Kontext
- Folgeteilnahme nicht gesichert

Als Infrastrukturanbieter

- z.B. Digitale Sammlungen, Anbieter elektronischer Dienste
- (Längerfristige) Finanzierung zur Stabilisierung der Dienste
- Moderate Weiterentwicklung (anwendungsorientierte Forschung)
- Direkter Kontakt mit verschiedenen Nutzern → Neue Forschungsfragen für Erweiterungen
- Verbesserung der Sichtbarkeit der Arbeiten und der Institution

Als Nutzer der Infrastruktur

- Einfacher Zugriff auf größere verteilte und teilweise geschützte Inhalte
- Großes Angebot an Diensten und deren Kombinationen
- Best Practice Guidelines, Schulungen, etc.
- Networking mit anderen Nutzern
- Direkter Kontakt zu den „Profis“



Dr. Thomas Risse
Leiter Elektronische Dienste
Tel. +49 69 798 39 905
Email: t.risse@ub.uni-frankfurt.de

Universitätsbibliothek J. C. Senckenberg
Bockenheimer Landstraße 134 - 138
60325 Frankfurt am Main
<http://www.ub.uni-frankfurt.de>