

Einmal Big Data - INSPIRE (hin und zurück)

Sven Schade

Institute for Environment and Sustainability
Digital Earth and Reference Data Unit

www.jrc.ec.europa.eu

*Serving society
Stimulating innovation
Supporting legislation*



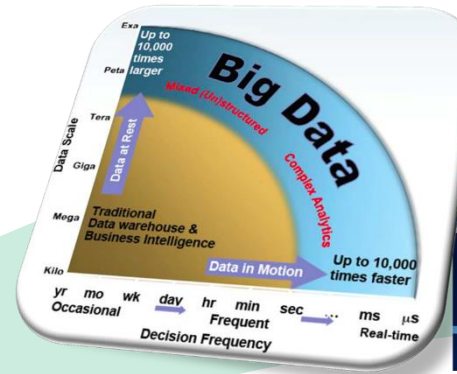
Reiseübersicht



- Big Data
- INSPIRE
- Coverages

- Big Data
- Noch mehr Big Data
- Big Geospatial Data
- **Big Data > INSPIRE**
- INSPIRE > Big Data
- Fazit

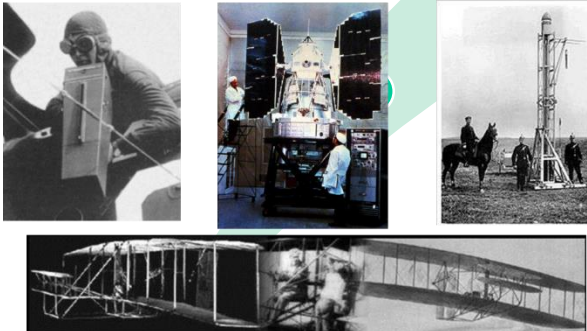
Entstehung des Phenomäns “Big Data”



“Big Data”

volume
variety
velocity
veracity
v...

Datenaustausch



Datengemeinschaften

Eine Flut von Bildern, Punktwolken und Standorten

European
Commission

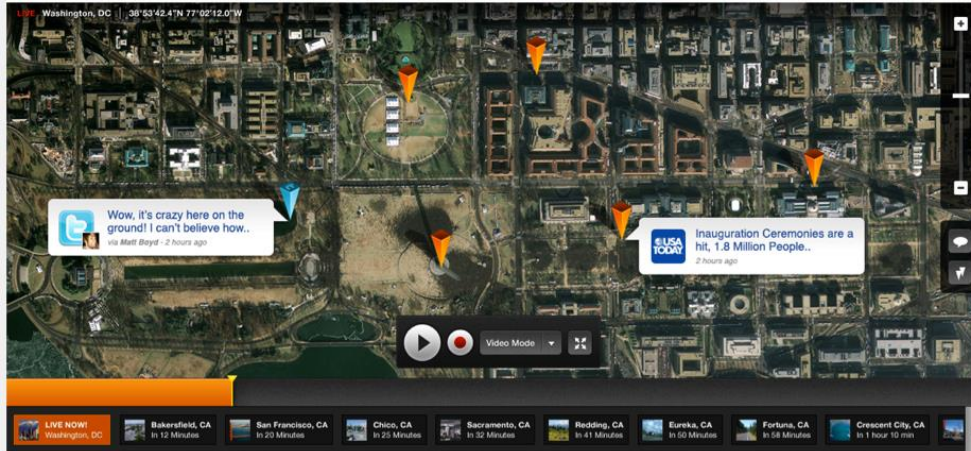
URTHECAST

Search for People, Places or Things...

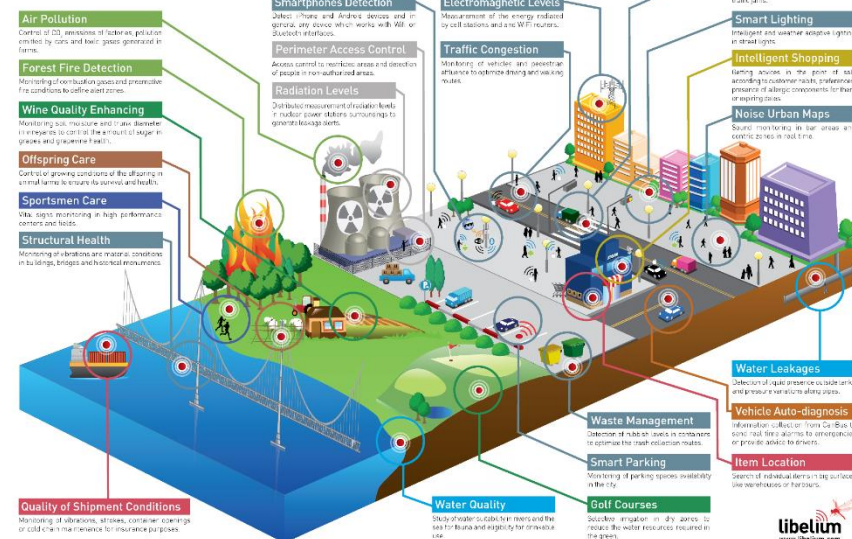
Create an Account

Login

We realize that a new site can be confusing at first so [here's a short tutorial to help you get started.](#)



Libelium Smart World



211,128 Instagram photos

514,048 Skype calls

2,058,005 Tweets

6,388,848 Dropbox files

19,566,471 Google searches

29,236,123 YouTube videos

34,322,697 Facebook likes

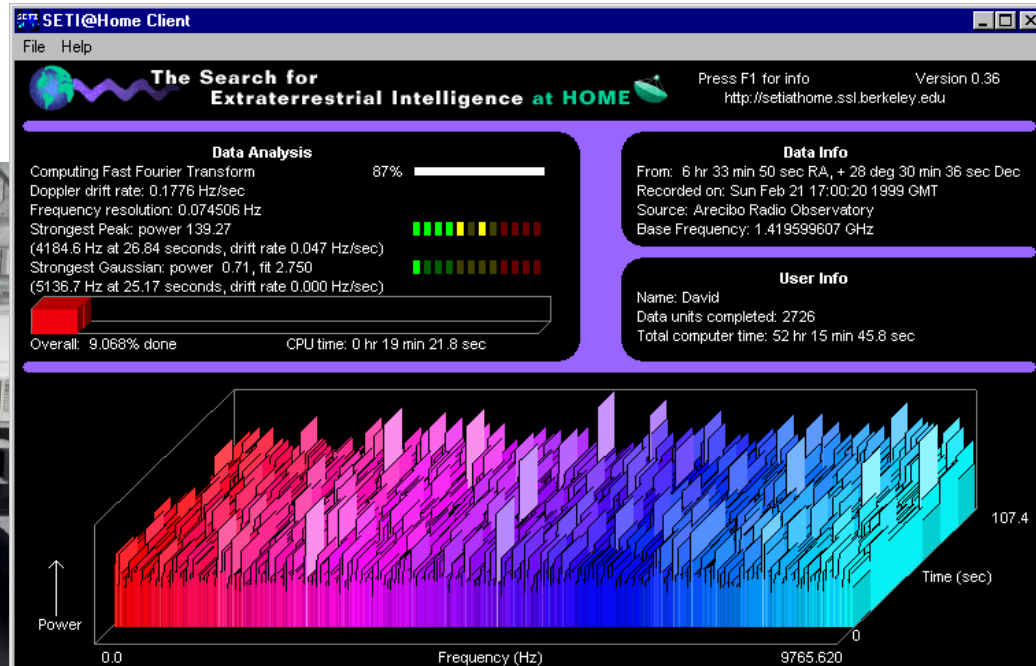
1,139,999,544 Emails

... in einer Sekunde

Joint
Research
Centre



All diese extremen **Datenmengen** treffen auf immer stärkere **Computerkapazitäten** (Could-Computing, Hochleistungsrechner, neue Netzwerkstrukturen)



BIG DATA LANDSCAPE, VERSION 3.0

Exited: Acquisition or IPO

Infrastructure

NoSQL Databases

FOUNDRATIONDB, DATASTAX, mongoDB, COUCHBASE, KEROPIKE, HYPERTABLE, bosh, CLOUDANT, ChmData, Neo4j, bones

Big Data

HADOOP On Prem, HADAPT, cloudera, splice, Zettaset, amazon, MAPR, Microsoft, Pivotal, IBM, Hortonworks, MORTAR, infochimps, Quobale, altiscale, amazon

NewSQL Databases

MarkLogic, TRANSILATICE, ROLIN, paradigms, deepdb, SkySQL, Clustrix, VoltDB, SQLFire

MPP Databases

TERADATA, InfiniDB, Kognitio, Netezza, SQL Server, Pivotal, PARACCEL

Graph Databases

Neo4j, aster data, InfiniteGraph

Crowd-sourcing

microtask, CROWD COMPUTING, servio, mobileworks

Data Transformation

TRIFACTA, Talend, KALIDO, everlytic, TRAIKON, syncsort

Security

DATA GUISE, Stormpath, IMPERA

Storage

Cleversafe, Panteras, simlstorage, Compuware

App Dev

CONCURRENT, CONTINUITY, wibedata

Analytics

Analytics Platforms

databricks, QuantCast, PERSASIVE, GUAVUS, Databricks, KARMACHINE, PRECOG, dataspora

For Business Analysts

STAT WING, CIRRO, TREPAREL, OrigamiLogic, ClearStory, DataGravity

Data Science Platforms / Tools

domino, Alpine, Sense, MORTAR, CONTINUUM, plooty, yhat, MOOSE

Unstructured Data

BASIS, ATTIVO, GENERAL SENTIMENT, semantix, Quid, Palantir

Data Visualization

tableau, ZoomData, visual.ly, Roambi, Chart.io, Quantum4D, ACTUATE, Kitenga, Ayasdi, ISS, DataHero, GEEKS4GEEKS

Machine Learning

SKYTREE, bigml, YOTTAWARE, wise.io, conduct relevant

Location / People / Events

RADIUS, FLIPTOP, COGATE, Place IO

Big Data Search

hp, LucidWorks, ontologr

Statistical Computing

sas, MATLAB, SPSS

Log Analytics

splunk, loggly, sumologic, Kibana

Crowd-Sourced

kaggle, DataKind

Real-Time

NETAMARKETS, amito, causata

SMB

RJMetrics, retention, GoSquared, sumall, custora

Applications

Ad Optimization

aggregate knowledge, rocketfuel, TAPAD, Matcha, thetradeesk, 33score

Publisher Tools

Chartbeat, Yieldex, yieldbot

Marketing

LATTICE ENGINES, Sailthru, spinsaker, gainsight, Kontera, RelateIQ, Teluspart, persado, bloomreach, CLICKFOX, Purway

Human Capital

evolv, entelo, gild

Legal

JUDICATA, RAVEL, Lex Machina

Finance

LendUp, Lenddo, BILL GUARD, wonga, cignifi, OnDeck

Government / Regulation

mark43, enigma, Priceline, PREDPOL

Security

ENIFY, siftscience, FEEDSCALE, feedzai

Educational Learning

KNEWTON, Geclara, PANORAMA, Clever

Health

Recombine, 23andMe, Ginger.io, FLATIRON, Coursera

Industries

tubular, OPOWER, NEXT BIG SOUND, SIGHT MACHINE, THE CLIMATE CORPORATION

Cross Infrastructure /

SAP, sas, IBM, Google, Microsoft, vmware, amazon, 1010data, talend, TERADATA, hp, NetApp

Open Source

Frame-work

Spark, Hadoop YARN, HDFS

Query / Data Flow

Query / Data Flow

Data Access

Cassandra, SciDB, ORACLE, HBASE, mongoDB, riak, Sqoop

Coordination / Work-flow

talend, ZKeeper

Real-Time

Storm

Stat Tools

SciPy, (P)

Machine Learning

ML

Cloud Deploy

Cloud Deploy

Search

Solr, ELASTICSEARCH, LUCENE

Data Sources

Data Mkts

Windows Azure, Chukra, DataMarket, factual

Data Sources

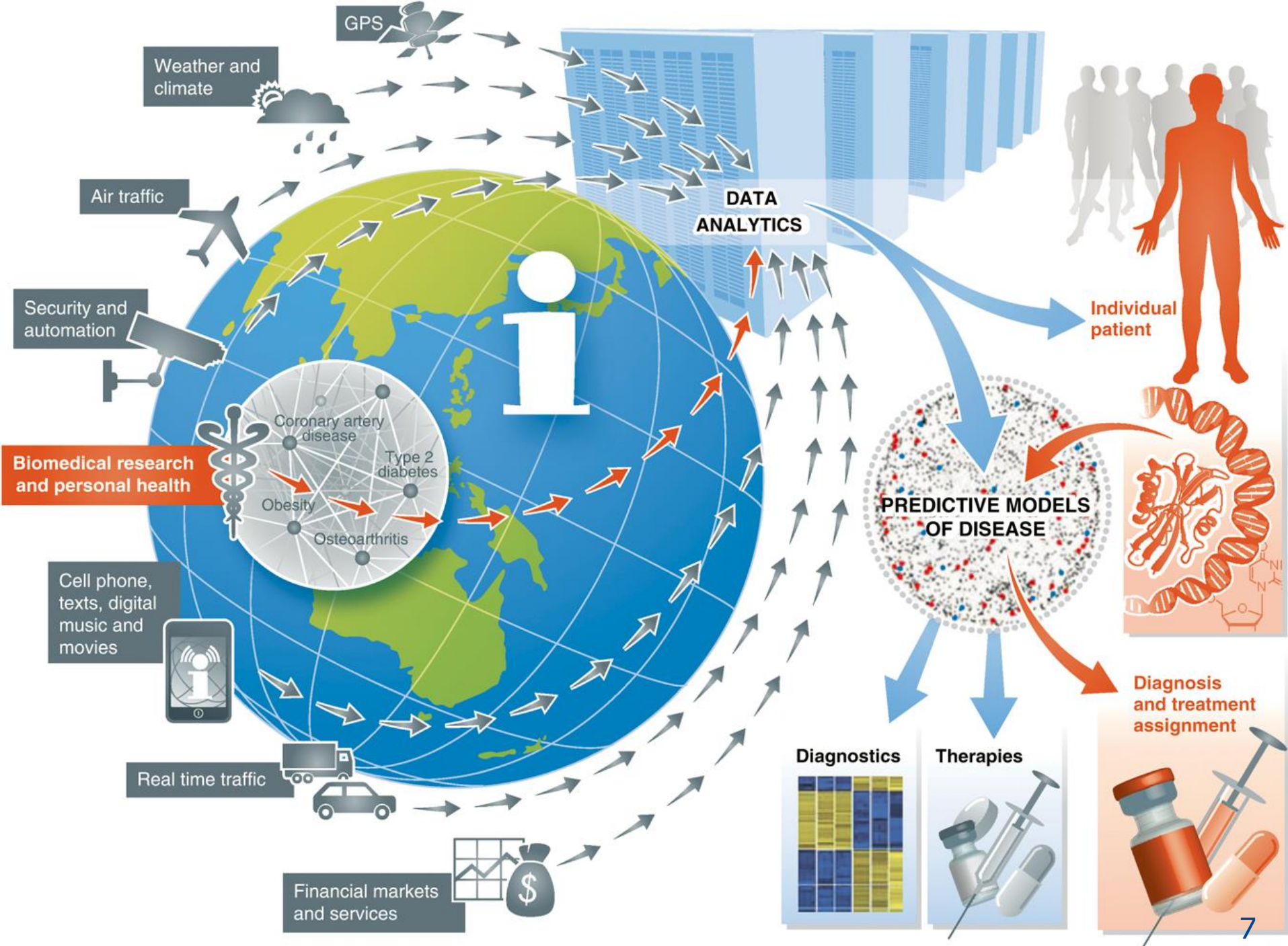
DATA.GOV, premise, YODLEE, xignite, ploid, quandl, STANDARD TREASURY, human/api

Sensor Data

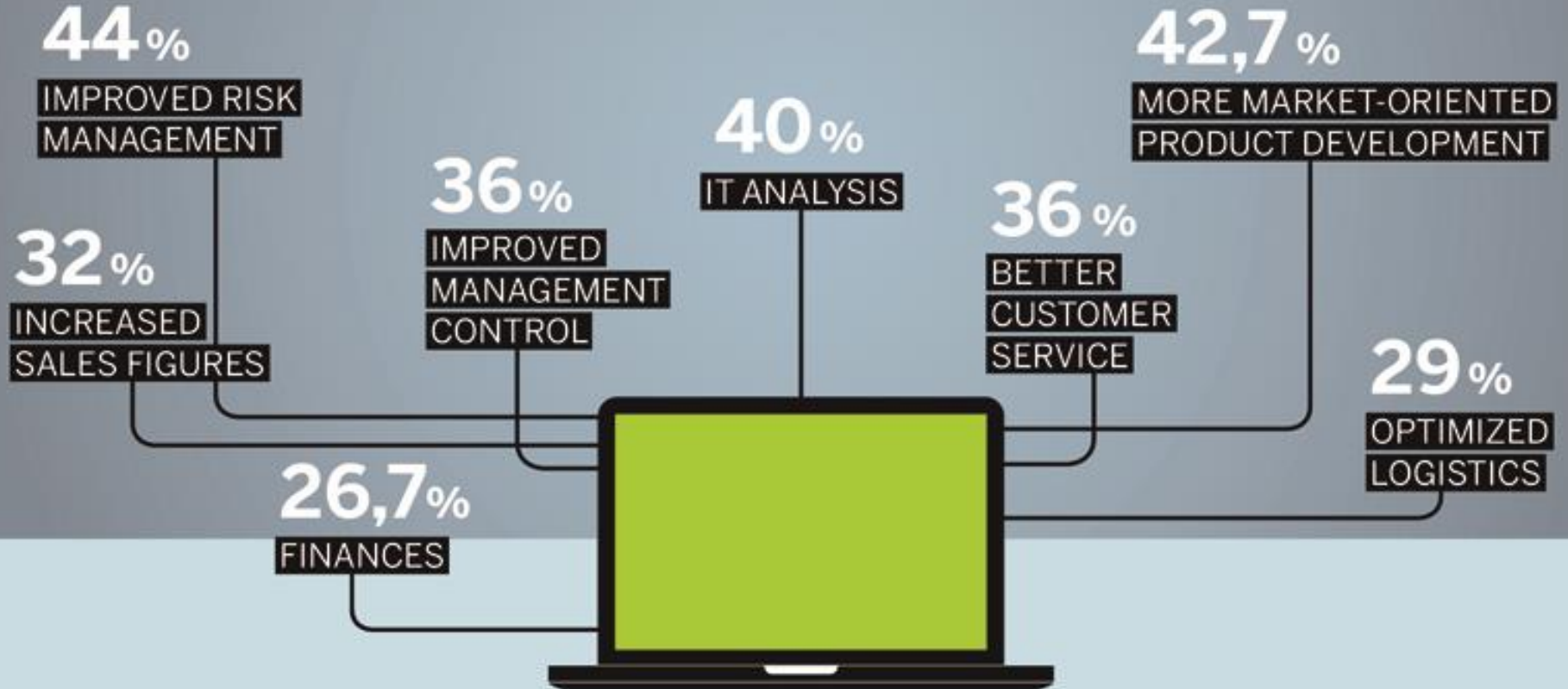
kinsa, STREETLINE, fitbit, RunKeeper, JAWBONE, SKYCATCH, LUMASENSE, Withings, BASIS, estimate

Incubators & School

zipfian, GA, INSIGHT, 6



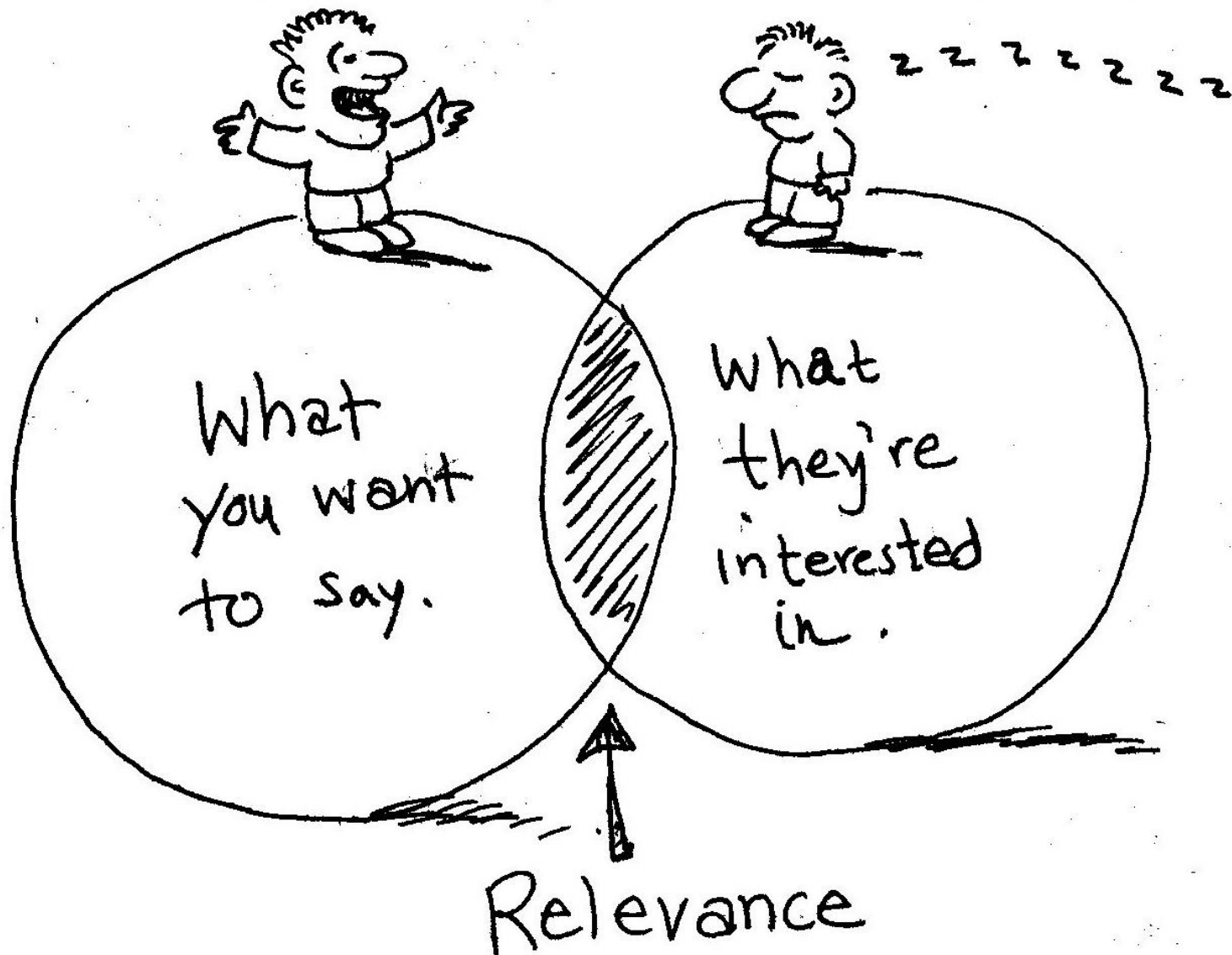
WHICH APPLICATION SCENARIOS WILL BIG DATA SOLUTIONS BE USED FOR IN THE FUTURE?



Der Geoinformations-Aspekt:



...Positionen, Orte, Karten, Bilder

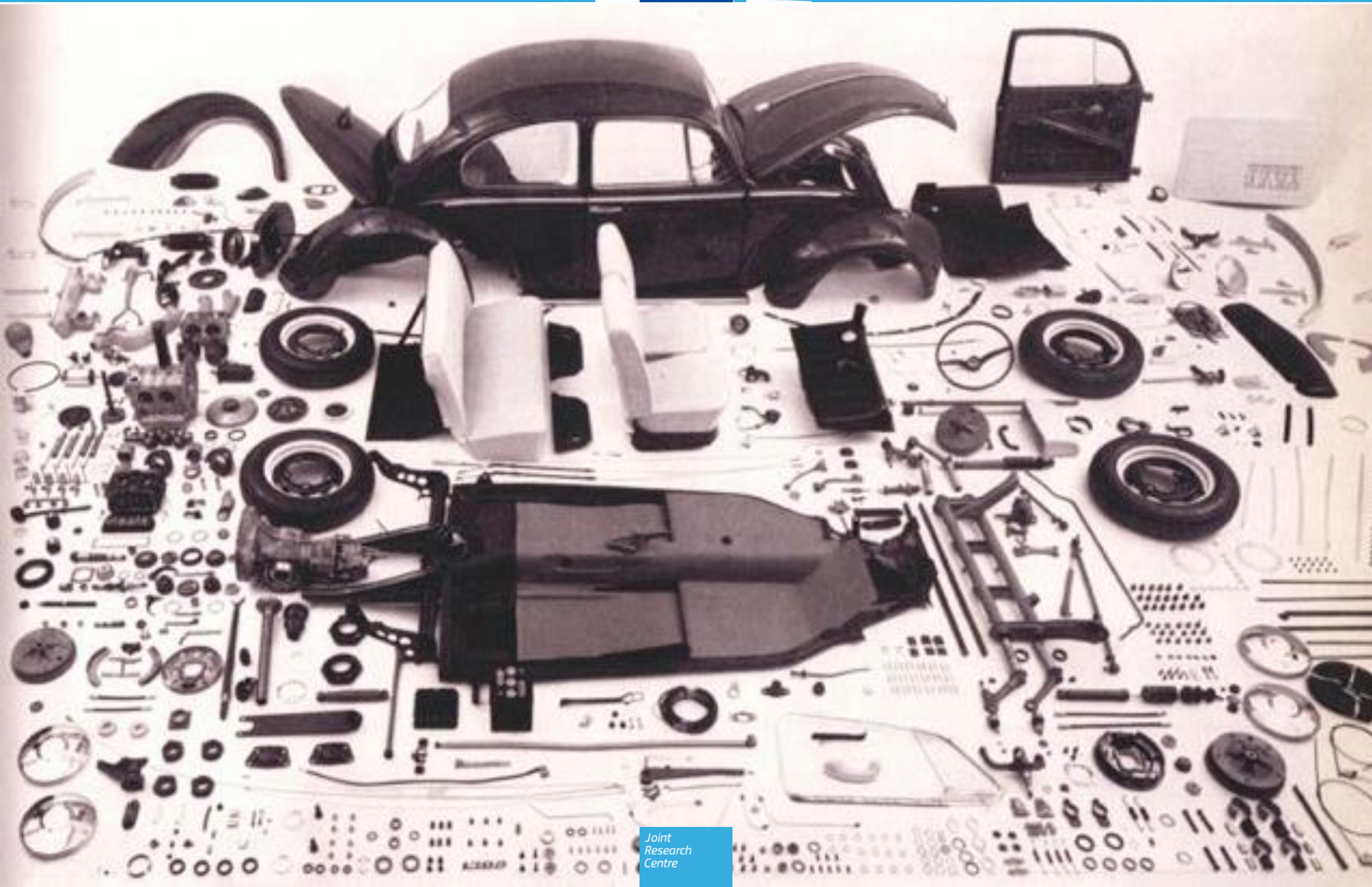


Die Relevanz von Big Data für INSPIRE

Auf den ersten Blick...

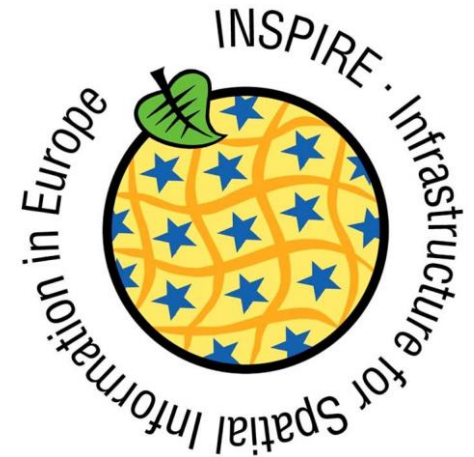


Detailbetrachtungen



“INSPIRE”? Was ist das eigentlich?

- Ein Prozess zur Harmonisierung des Austausches von Umweltdaten
(**34 Themen**, 3 Anhänge) in Europa
- Eine **EU Richtlinie**
- Implementing Rules & Technical Guidelines
- **Addressiert einen komplexen Bereich**
- Gemeinschaftliche Aufgabe
- Ein Erfolg der über die Grenzen Europas hinaus anerkannt und angewendet wird!



Big-Data-Probleme und INSPIRE

Variety (Vielfalt): Motivation und Geltungsbereich (Dateninteroperabilität), Themen, Interessenvertreter, beteiligte Institutionen, Status, Herangehensweise

Veracity (Glaubwürdigkeit): Validierung noch in Diskussion, Metadaten für das Auffinden von Daten und Diensten - nicht für diverse Qualitätsmerkmale

Velocity (Schnelligkeit): Aktualisierungszykel in den Mitgliedsstaaten und der EC, technologischer Wandel, Schwierigkeiten zu synchronisieren und mitzuhalten, Sensordaten (an Ort und Stelle, und aus der Fernerkundung)

Volume (Umfang): Anzahl an involvierten Interessenvertretern und Fachleuten (länderabhängig), könnte für Daten in mehreren Themenbereichen eine Rolle spielen

Big-Data-Probleme und INSPIRE

Variety (Vielfalt): Motivation und Geltungsbereich (Dateninteroperabilität), Themen, Interessenvertreter, beteiligte Institutionen, Status, Herangehensweise

Veracity (Glaubwürdigkeit): Validierung noch in Diskussion, **Metadaten** für das Auffinden von Daten und Diensten - nicht **für diverse Qualitätsmerkmale**

Velocity (Schnelligkeit): Aktualisierungszykel in den Mitgliedsstaaten und der EC, technologischer Wandel, Schwierigkeiten zu synchronisieren und mitzuhalten, **Sensordaten (an Ort und Stelle, und aus der Fernerkundung)**

Volume (Umfang): Anzahl an involvierten Interessenvertretern und Fachleuten (länderabhängig), **könnte für Daten in mehreren Themenbereichen eine Rolle spielen**



Sensordaten als ein Kernthema: Datenmodell

- Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE Annex II and III data specification development
including discussion of including images (read coverages) in GML envelopes (Annex B)
- Similar guidelines for the use of Coverages desirable

Annex II

Items per page 50 ▾		Filter coverage
Label	Theme	Status
Elevation Grid Coverage*	Elevation	Valid
Land Cover Grid Coverage	Land Cover	Valid
Orthoimage Coverage*	Orthoimagery	Valid

Annex III

Soil Theme Coverage	Soil	Valid
Soil Theme Descriptive Coverage	Soil	Valid
Exposed Element Coverage	Natural Risk Zones	Valid
Hazard Coverage	Natural Risk Zones	Valid
Observed Event Coverage	Natural Risk Zones	Valid
Risk coverage	Natural Risk Zones	Valid
Renewable And Waste Potential Coverage	Energy Resources	Valid

*

DTs jointly work on

encoding in TIFF

(DS Elevation Annex E)



Sensordaten als ein Kernthema: Dienstemodell

- INSPIRE Download Services
- TG includes OGC-WFS-based solution (for Direct Access Download Services)
- This is now being extended to include OGC-SOS for in-situ measurements (see also presentation at FOSS4GEurope)
- **Similar approach could be imagined for OGC-WCS**

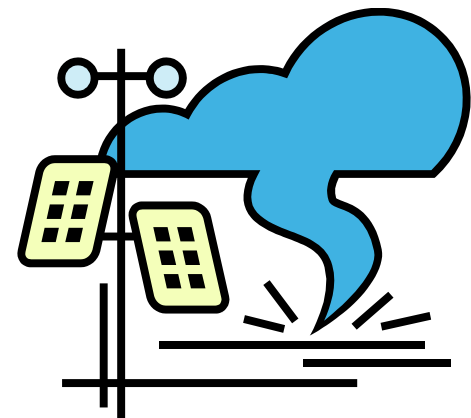
Collaboration space on download services TG:

<https://ies-svn.jrc.ec.europa.eu/projects/download-services-tg>



Sensordaten als ein Kernthema: Prozess

- **ARe³NA** study on proposed TG update and open source prototype (reference implementation) for OGC-SOS
- Study results provided to MIG
- MIG is now maintaining the results and plans to update TG on INSPIRE Download Services (as part of **MIWP-7**)
- **Desire to do similar activity for OGC-WCS**
- **Teleconference foreseen for in July 2014**
- Back to Peter





**INSPIRE & Big Data =
Alter Wein in neuen Schläuchen?**

INSPIRE bekommt...

- Neue Daten
- (Neue Technologien)
- Neue Interessenten
- **Neue Möglichkeiten**
- Evtl. neue Datenquellen

INSPIRE behält...

- Mandat
- Herausforderungen
- Kern-Zielgruppe



Die Relevanz von INSPIRE für Big Data

Auf den ersten Blick...

You Are Here

A composite image of the Whirlpool Galaxy (M51) against a starry background. The galaxy is shown in a tilted, edge-on perspective, with a bright yellowish-white central core and reddish-brown dust lanes. Two bright blue-white spots are visible: one on the left side of the galaxy's main body and another in the lower foreground. A white speech bubble with the text 'You Are Here' points to the blue spot on the galaxy's left side.

... aber!





INSPIRE als Fallbeispiel und Teil des Rahmenwerks

- INSPIRE **löst zentrale Probleme** der (Geo-)Datenvielfalt und Interoperabilität (im öffentlichen Sektor in Europa)
- Der allgemeine Lösungsansatz kann auf andere Bereiche **übertragen** werden u.A. in Erweiterungen innerhalb des Umweltbereiches (siehe Luftqualität), und im Brückenschlag mit anderen Bereichen (z.B. Energie)
- Andere Sektoren könnten Teile übernehmen (z.B. die Wissenschaft)
- INSPIRE kann auch ausserhalb Europas eingesetzt werden

ABER: Verbindungen mit anderen Interessengruppen müssen geschaffen und gepflegt werden (auf organisatorischer, inhaltlicher und technischer Ebene)

Zusammenfassung

Friedliches Zusammenleben ist möglich!

Keine bahnbrechenden Veränderungen erwartet.

INSPIRE löst seit jeher Big-Data-Probleme

- Vor Allem hinsichtlich Daten-Vielfalt
- Und das in einem höchst komplexen Bereich

Beide Seiten sollten noch von einander lernen

- **Anpassung** der Richtlinien **an neue Lösungen**
- **Weitere Öffnung von INSPIRE für neue Nutzer**
- Anwendung von INSPIRE-Prozessen für Dateninteroperabilität



Was machen “wir” in Sachen INSPIRE & Big Data?

- INSPIRE: Implementierung, Instandhaltung und Erweiterung
- Open Data Strategie des JRC
- **Big Data: Einflüsse auf Politik und Wissenschaft**



Das Potential von Big Data für neue Wissenschaftsformen

FOSS4G Europe 2014 (Academic Track)

Mittwoch 16. Juli

16:30 (Raum 3)

Gleichzeitig (für die Praktischer-Orientierten):

“An Open Source SOS-Based INSPIRE

Download Service Implementation”

Simon Jirka, 16:00 (Raum 1)



Siehe auch: <http://tinyurl.com/are3na-joinup>





Vielen Dank für ihre Aufmerksamkeit!

