SoBigData EUROPE
Research Infrastructure
on Social Mining &
Big Data Ecosystem

Thomas Risse

2nd International Alexandria Workshop 2015

We inhabit a measurable global society nowadays

The increasing wealth of data is a chance to disentangle the social complexity
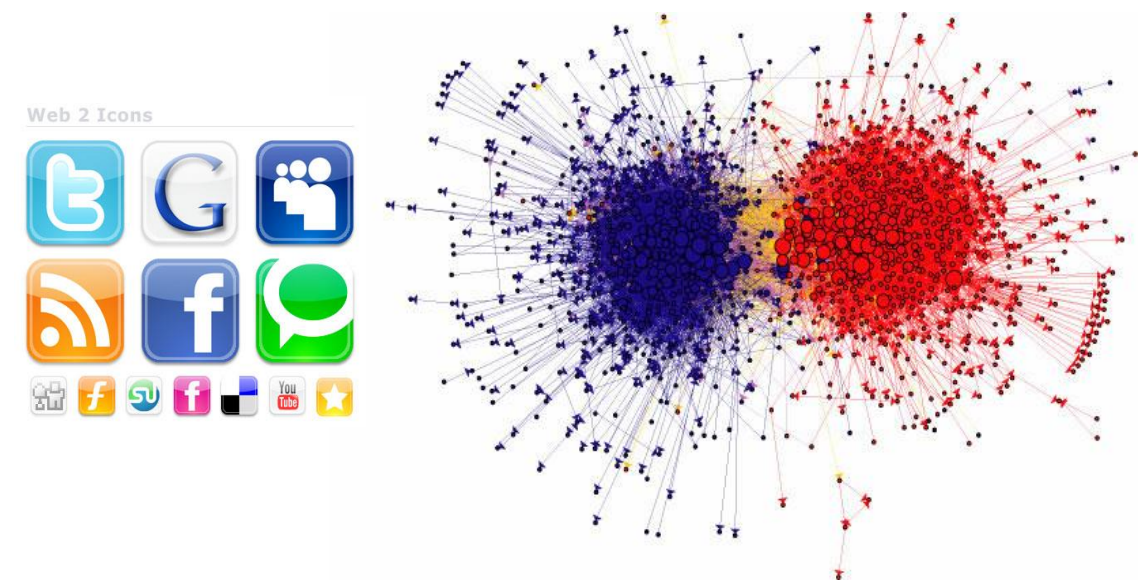
Big data "proxies" of social life
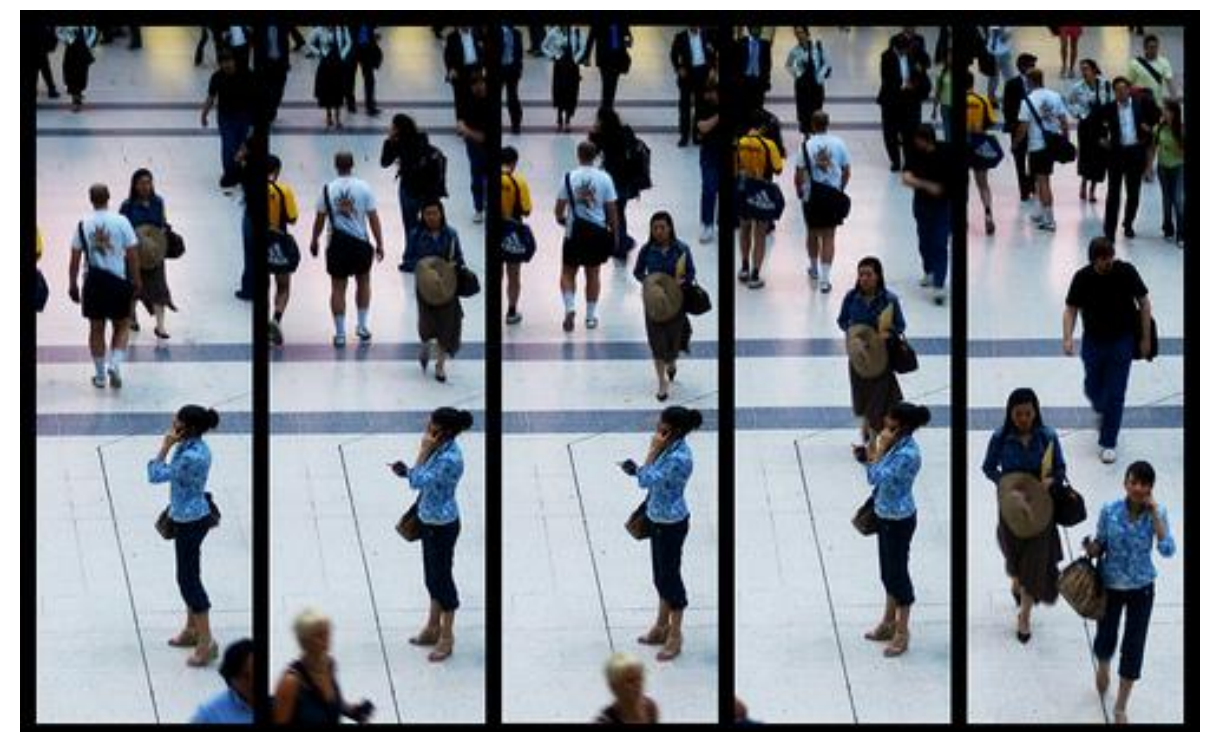
Shopping patterns & lifestyle

Relationships & social ties

Desires, opinions, sentiments

Movements

Big Data Analytics & Social Mining

A TOOL TO MEASURE, UNDERSTAND, AND POSSIBLY PREDICT HUMAN BEHAVIOR

Systematically analyze valuable information from social data for

- monitoring well-being and resilience of society based on behavioral patterns
- discovering surprising correlations between the socio-economic status and big data measurements,
- nowcasting happiness, trust, the patterns of human mobility,
- Etc.

Data+ Knowledge+People Infrastructure

ETHICAL VALUES: privacy, trust, transparency

EDUCATION: data literacy – data scientists
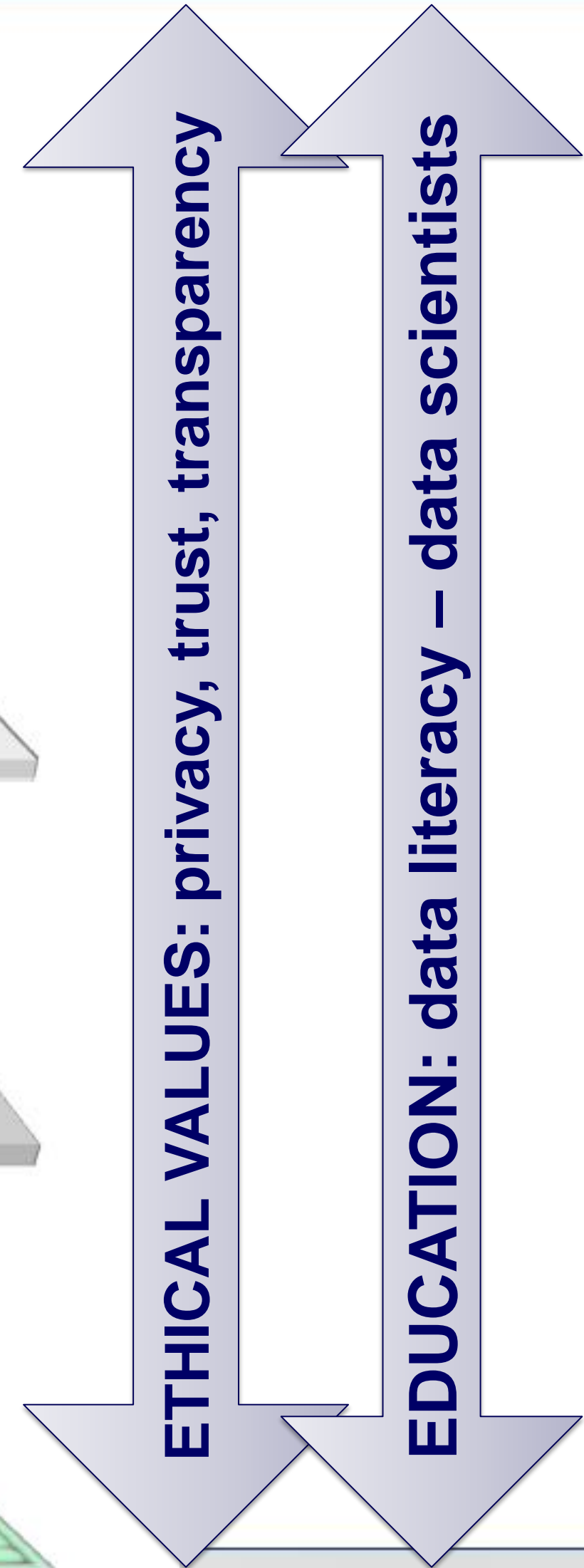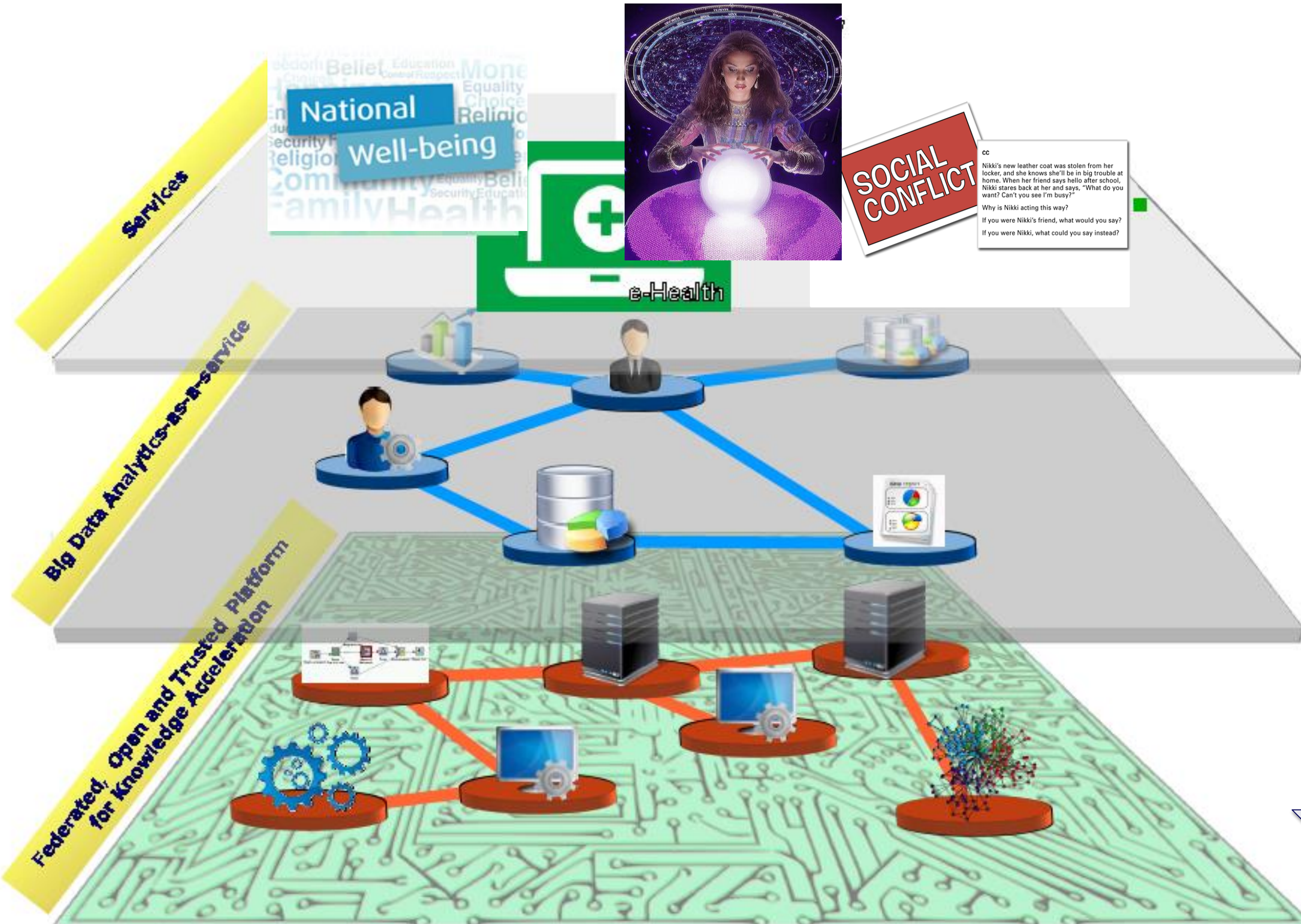
Existing Research Centers

- multiple sources and collections of data of various quality and quantity

- multiple social sensing methods to procure, manage and curate big data sources

- multiple social mining methods to make sense of these data and extract relevant social knowledge from them;

- Multiple visualization and story-telling tools for sustaining the discovery and the interpretation of social knowledge and turning it into actionable policies.

But

Expertise and tools are fragmented, due to the difficulties in sharing experimental data and methods

SoBigData - The Social Mining & Big Data Ecosystem

Research infrastructure (RI) for ethic-sensitive scientific discoveries and advanced applications of **social data mining** to the various dimensions of social life, as recorded by "big data".

Pillars of SoBigData

- An ever-growing, distributed **data ecosystem** for procurement, access and curation of big social data.

- An distributed platform of **interoperable, social data mining methods** and associated skills

- A **starting community** of scientific, industrial, and other stakeholders (e.g. policy makers), supported by transnational and virtual access activities

Integrating key national infrastructures and centers of excellence

Create on European level a networked virtual ecosystem in big data analytics and social mining - the **SoBigData** RI.
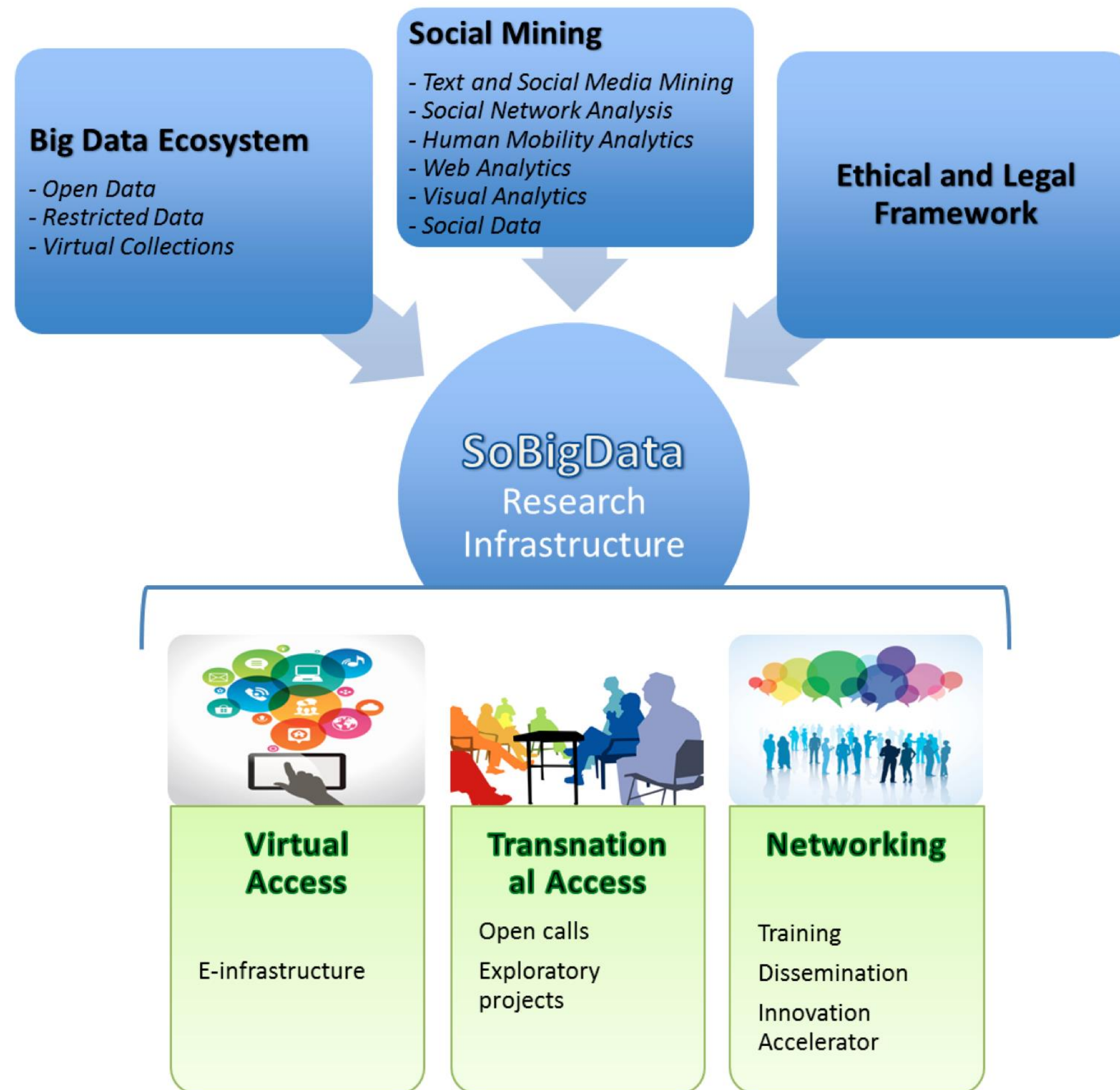
Infrastructures to be integrated

- **SoBigData.it (CNR, Uni Pisa)**: Big Data Analytics and Social Mining

- **GATE Cloud (Uni Sheffield)**: Cloud-based infrastructure for large-scale, data intensive natural language processing and text mining

- **IVAS (FhG IGD)**: Infrastructure for Information Visualization and Visual Analytics

- **Alexandria (L3S)**: Infrastructure for Web Science offering data, services and expertise on Web Archives

- **AALTO (Uni Aalto)**: data, services and competences on social network analysis

- **E-Gov.data (Uni Tartu)**: Estonian e-government and ehealth data over 10 years

- **Living Archive for Open Data (ETH Zürich):** Search engine for Open Data by the Department of Sociology at ETH Zurich

Conceptual architecture of SoBigData

Social Mining

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

Big Data Ecosystem

- *Open Data*
- *Restricted Data*
- *Virtual Collections*

Ethical and Legal Framework

SoBigData
Research
Infrastructure

Virtual Access

E-infrastructure

Transnational Access

Open calls

Exploratory projects

Networking

Training

Dissemination

Innovation Accelerator

Many different pieces…

Research Publication

Method Catalogue

Open Call system

Analytical Platform

Data Catalogue

Privacy and Sharing Policy

Projects

Physical resources

SoBigData.eu thematic clusters

There are six *thematic clusters* of competences and services

- ■ Text and Social Media Mining
- ■ Social Network Analysis
- ■ Human Mobility Analytics
- ■ Web Analytics
- ■ Visual Analytics
- ■ Social Data
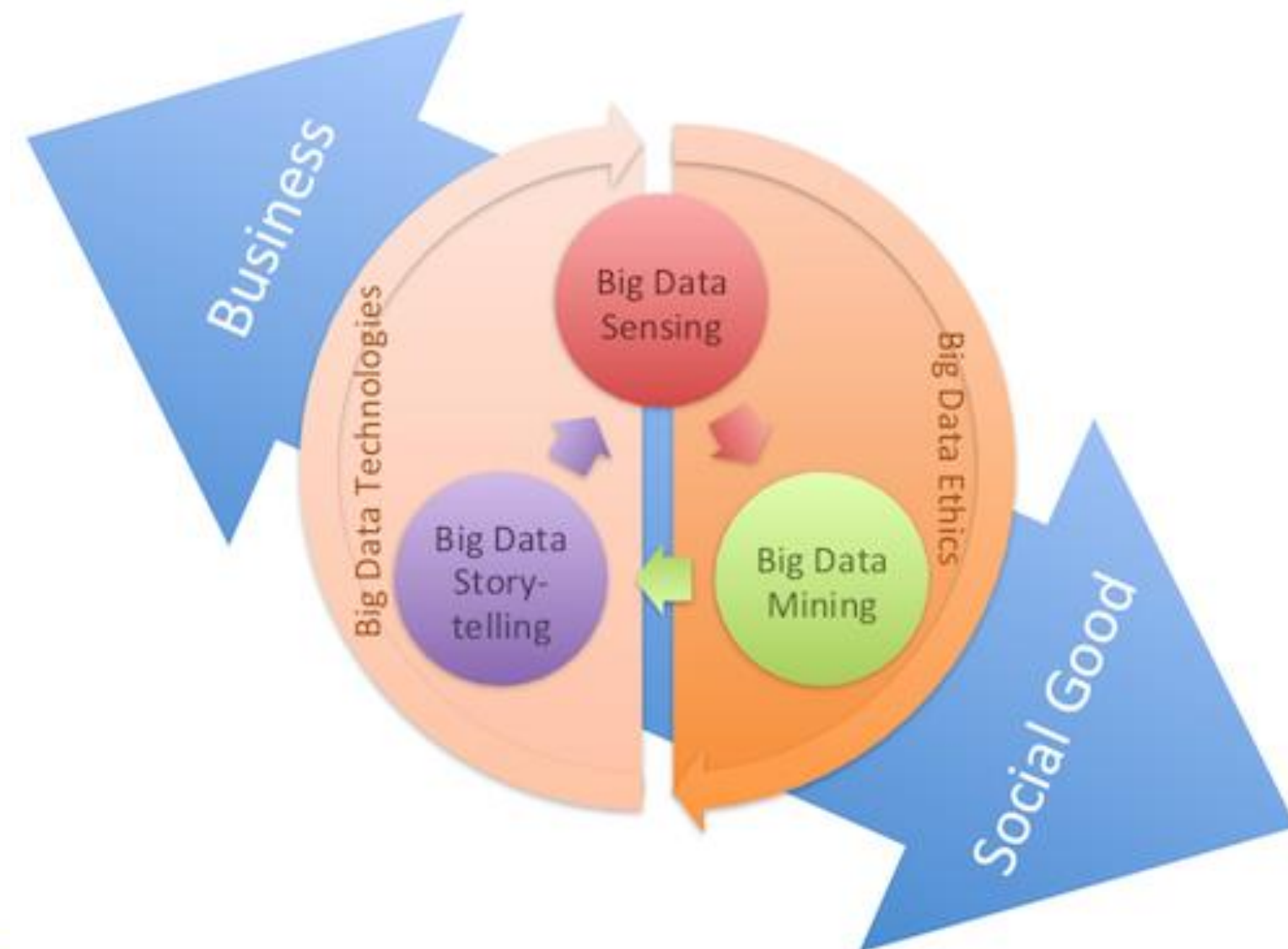
SoBigData.eu Access
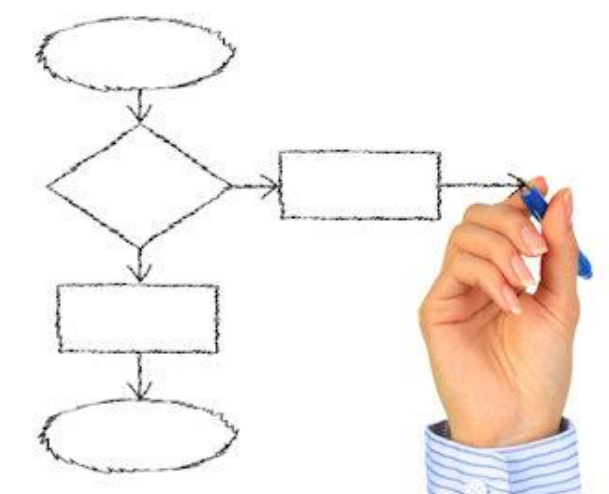
There are two access modalities to data and methods:

Transnational Access

- Exploratory Projects
- Blue-sky projects

Virtual Access

- Data and Methods Catalogue
- Modular virtual research environment

Target Communities

- **Big data analysts and social informatics researchers**
  - enhance their algorithms to deal with social data
  - gain multi-disciplinary research skills
  - harmonize existing data and analytics infrastructures
- **Economists, social science and humanities researchers, journalists, policy and law makers**
  - To analyze the (big) social data, in order to gain insight and actionable knowledge
- **Researchers in related communities**
  - who would like to use the algorithms, the analytical competences and data infrastructure;
- **The public as a whole**
  - For understand their role in the production, consumption and value-creating of social data.
- User Partners
  - Scuola IMT, Lucca
  - King's College London
  - Scuola Normale Superiore di Pisa

Partner

1 – CNR-ISTI: Consiglio Nazionale delle Ricerche (PI: Fosca Giannotti) Italy

2 – USFD: The University of Sheffield (PI: Hamish Cunningham) UK

3 – UNIPI: Università di Pisa (PI: Dino Pedreschi) Italy

4 – FRH: Fraunhofer IAIS and IGD (PI: Gennady Andrienko) Germany

5 – UT: Tartu Ulikool (PI: Marlon Dumas) Estonia

6 – IMT: Scuola IMT (Istituzioni, Mercati, Tecnologie) Lucca (PI: Guido Caldarelli), Italy

7 – LUH: Gottfried Wilhelm Leibniz Universitaet Hannover (PI: Wolfgang Nejdl), Germany

8 – KCL: King's College London (PI: Tobias Blanke) UK

9 – SNS: Scuola Normale Superiore di Pisa (PI: Fabrizio Lillo) Italy

10 – AALTO: Aalto University (PI: Aristides Gionis) Finland

11 – ETHZ: ETH Zurich (PI: Dirk Helbing) Switzerland

12 – TUDelft: Technische Universiteit Delft (PI: Jeroen Van Den Hoven) Netherlands

Summary

Integrating key national infrastructures and centres of excellence at European level in big data analytics and social mining.

SoBigData will leverage these rich scientific assets (big data, analytical tools andservices, and skills), to enable cutting-edge, multi-disciplinary social mining experiments

Granting access (both virtual and trans-national on-site) to the SoBigData RI to multidisciplinary scientists, innovators, public bodies, citizen organizations, SMEs, as well as data science students at any level of education

First version will be demonstrated at the **SoBigData 2016 workshop** in conjunction with **TPDL 2016, 5-9 September 2016, Hannover**

Thank you!

Dr. Thomas Risse
Forschungszentrum L3S
Appelstrasse 9a
30167 Hannover

E-Mail: risse@L3S.de
Telefon: 0511 – 762 17764
Telefax: 0511 – 762 17779