Leveraging the data potential in Europe

European Data Forum 6 June 2012, Copenhagen

On behalf of Khalil Rouhana, EC

Márta Nagy-Rothengass

OUTLINE

1. The digital dilemma
2. Virtuous cycle of data
3. European data value chain
4. Trends

EUROPEAN COMMISSION

The Information Society and Media Directorate General

Directorate Digital Content & Cognitive Systems (based in Luxembourg)

E1: Language Technologies, Machine Translation
E2: Technologies for Information Management
E3: Cultural Heritage & Technology Enhanced Learning
E4: Access to Information
E5: Cognitive Systems & Robotics
E6: eContent and Safer Internet
E7: Administration and Finance

EUROPEAN COMMISSION

The Communication Networks, Content & Technology Directorate General

Directorate Media and Data

G1: Converging Media and Content
G2: Creativity

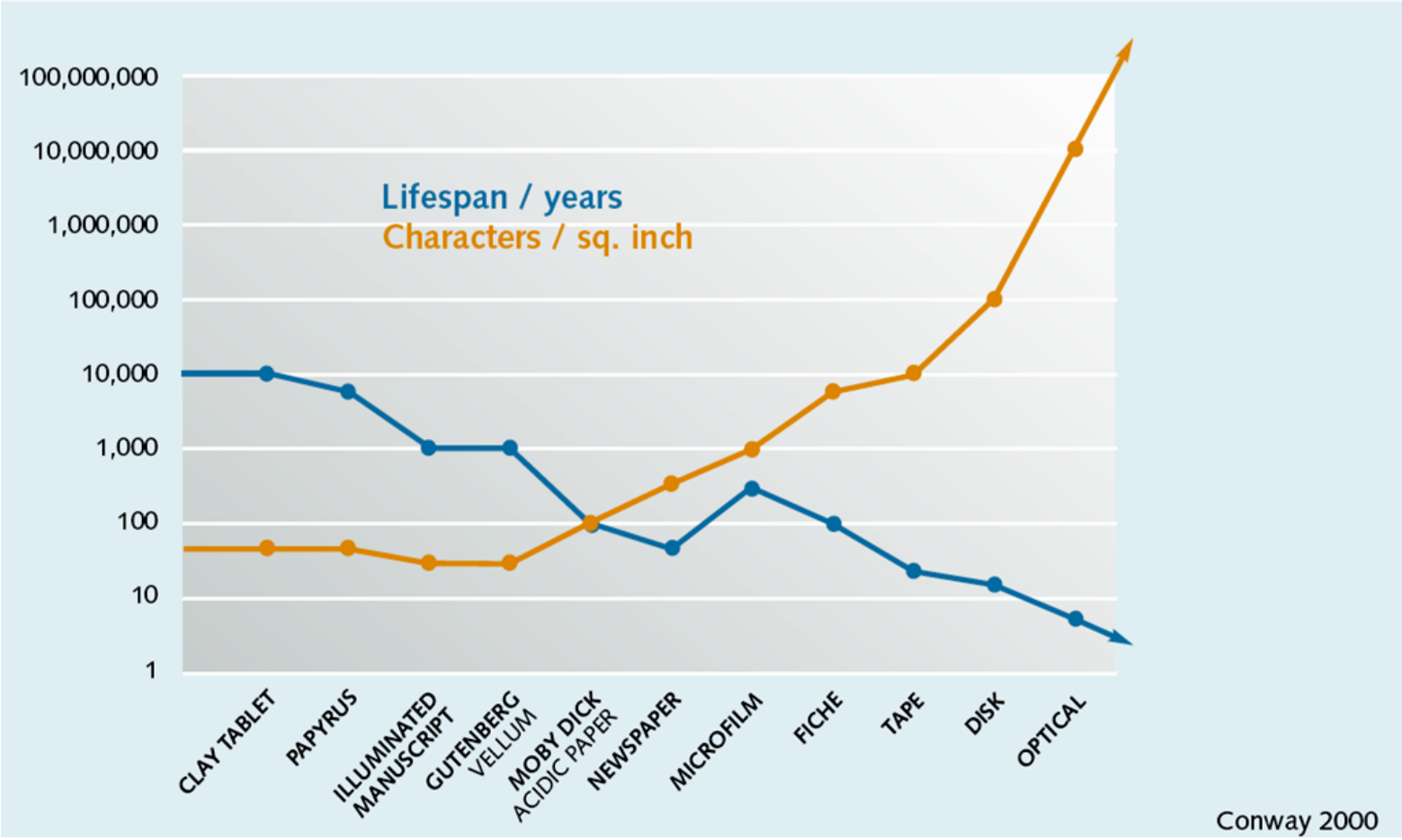
G3: **Data Value Chain**

G4: Inclusion, Skills and Youth
G5: Administration and Finance

As from July 1st 2012

European Commission

The digital dilemma



"DATA Value" Vision

*to **enable** and **foster** best possible social and commercial **added value***

*based on intelligent use, management and re-use **of data sources** in Europe.*

This will lead to

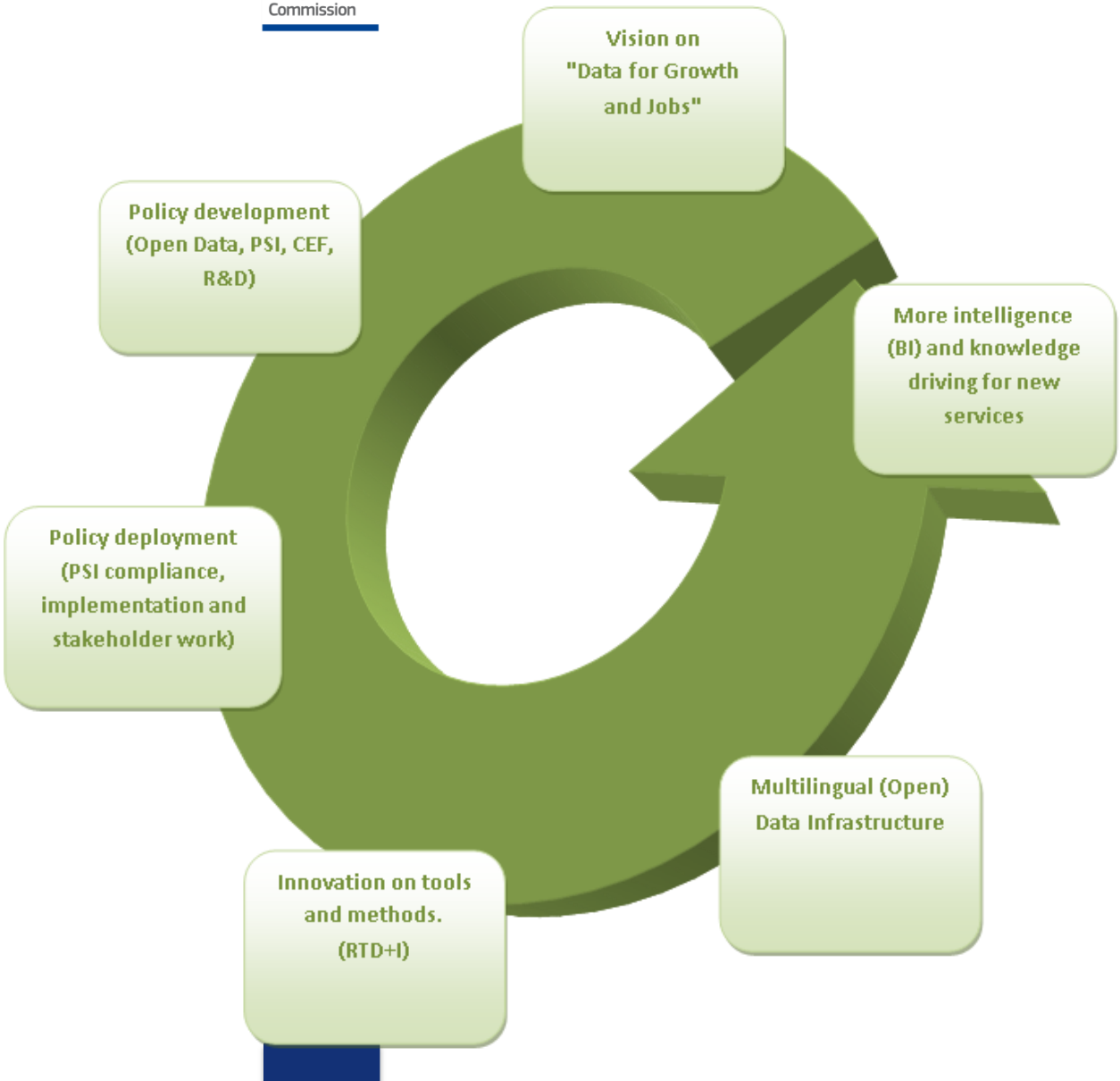
*- **increased business intelligence** and **efficiency** of private and public sectors*

*- **world class applications***

***- new business opportunities** involving SMEs*

*- (open) data friendly **policy and business environment***

Virtuous cycle of data



Main elements of the data cycle (1)

Creation of "data value chain friendly" policy environment:

- Fostering of (Open) **Data policy**

- Adoption of the **revised Directive on the re-use of Public Sector Information** (PSI) and the **Commission decision on re-use of its own information**

- **Implementation of PSI** policy across Europe by ensuring compliance and the development of **soft law instruments** (e.g. guidelines on licensing and charging)

- **Stakeholder** involvement and engagement

Open (public) data: Why does it matter for Europe?

1. **Untapped business and economic opportunities**: data is the new gold; possible direct and indirect gains of €140bln across the EU27

2. **Better governance and citizen empowerment**: open data increases transparency, citizen participation and administrative efficiency and accountability

3. **Addressing societal challenges**: data can enhance sustainability of health care systems; essential for tackling environmental challenges

4. **Accelerating scientific progress**: e-science essential for meeting the challenges of the 21st century in scientific discovery and learning.

Open data strategy

- *Publicly funded data must be available for all*
- *What data are we talking about?*
  - Statistics, geographical data, meteorological data, business information, legal information
- *Dutch geo-sector in 2008: 15.000 jobs*
- *More is possible: unused potential of some 40 billion euro a year across the EU*

How to re-use government data?

- *Combination of different types of data (e.g. geo, traffic and tourism)*
- *EU-wide applications and services*
  - Capitalise on the size of the internal market
- *Apps*
- *Systems that facilitate decision making by companies*

Revised Directive enables easier re-use of PSI

- Creation of a **genuine right to re-use public data**: all public data not covered by an exception is to be re-usable

- Invitation for public bodies to make their **documents** available in a **machine-readable format** and together **with their metadata** where possible and appropriate

- **Charges** shall be limited to the **marginal costs** of reproduction and dissemination (**does not apply to cultural institutions**)

Main elements of the data cycle (2)

Multilingual (Open) Data infrastructure

- Development of European Digital Service Infrastructure and fostering new services in relation to
  - Open Data portals at local, regional and national and European level
  - "Multilingual access to online services"

- Leading by best practice examples …

… the *European data portals*

- ***European Commission data portal*** *(2012)*

- ***Pan-European open data portal*** *(2013)*
  - Multilingual access point to data from across the EU, to be funded through CEF (Connecting Facility for Europe)

- ***Benefits***
  - Scale
  - Interoperability of datasets
  - Easy to find across languages
  - Similar basic use conditions

Financing and support measures for data infrastructures

- Hackatons and competitions to foster setting up of data portals and applications: prices, FP7 and CIP funding
- Pan-European data portal: single access point to datasets from across the EU, expected launch 2013 (building upon FP7 funded R&D work)
- Support for inception phase (2012-2013): CIP calls
- Support 2014-2020: Connecting Europe Facility (expected total funds for ICT/Digital: €9.2 billion)

CEF digital service infrastructure for data

- ***"Core service platform"***
  - Distributed system
  - Query and visualization tools
  - Open source
  - Governance model involving the data providers
- ***"Generic services"***
  - Aggregation of datasets
  - Interoperability of datasets
  - Interface to open data infrastructures in third countries
  - Data repositories and long-term preservation services

Main elements of the data cycle (3)

Supporting Research and innovation which fosters the intelligent use, management and reuse of **complex and large amount of data**

for

- better decision making
- efficiency
- knowledge management
- extraction of embedded intelligence and data insights.

including

- R&D in Multilingual data and content analytics
- Innovation in Data driven intelligence and knowledge management in data intensive sectors

Main research challenges in
(Calls 1+3+5+SME+8; 81 projects)

- Content creation & processing (including multimedia and games)
- Development of media post-production tools
- Integration of social software & semantics
- Personalisation & summarisation
- Semantic foundations
- **Reasoning (temporal, dimensional and uncertainty, approximate & incomplete reasoning)**
- **Knowledge management in business & public-interest domains**
- **Copying with data explosion** ("big data" + real-time)

Closed FP7 Calls in figures

Inputs:	***Call 1***	***Call 3***	***Call 5***	***Call SME-DCL***	***Call 8***
Proposals	*148*	*252*	*169*	*343***	*139*
Participants	*210*	*2017*	*1387*	*1748*	*1128*
Request M€	*473*	*817*	*611*	*536*	*470*
Available M€	*51*	*50*	*70*	*35*	*50*
Outputs:	***Call 1***	***Call 3***	***Call 5***	***Call SME-DCL***	***Call 8***
Projects	*15*	*13*	*17**	*20***	*16****
Participants	*128*	*106*	*148*	*119*	*118*
Countries	*21*	*21*	*22*	*24*	*20*

***Total**: 81 projects, 619 contractors, 256 M€*

** additional 3 enlargements of existing projects (10p)*

*** joint call with INFSO E1 organised in two stages*

**** proposals retained for negotiation*

FP7 Call SME-DCL focus:
Intelligent Information Management

Key work programme themes :

- **Bootstrapping a data economy**
- Community building and best practices
- **Sharing language resources**
- Building consensus and common services

Key dimensions:

data pooling for new service + focus on SME participation

FP7 data market experiment

*Framework Programme 7, ICT-2011.4.1 special call for **SME Digital Content and Languages**, budget 31M€*

344 stage 1 proposals (28 April 2011) 235 for data, 109 for language resources

65 stage 2 proposals (28 September 2011) 44 for data, 21 for language resources

***Funded**: 12 for data (2 support actions), 8 for language resources*

In a nut shell:

Application domains:

Business sector

- **Marketing**
- **Logistics**
- **Product dev.**
- **Financial dev.**
- **…**

Personal and social sphere

- Entertainment
- Personal applications
- Social networks
- …

Public services

- Health
- Education
- Culture
- Emergency management
- …

Societal challenges

- Science
- Transport
- Environment
- Smart cities
- GIS
- …

Technology challenges:

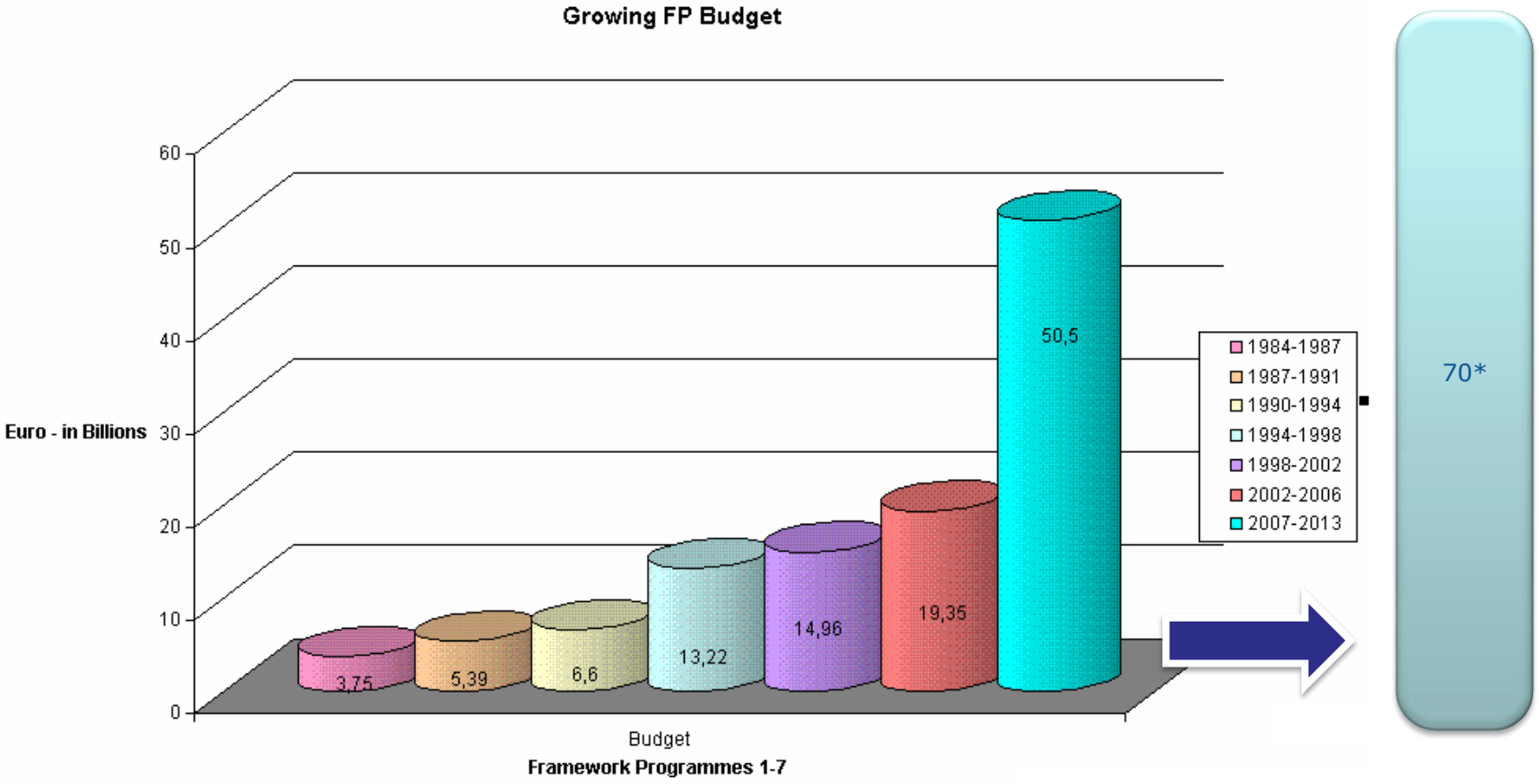
- Big data
- Semantics and reasoning
- Collaboration tools
- Multimedia and multimodal content

Financing and support measures of *R&D&I* enhancing new data-handling technologies:

- 2011-2013: ~ € 100 million
- one of priority areas envisaged for ICT in Horizon 2020 (2014-2020)
- support for ***technology innovation and uptake*** (pilot actions, testing, showcasing innovative applications): CIP-ICT PSP (2012)-2013 and Horizon 2020

Community Framework Programmes

Main EU instrument to fund Community research



***Proposal of the EC for 2014-2020**

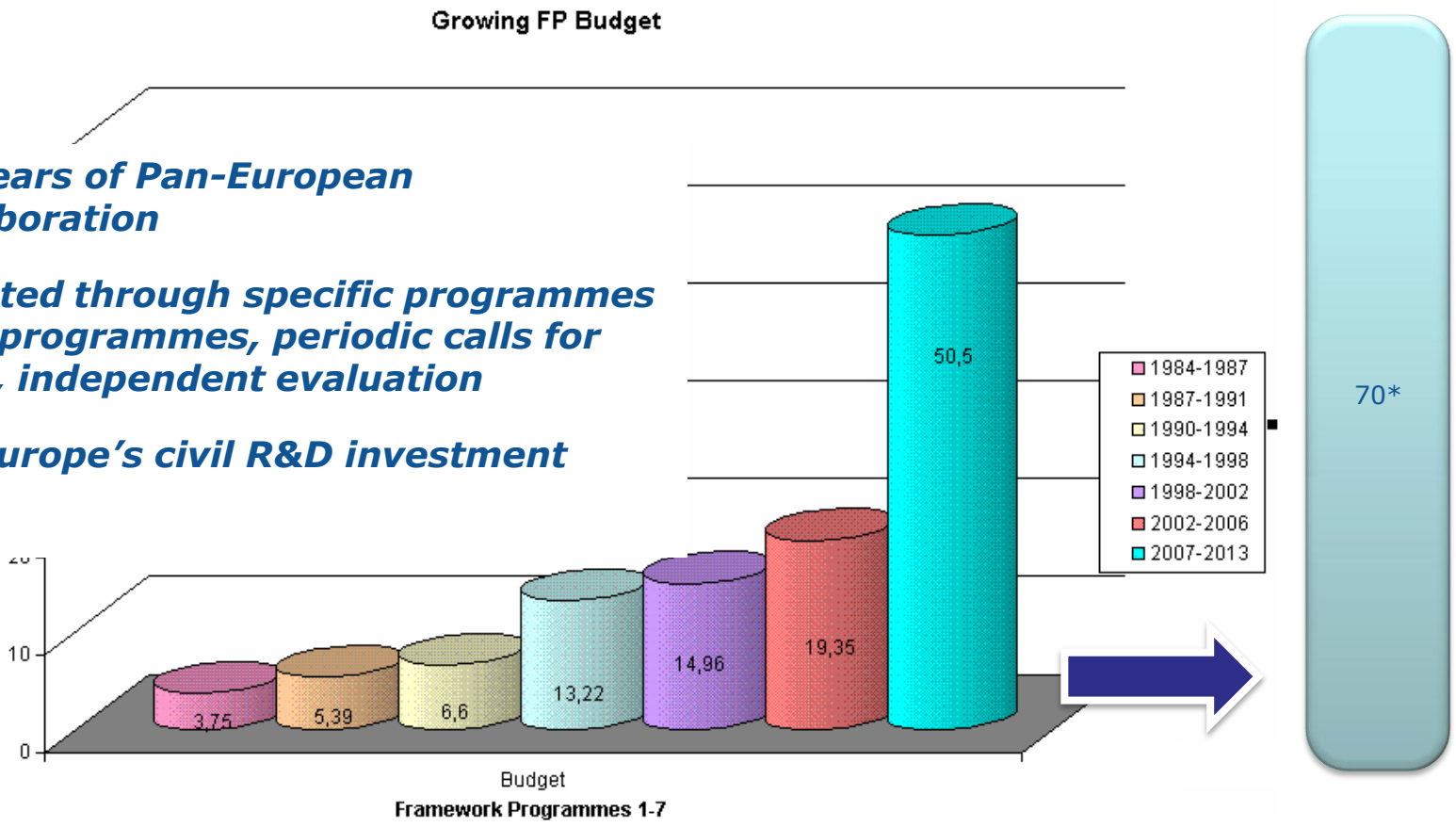
Community Framework Programmes

Main EU instrument to fund Community research

Over 20 years of Pan-European R&D collaboration

Implemented through specific programmes and work programmes, periodic calls for proposals, independent evaluation

~ 6% of Europe's civil R&D investment



***Proposal of the EC for 2014-2020**

Outlook - Horizon 2020

- Period from **1014 until 2020**
- 40% **budget**ary increase (Commission proposal)
- **Administrative simplification**: simpler funding rules; Open, light and fast schemes
- Higher **integration** between **R&D** and **innovation**
- **Roadmap** based research (from projects to programmes)
- "Data" will have even more importance

Main areas of Horizon 2020:

- Excellent Science

- Competitive Industries

- Better Society

Further important elements:

- facilitate access to risk finance;

- provide Union wide support for innovation in SMEs

EC proposal:

http://ec.europa.eu/research/horizon2020/index_en.cfm

European Data Value Chain

Data Value Chain

Who/what is producing data in the EU?

How is data captured?

How is data shared?

How is data reused?

What is the relevant maturity of these technologies and processes?

How deep and diverse is the EU data supply chain?

Where EU Data comes from

Finance
Geospatial
Energy
Logistics
Transportation
Science
….

Telecommunications
Pharma
Manufacturing
Retail
Government (open data)
Personal/Social
…

Where EU Data comes from

Humans (growing slowly)

- **Professional**
- **Personal/social**

Machines (growing fast)

- **Earth observation (satellite/env monitoring)**
- **Scientific equipment (DNA sequencers, telescopes, particle accelerators, …)**
- **Transport (air, trains, ships, cars, …)**
- **Sensors (Internet of Things, Smart Cities)**
- **Industrial machinery**
- **Robots (UAVs, smart spaces)**

How is data captured?

Humans (growing slowly)

- Benefits from **agreement on formats and semantics**
- Need for **better software support for data editing, validation, linking**

Machines (growing fast)

- Benefits from **agreement on formats and semantics**
- Benefits from **massive cross-industry exchange/reuse of entity identifiers (e.g. OKKAM)**

How is data shared?

Open data / Open access

- **Publication of (non private) government data for transparency and reuse (e.g. UK's Open Data Institute http://theodi.org/drupal7/node/6)**

- **Publication of scientific data for replication and reuse (e.g. http://www.openphacts.org)**

How is data shared?

Data Markets

- **When N actors all need the same data resource** (e.g. list of all hospitals in EU)…
- **…building the resource N times is wasteful**
- Better build/maintain it only ONCE and **bring it to a data market where N buyers can buy** it at a fraction of production cost
- **For a EU common data market** we need to solve additional problem of data **language barriers**

EU Data Supply Chain

How deep and diverse is the EU data supply chain? Fostering healthy EU industrial data ecology

Many cross border suppliers/buyers of:

- **Raw data**
- **data storage**
- **data management, refinement, linking**
- **(real time) analytics, recommendations**
- **Visualization**

Healthy environment for SME innovation

EU Data Supply Chain

- A richly EU-wide connected data supply chain is a strategic objective

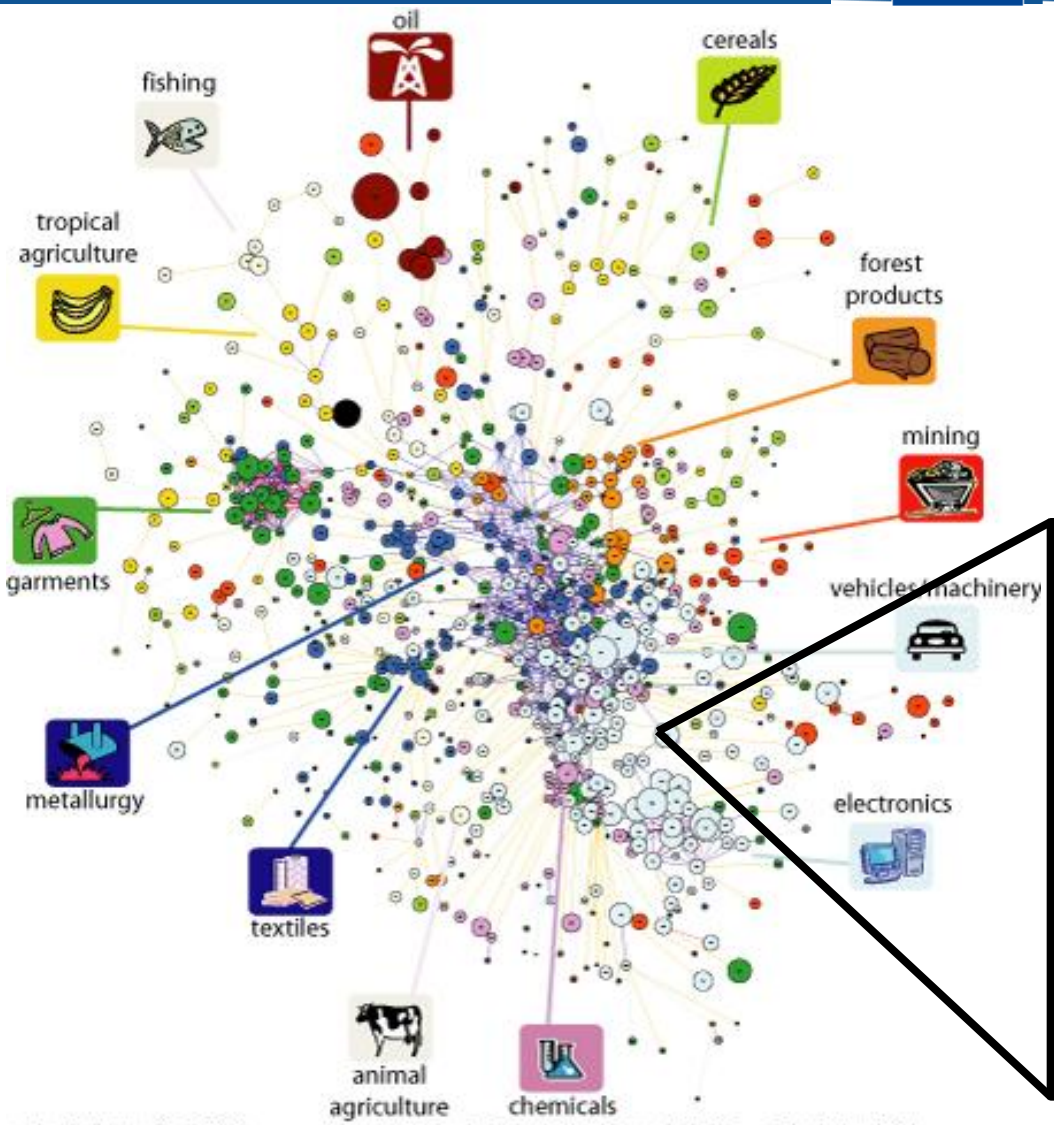
- Innovation means hopping quickly from established products to new product opportunities

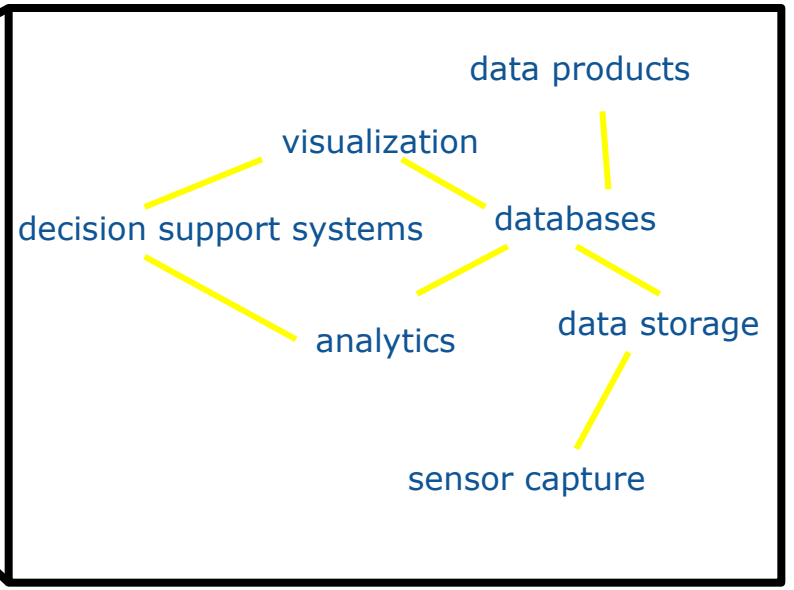
- Hopping requires availability of data skills, resources, infrastructures

- Future multilingual cross border services

Data Supply Chain and Product Spaces

Need a high value added data cluster





Trends

Decision makers rely on countless heterogeneous complex ICT systems …

Finance

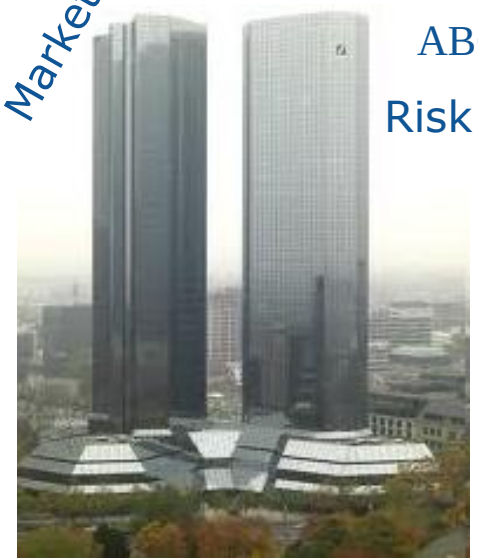
Stock markets

Accounting

Taxes *GL*

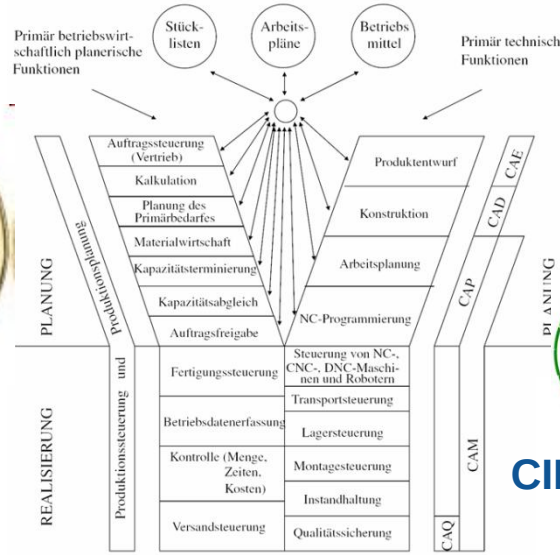
Cash-flow

Marketing



ABC / ABM

Risk analysis



Human resources

MTM

Assets management

CRM

DRM

Services

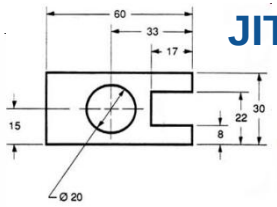
Bank accounts

PERT-CPM

SCADA

Industry

JIT



MPS

MRP

BOM

ERP

SPC

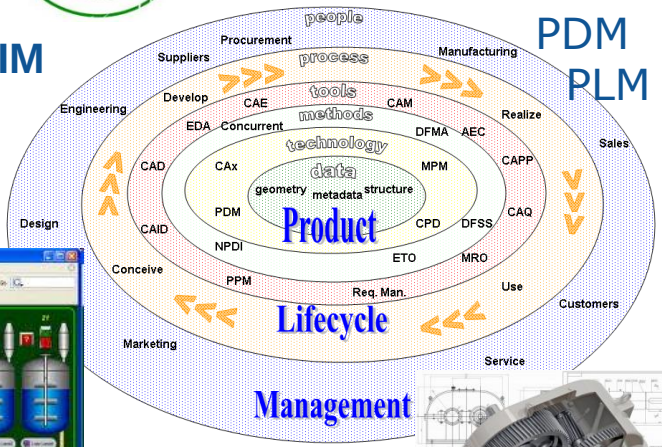
TQM

CIM

PDP

PDM

PLM



STEP XPDL

OEE CMMS

… but structured information represents only a small fraction of the whole picture

15%

85%

More than **85%** of all business **valuable information** exists in the form of e-mails, memos, notes from call-centres, news, user groups, chats, reports, web-pages, presentations, image-files, video-files, marketing material and news.

Source: Merrill Lynch

Main ICT challenges for data business

- Growth of organisational information
- Large and growing data quantity
- Multimodal information
- Unstructured data
- Heterogeneity of data and data sources
- Complexity
- Interoperability
- Enriching business data with open data
- External shocks, e.g. financial crisis

Trends

- Information overabundance creates fantastic **opportunities** for new business but also some **threats**
- The business champions of the future will be the most successful companies in **coping with data flood**
- Yet **technology solutions lags far behind** the complexity of information problems
- Legal issues (privacy, licensing, reuse) needs European actions
- The **EC is committed to support** to improve European data driven competitiveness

Proposers' day

Networking and information gathering event for all ICT calls of 2013

http://ec.europa.eu/information_society/events/ictproposersday/2012/index_en.htm

Conclusion

- *'Big data' has the future*
  - Applications and services + re-use
- *Open data strategy: towards a better use of publicly funded data in Europe*
- *European digital service infrastructure for data will help unleashing the potential*

Further info

- ***ICT under FP7***
  http://cordis.europa.eu/fp7/ict/
- ***Experts data base:***
  https://cordis.europa.eu/emmfp7/
- ***Unit*** *– Technologies for Information Management*

  URL**: http://cordis.europa.eu/info-management/**

  eMail to**: infso-e2@ec.europa.eu**

Thank you!