Co-funded by the European Union under FP7-ICT-2009-6

APA RSEN

Alliance Permanent Access to the Records of Science in Europe Network

Alliance Permanent Access to the Records of Science in Europe Network

Digital Preservation Sustainability on EU policy level

London, Sept 2014

David Giaretta

APARSEN/APA

Co-funded by the European Union under FP7-ICT-2009-6

EC policy – a brief history – a personal view

- EC support for
  - DP research
  - for creating digital objects
    - Data
    - Digitisation
  - e-Infrastructure

  to
  - Digital Agenda
- National funding
  - Significantly more than EC funding

What is the EC role?

DP research: approx 100M€ from EC

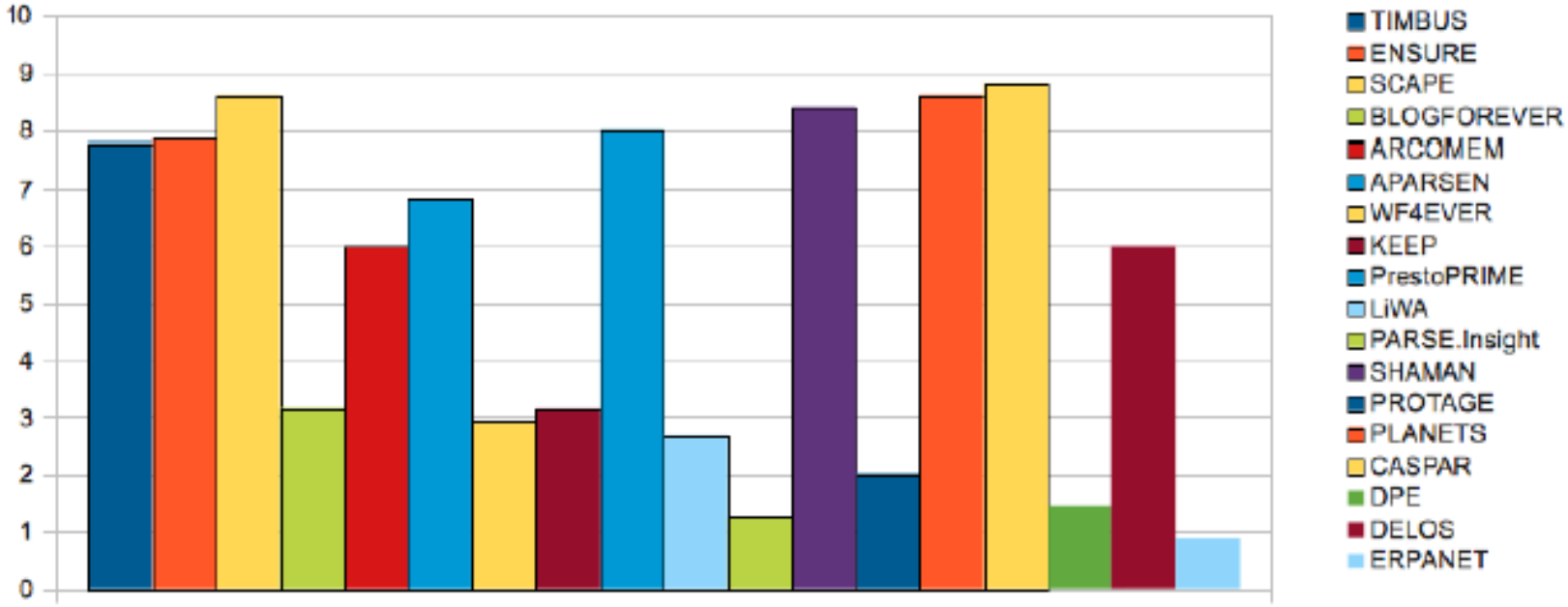


Figure 1: EU funding for all digital preservation projects in (mio) EUR

From *Research on Digital Preservation within projects co-funded by the European Union in the ICT programme, 2011, Stephan Strodl et al* http://cordis.europa.eu/fp7/ict/creativity/report-research-digital-preservation_en.pdf

Situation now

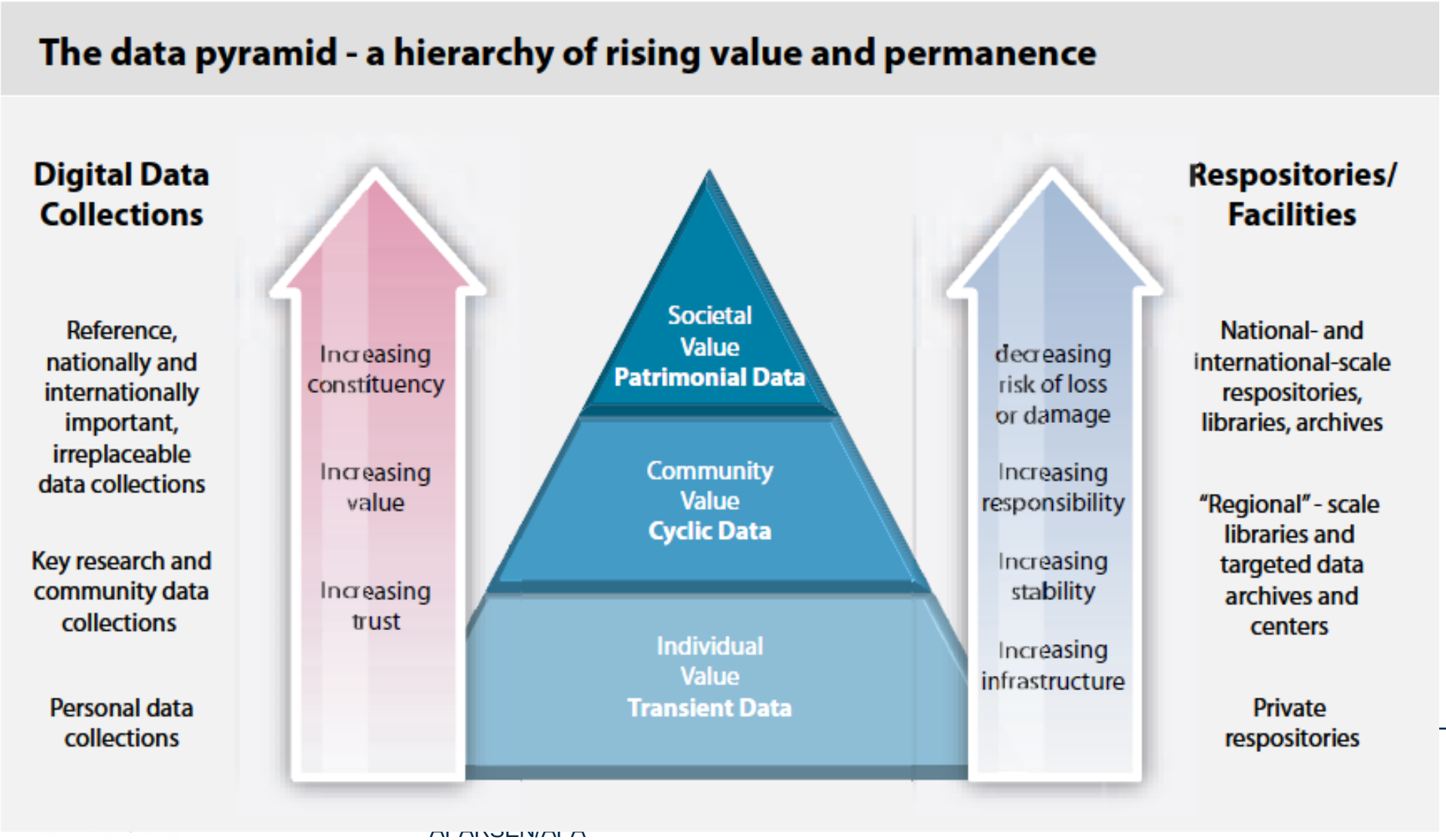
- As a community we have failed in persuading the EC that there is need for more funding for DP research
  - We do not have a consistent story about:
    - Costs
    - Rights
    - Methods etc
      - + “Emulate or Migrate” inadequate!
    - Who is doing it right
- Luxembourg unit which previously funded DP research – name changed to “Creativity” - now shows **no funding** for digital preservation research
- EC expects results from the previous 100 M € research

Digital Preservation – some quotes:

- Head of unit funding the Digital Preservation projects asked repeatedly:
  - "Who pays and why?"
- NSF colleague:
  - "Digital preservation is like VAT – people don't like it"

Value pyramid

The data pyramid - a hierarchy of rising value and permanence



"The Digital Agenda for Europe outlines policies and actions to maximise the benefit of the digital revolution for all. Supporting research and innovation is a key priority of the Agenda, essential if we want to establish a flourishing digital economy."

Neelie Kroes,

Vice-President of the EC, responsible for the Digital Agenda

Data is the new gold.

"We have a huge goldmine... Let's start mining it."

Neelie Kroes

That is the magic to find value amid the mass of data. The right infrastructure, the right networks, the right computing capacity and, last but not least, the right analysis methods and algorithms help us break through the mountains of rock to find the gold within.

…….but

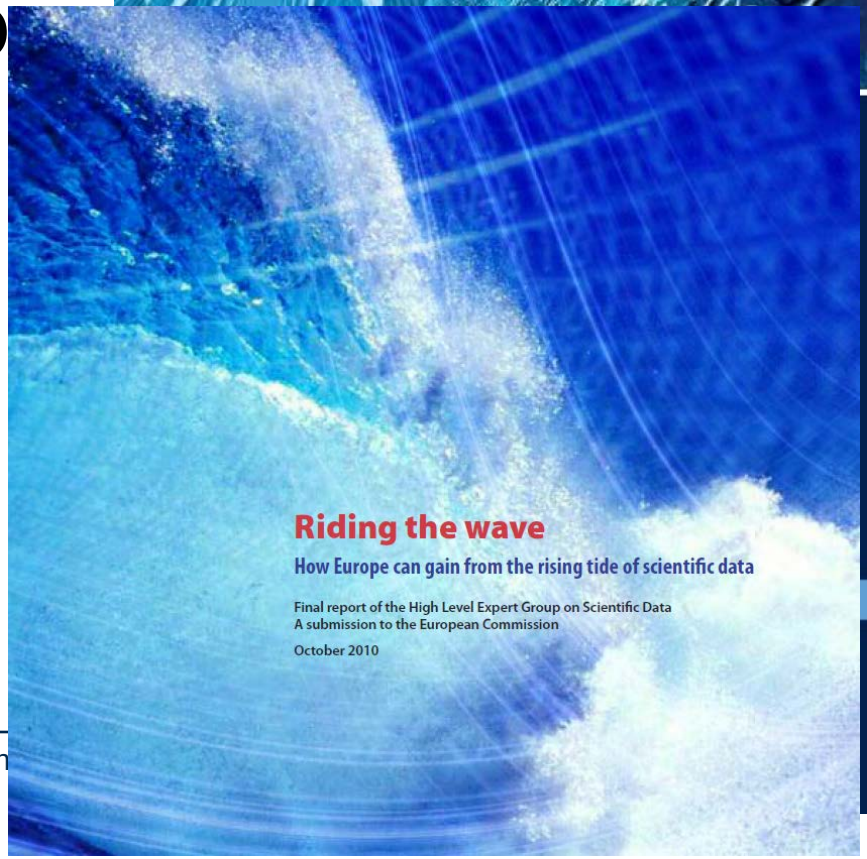
- Gold is precious because
  - it is rare
  - it does not combine with other elements
  - it does not perish

……..but…….

- Data is valuable because
  - there is so much of it
  - it is more valuable when it is combined together
  - BUT it is far from imperishable

Co-funded by the European Union under FP7-ICT-2009-6

O



TRAINING TRAINING TRAINING TRAINING
terminology plus extension
Business case
Business Model
Preservation
Usability
Value
Services
TOOLS
COSTS
SELECTION
GOVERNANCE
RIGHTS
AUTHENTICI
INGEST
STORE
PLAN
REP. INFORMATION
TRANSFORM
HANDOVER
Dispose
Use
Audit

H2020 EINFRA-1-2014 – Managing, preserving and computing with big research data

- **Specific challenge**: Development and deployment of integrated, secure, permanent, on-demand service-driven, privacy-compliant and sustainable e-infrastructures incorporating advanced computing resources and software are essential in order to increase the capacity to manage, store and **analyse extremely large, heterogeneous and complex datasets**, including text mining of large corpora. These e-infrastructures need to provide services **cutting across a wide-range of scientific communities** and addressing a diversity of computational requirements, legal constraints and requirements, system and service architectures, formats, types, vocabularies and legacy practices of scientific communities that generate, analyse and use the data.

- Establishing a **federated** pan-European **data e-infrastructure** to provide **cost-effective** and **interoperable** solutions for data management and **long term preservation**.
- Services to ensure the quality and reliability of the e-infrastructure, including **certification mechanisms for repositories** and certification services to test and benchmark capabilities in terms of resilience and service continuity of e-infrastructures;

Impact expected

- Increased availability of scientific data for scientific communities independently of them having already embraced or not e-science; **this will be measured by cross-border data traffic over the research networks in Europe as a proxy**.

Co-funded by the European Union under FP7-ICT-2009-6

So what next

- EC will not commit to sustaining digital preservation over the long term
- They expect “Data is the new gold” / Fourth Paradigm/ Riding the Wave
- We need to be part of the solution to value/benefits
  - We do not have to show immediate value for every single digital object.
  - A rising tide floats all boats