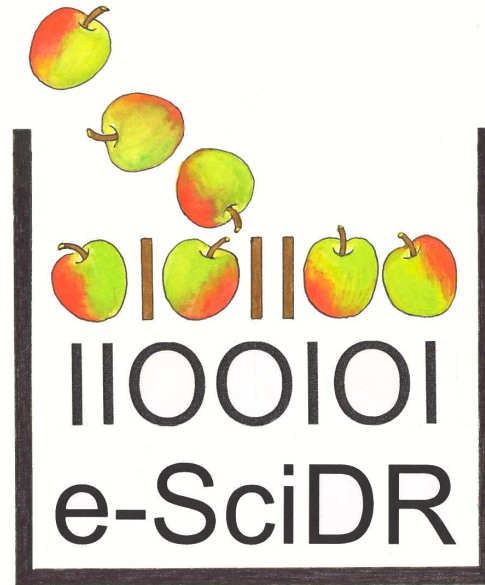# Towards a European e-Infrastructure for **e-Sci**ence **D**igital **R**epositories

Harvesting and seeding the fruits of e-Science

e-IRG Workshop, Lisbon, 11-12th October, 2007

e-SciDR study background

- Sponsored by DG INFSO by the European Commission

- To provide an overview of the e-Science digital repository situation in Europe

- Recommend policy options for supporting and advancing the use of digital repositories

- Contexts:
  - i2010 Action Plan
  - Building the Single European Information Space
  - Input into the FP7 Capacities Programme

The study team

- Digital Archiving Consultancy (DAC)
- Charles Beagrie Limited
- GridwiseTech
- National e-Science Centre (NeSC) University of Glasgow
- Internet Centre, Imperial College
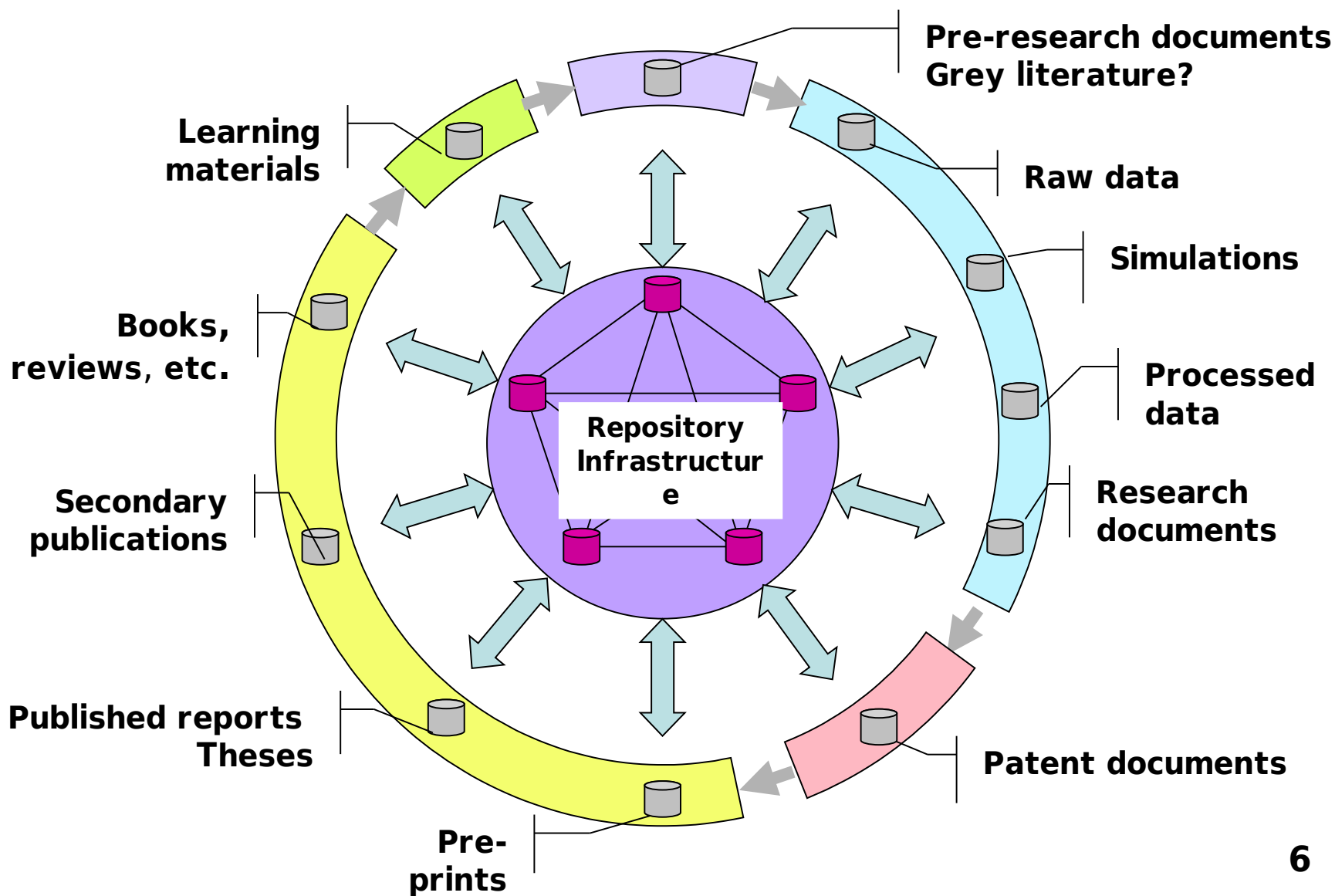- Com'tou sarl

- European Commission

Study timetable

- Started in January 2007
- Initial 3 workshops with invited experts
- Interviews with key informants
- Desk research
- An on-line public consultation (questionnaire), July-August
- Workshop 4th September in Lisbon to consider outcomes
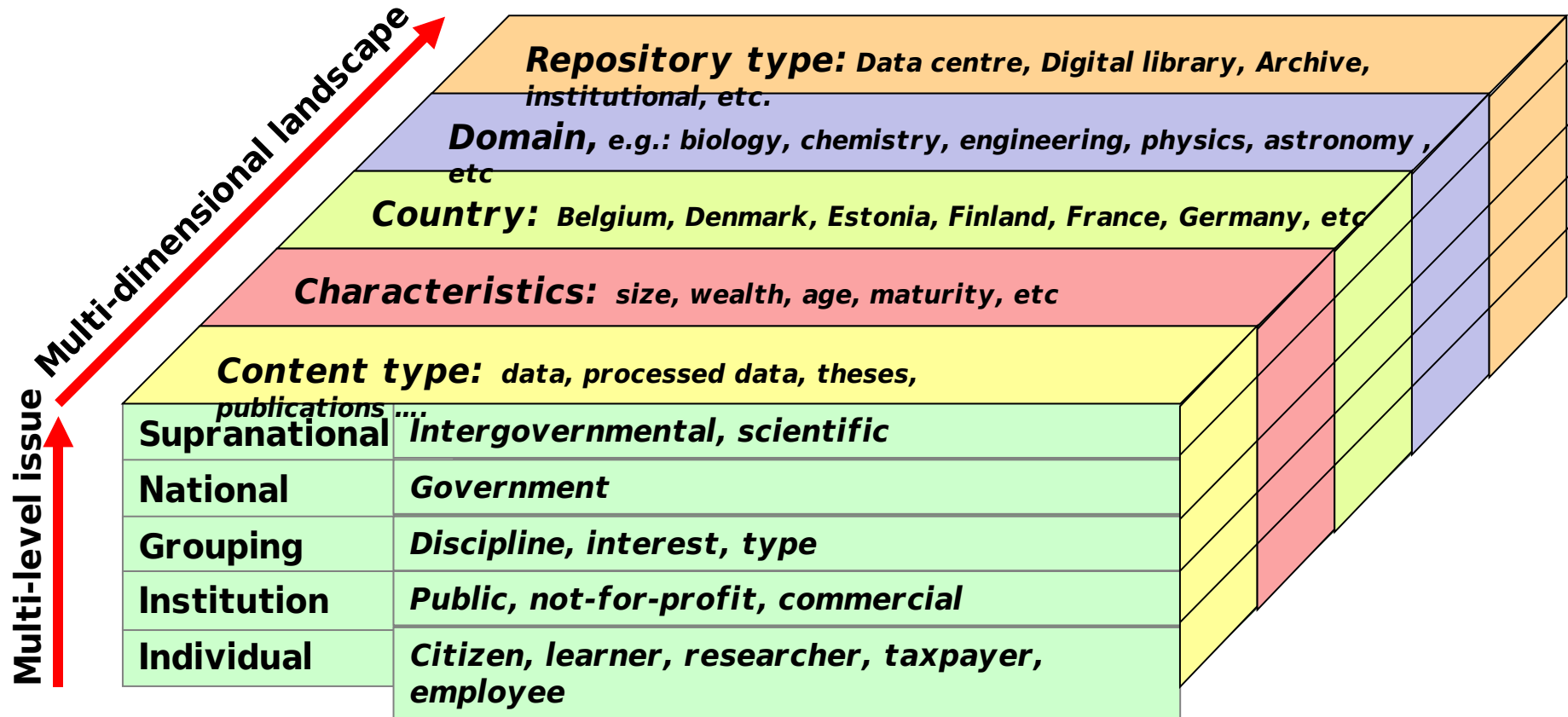- Final report near completion

Study scope

- A **top-level** study of a **vast area**
- Scope widely drawn:
  - All repository types
  - All content types
  - The full research cycle
  - Wissenschaft – the full range of research disciplines
- e-Science and e-Infrastructures
- Issues up front – sustainability and interoperability

The full research cycle

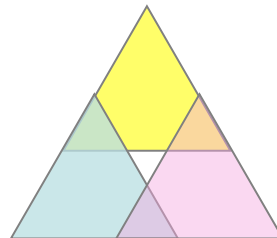Pre-research documents
Grey literature?

Raw data

Simulations

Processed data

Research documents

Patent documents

Pre-prints

Published reports
Theses

Secondary publications

Books, reviews, etc.

Learning materials

Repository Infrastructure

Multiplicity of stakeholders

Multi-dimensional landscape

Multi-level issue

Repository type: *Data centre, Digital library, Archive, institutional, etc.*

Domain, *e.g.: biology, chemistry, engineering, physics, astronomy , etc*

Country: *Belgium, Denmark, Estonia, Finland, France, Germany, etc*

Characteristics: *size, wealth, age, maturity, etc*

Content type: *data, processed data, theses, publications ….*

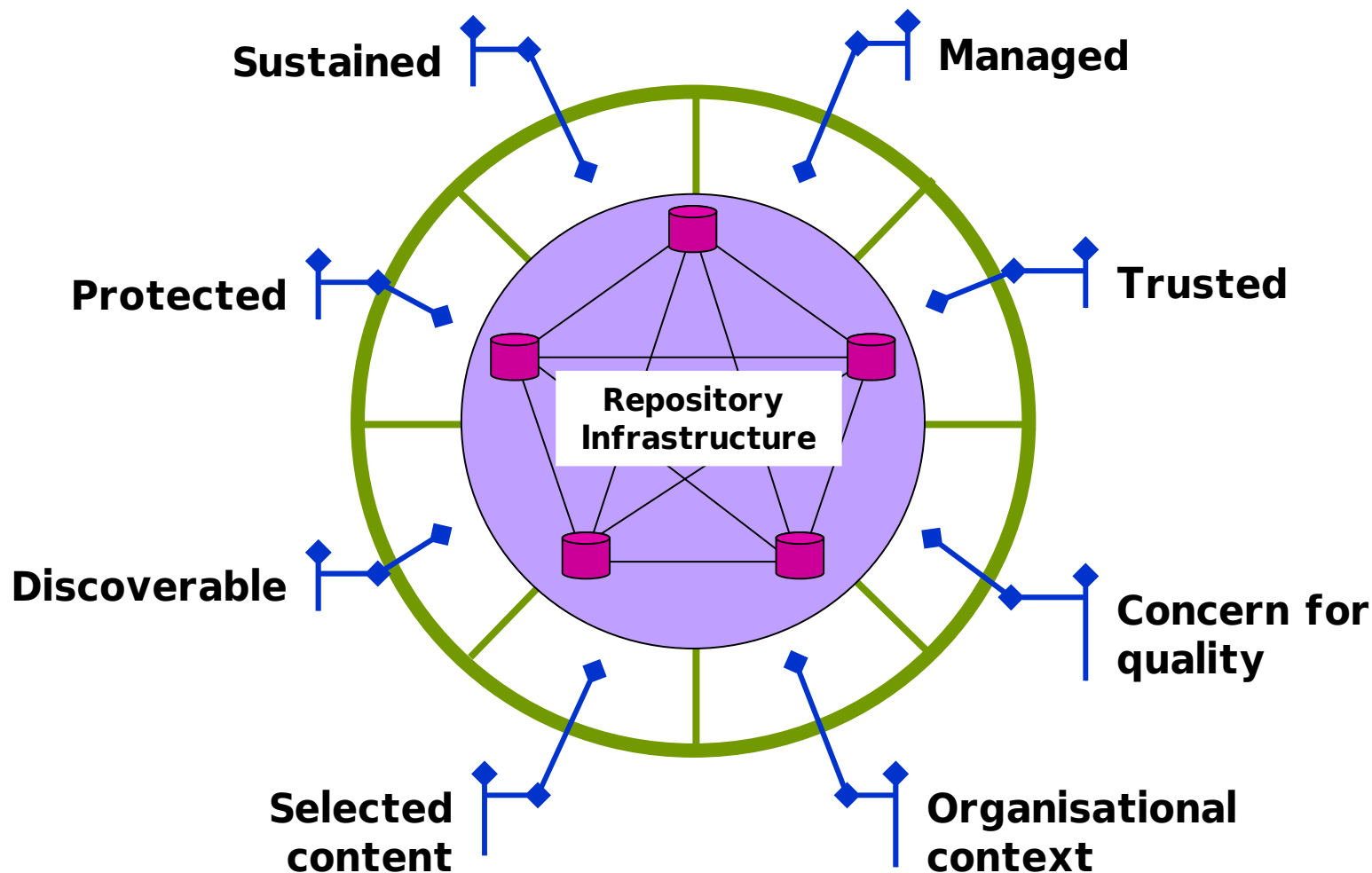| Supranational	*Intergovernmental, scientific*
National	*Government*
Grouping	*Discipline, interest, type*
Institution	*Public, not-for-profit, commercial*
Individual	*Citizen, learner, researcher, taxpayer, employee*

What is a repository?

- Not easy to define!
- Some basic (non-defining) architectural components:
  - Storage for content and metadata
  - Input / output and access controls
  - Connectivity
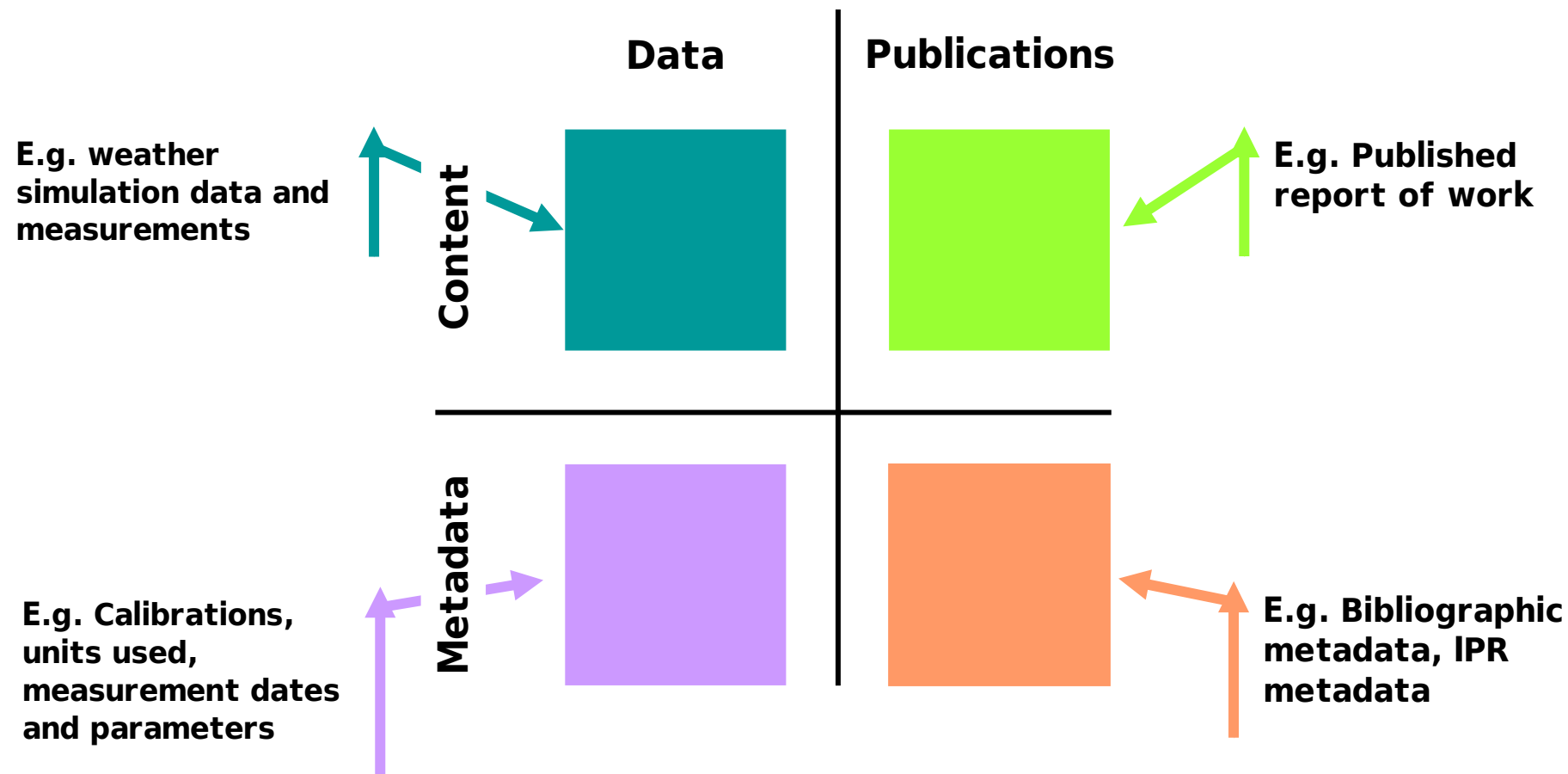- Many characteristics – no one defining subset:

Defining characteristics

What information types?

	Data	**Publications**
Content		
Metadata		

E.g. weather simulation data and measurements

E.g. Published report of work

E.g. Calibrations, units used, measurement dates and parameters

E.g. Bibliographic metadata, IPR metadata

Public consultation

- Mid July to end August
- Dissemination by mailings, lists, websites
- 426 responses, world-wide
- Excellent range of responses, from policy makers, experts, managers, users
- High-quality free-text feedback
- High-level people

Consultation highlights

- Primarily users of repositories
- Researchers the largest group
- Geographical hot spots?
- Only a minority had training
- Difficulties in repository use were:
  - Time-consuming to use
  - Do not know where to look for information
  - More accurate search tools needed
  - Metadata generation tools needed

Findings : issues

2. Funding models for repositories are in general inappropriate, resulting in inefficient use of resources, overworked & under-resourced staff

3. Dark holes: Use of repositories is hampered by difficulties finding information, accessibility and ease of use

4. Capture of materials for further use: materials are being lost. On the other hand, the increasing data (and metadata) deluge challenges management & sustainability

5. There are concerns about the quality as well as quantity of data captured – particularly capturing all-important metadata

Findings and issues

1. Incentives and cultural constraints inhibit deposit; and wider training is needed, including across professions (librarians, computer scientists, scientists)

2. Valuable scientific outputs will be lost unless sustainability and preservation frameworks are developed

3. Organisational immaturity and fragmentation inhibit use and trust

4. Repository curators, manager lack professional incentives and recognition of the importance of their work

Findings and issues

1. Science and information exchange operate globally, but at the same time there are areas of inadequate in provision in Europe.

- Data collections are often global, so European repositories part of global framework

Draft recommendations: Vision - 1

- Support the scientist at all points in the research cycle by providing easy, cost-effective access in a joined-up fashion to materials of all types that are already available (subject to well understood precautions in respect of ownership, privacy and ethical use)

- Support easy and reliable deposit of materials for science, research and learning into known, trusted repositories through the whole research cycle, providing confidence that the materials will be well maintained, and not abused.

Draft recommendations: Vision - 2

- The collections in repositories are expertly maintained
- The repositories should have a capacity or associated framework to:
  - support the long-term sustainability of information
  - be trusted
  - guarantee the authenticity of stored materials and cope with future demand
- The infrastructure delivers or supports services equally across the whole of Europe and participates as leaders and partners in the wider global e-science information infrastructure

Draft recommendations: Vision - 3

■ The various stakeholders - administrations, the scientific community, the private sector and the public - have well-founded confidence that the infrastructure is:

- reliable, delivers value for money
- can adapt to change as technologies and science move on and
- that it continues to collect and preserve securely Europe's great scientific heritage.

Policy recommendations

- Highly cross-cutting
- 20 areas identified, grouped and prioritised under major policy headings
- To be cast as policies for the European Commission's consideration and action
- Grouped by:
  - Funding reform
  - A European Research Information Space (ERIS)
  - Harmonisation – internal and external

Recommendation headings*

- Funding on a long-term or rolling basis, specifically for repositories; including funding for tools, services

- Reporting, governance and awareness actions, for sustainability

- Actions and funding to support data producers – for good data management

- Publicly funded activity should mandate **submission** of outputs to accessible repositories

(* As discussed at the final workshop, Lisbon, September 4th 2007.)

Recommendation headings

1. Establish selection and appraisal techniques and tools
2. Support for data preservation activities
3. A European multi-lingual gateway to repository resources
4. Establish centres of repository excellence
5. A European-level repository facility

Recommendation headings

1. Action to improve discovery across and at repository, collection and item levels
2. Support work to enable linking through the whole research chain from raw data to publication
3. Help establish data citation mechanisms
4. Harmonisation of access, authorisation and authentication across Europe
5. Promote & provide training activities

Actions recommended

1. Establish career structures for repository practitioners and managers
2. Establish networking and information exchange facilities regarding repositories
3. IPR training & advice; harmonisation of applicable laws , e.g. copyright
4. Institute certification schemes for repositories
5. Repository integration with the wider world

Towards a European e-Infrastructure for **e-Sci**ence **D**igital **R**epositories
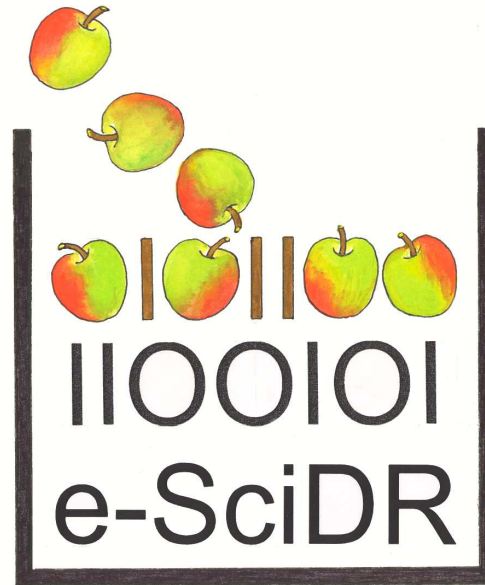
Harvesting and seeding the fruits of e-Science

© **Digital Archiving Consultancy Limited**