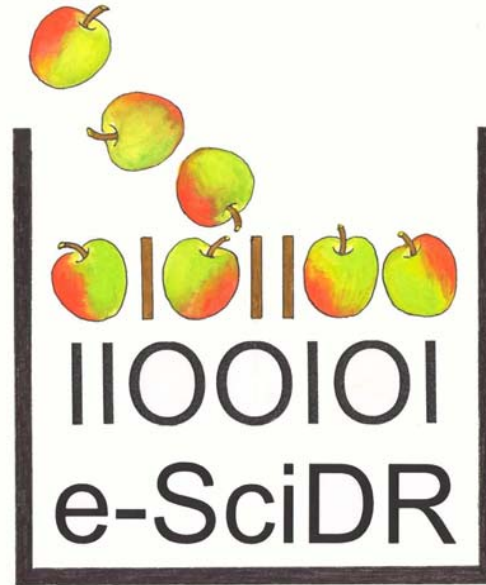




Towards a European e-Infrastructure for **e-Science Digital Repositories**



Harvesting the fruits of e-Science

Final Workshop, Lisbon, 4th September, 2007



The study team

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

- European Commission



People

- Alison Macdonald
- Brian Fuchs
- Damian Counsell
- Daphne Charles
- Dolores Iorizzo
- Hanne Mostecky
- John Darlington
- Krzysztof Wilk
- Melanie Dulong de Rosnay
- Neil Beagrie
- Philip Lord
- Pawel Jarosz
- Pawel Plaszczyk
- Richard Sinnott

European Commission:

- Carlos Morais-Pires
- Elina Zicmane
- Mario Campolargo

By first name



e-SciDR study background

- To provide an overview of the e-Science digital repository situation in Europe
- Recommend policy options for advancing repositories for consideration by the European Commission
- Contexts:
 - i2010 Action Plan
 - Building the European Information Space
 - Input into the FP7 Capacity Programme



Method

- Started in January 2007
- Initial 3 workshops with invited experts
- Interviews with key informants and desk research
- An on-line public consultation (questionnaire)
- This workshop to consider outcomes
- Final report by end September 2007 followed by dissemination

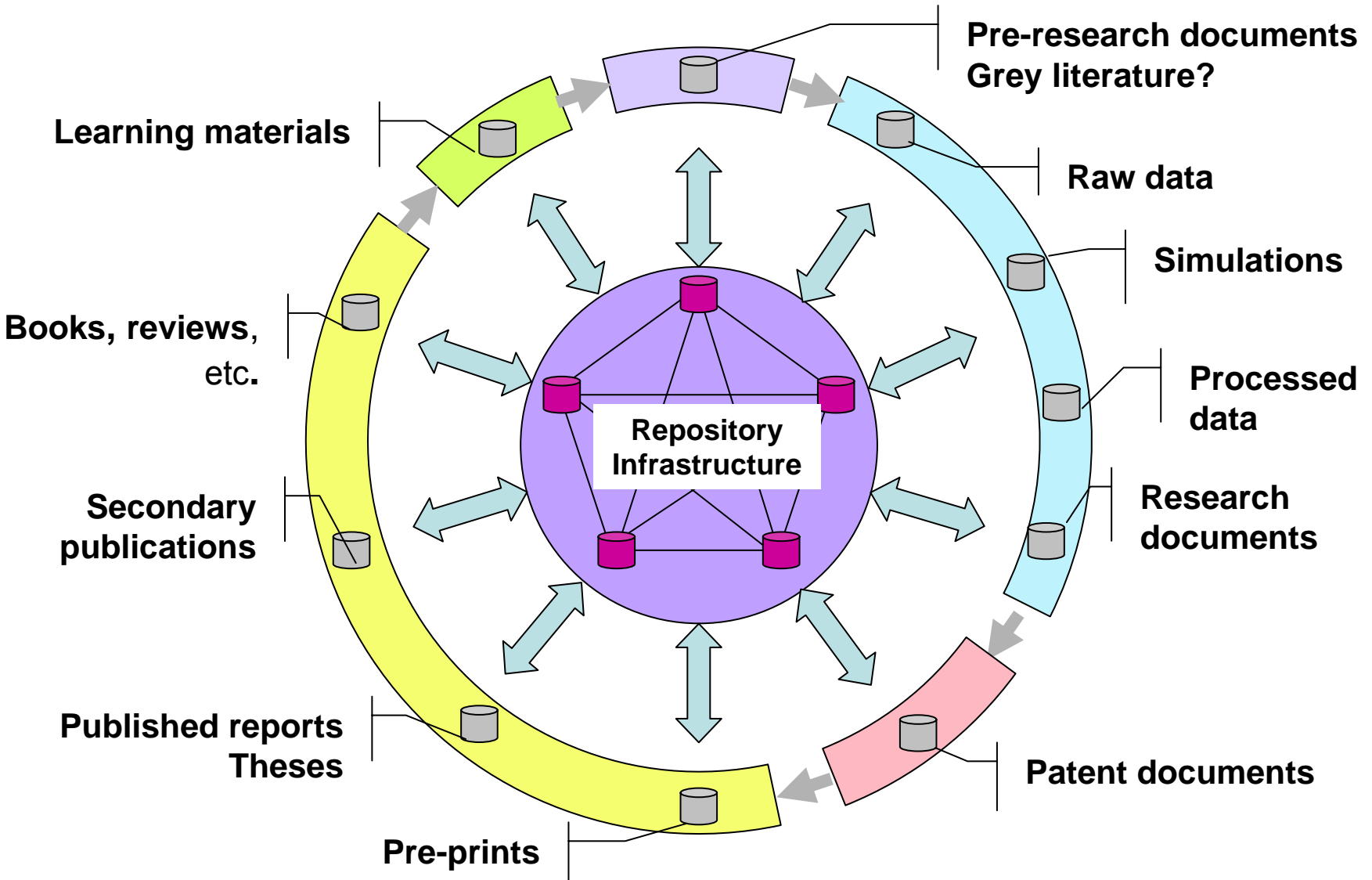


Study scope

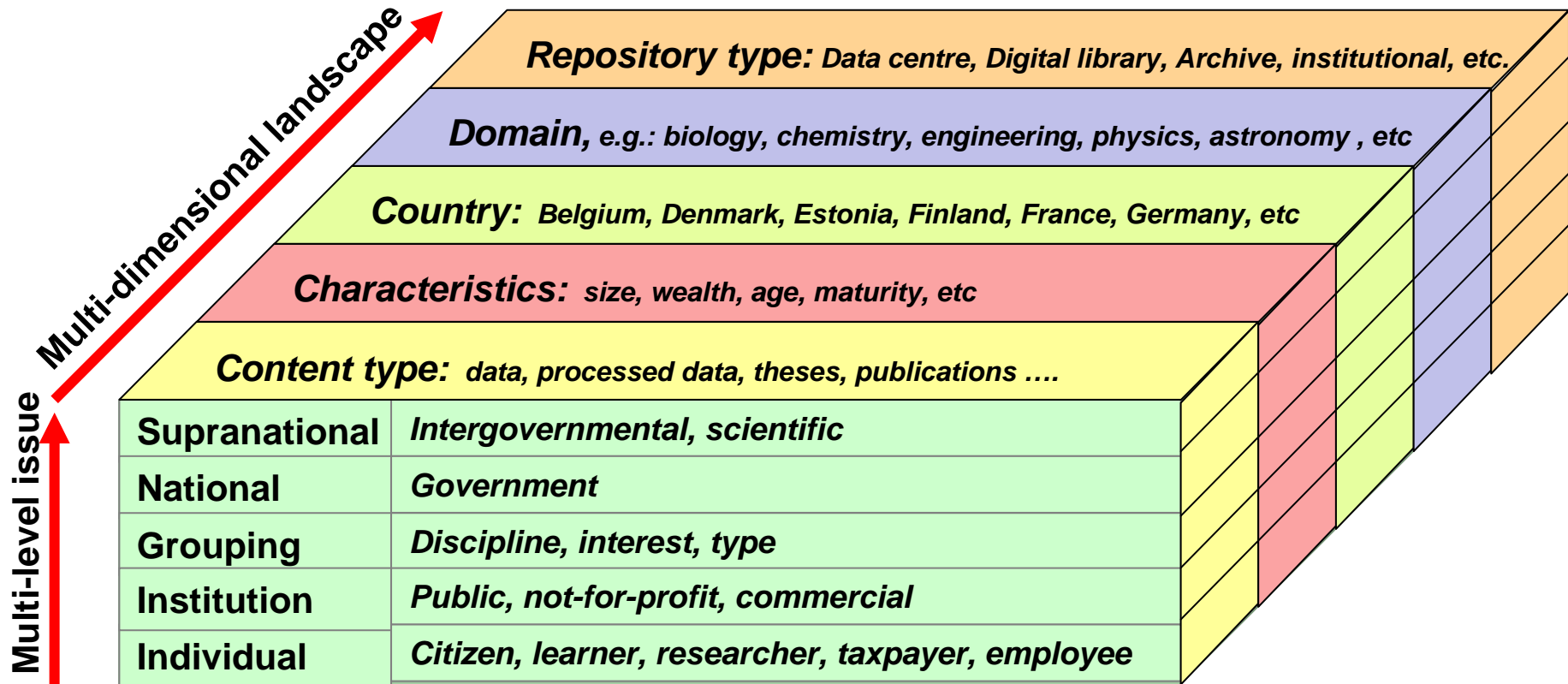
- **A top-level study of a vast area**
- **Scope widely drawn:**
 - All repository types
 - All content types
 - The full research cycle
 - Wissenschaft – the full range of research disciplines
- **e-Science and e-Infrastructures**



Research cycle & repositories

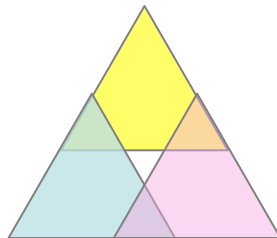


Multiplicity of stakeholders

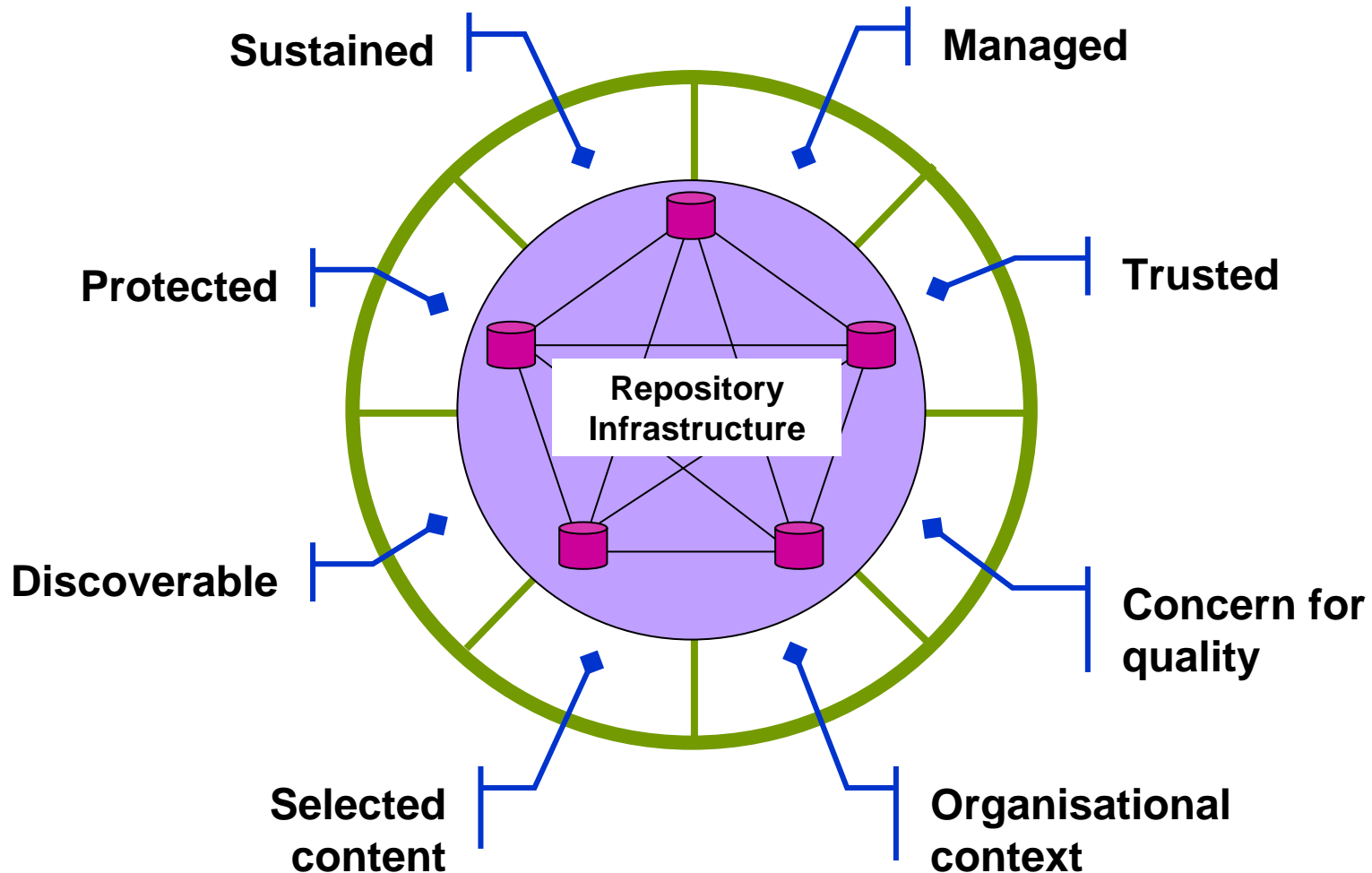


What is a repository?

- 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:

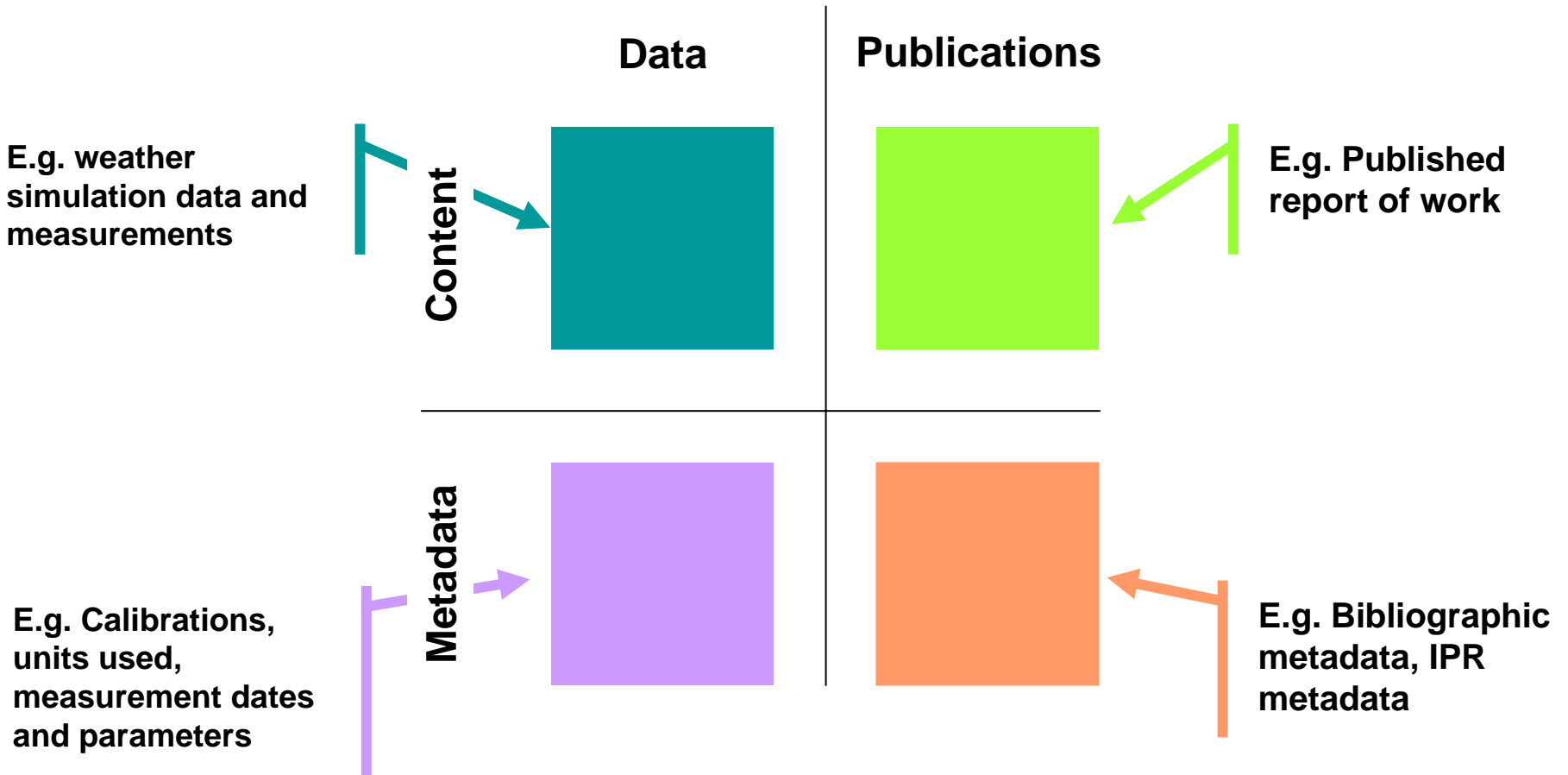


Defining characteristics



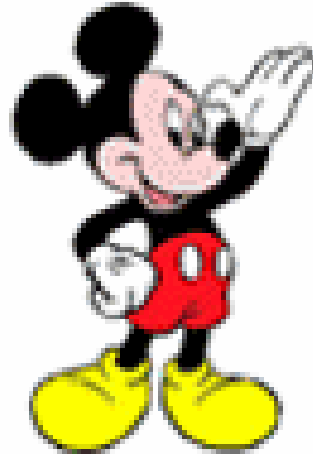


What information types?



Public consultation

- Mid July to end August
- Dissemination by mailings, lists, websites
- 426 responses, world-wide
- High-quality free-text feedback
- High-level people



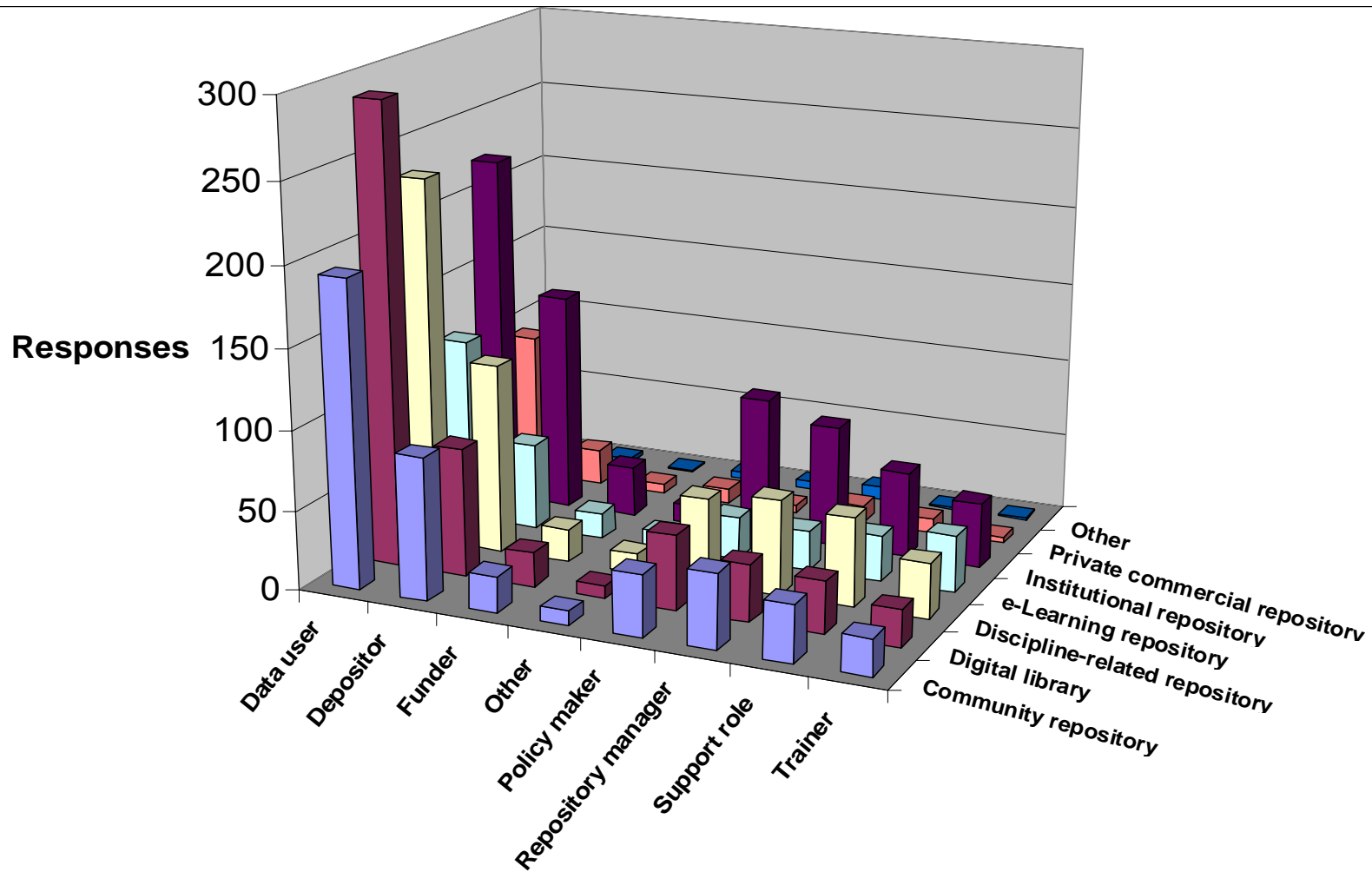


Consultation highlights

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



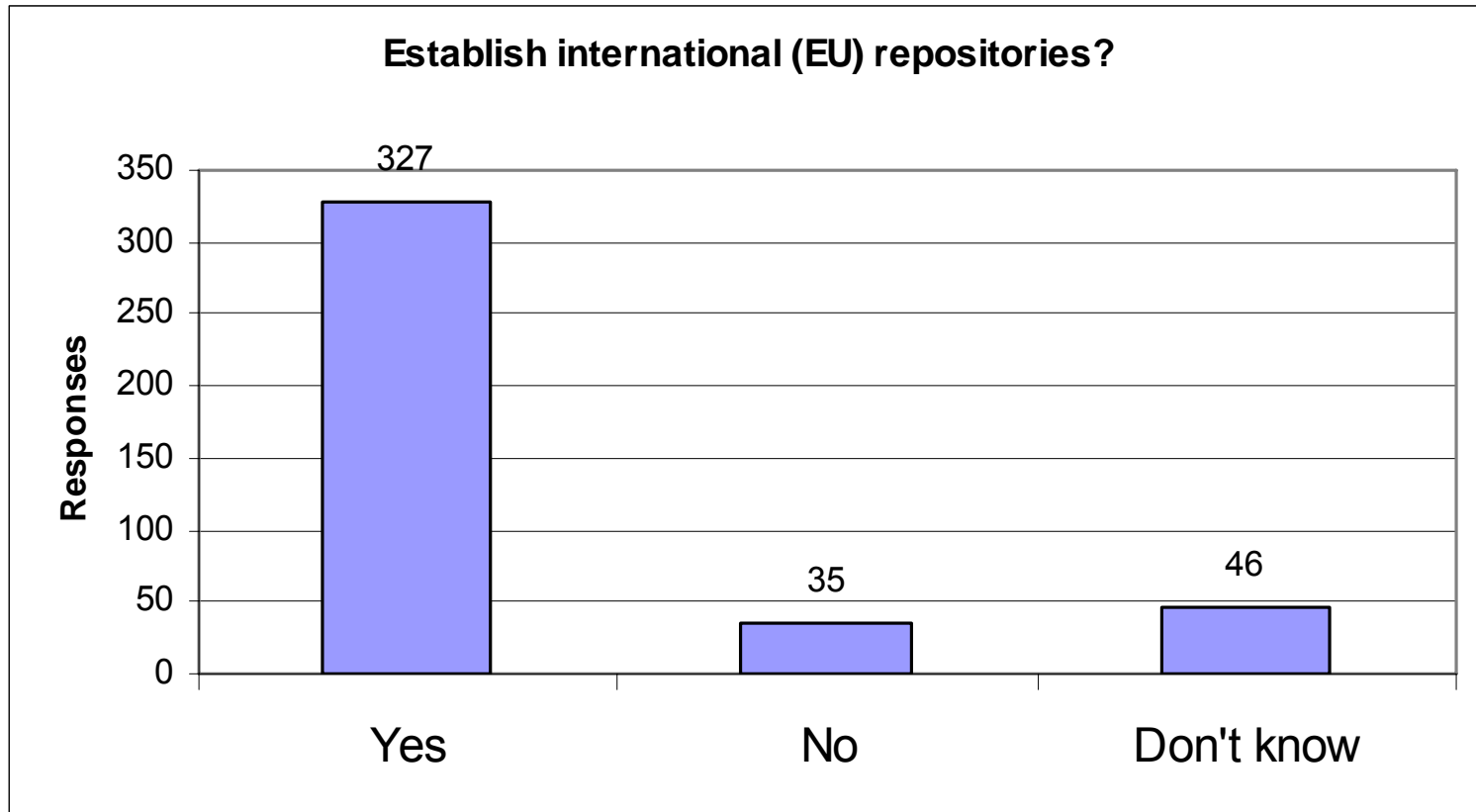
Repository role by type





European-level repository?

- 78% support for this





Issues raised - findings

- 23 areas identified
- A wide range of issues were raised from the findings
- Issues relate to:
 - Funding
 - Accessibility, searching and ease of use
 - Cultural and behavioural issues, including incentives, training
 - Sustainability and preservation
 - Policies, organisational structures, roles
 - Data management
 - Trustworthiness and authenticity
 - Selection and appraisal – dealing with the data deluge



Actions recommended

- 20 areas identified, to be group and prioritised under major policy headings in the final report
- To be cast as policies for the European Commission's consideration and action
- To be debated at this workshop before finalisation



Actions recommended

1. Funding on a long-term or rolling basis
2. Reporting, governance and awareness actions
3. Funding for services and tools on the same basis
4. Actions to support data producers
5. Publicly funded activity should mandate submission of outputs to accessible repositories



Actions recommended

6. A European multi-lingual gateway to repository resources
7. Establish centres of repository excellence
8. A European-level repository facility
9. Support for data preservation activities
10. Establish selection and appraisal techniques and tools



Actions recommended

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



Actions recommended

16. Establish career structures for repository practitioners and managers
17. Institute certification schemes for repositories
18. Establish networking and information exchange facilities regarding repositories
19. IPR training & advice; harmonisation of applicable laws , eg copyright
20. Provide repository support to the developing world



Other recommendations

- Rights management research – endorse fundamental review of copyright in digital information age
- Ease of use for range of users, lay to expert
- Micro-pricing study for identification & quantification of costs
- World data heritage designation



Parallel sessions

- Three parallel sessions, A, B C
- Colour coded:
 - Group A ●
 - Group B ●
 - Group C ●
- Chairs and rapporteurs appointed
- Tasks: See separate programme notes
- Time : 11:00 to 12:45 (after coffee break)
- Report back after lunch



Infrastructure

- “Infrastructure never gets adequately funded because it cuts across disciplinary boundaries, it doesn’t benefit particular groups. It is a prerequisite to great leaps forward and is thus never captured within disciplinary funding, or normal government operations. We need to revise radically our concept of cyberinfrastructure. ... It is a set of structures that network different areas of knowledge ... And that is software and social engineering, not fibre-optic cable. The superhighways of the biological information age ... need to be structures of knowledge.”
- *Professor James Boyle, Duke University Law School, USA*