

# PARSE.insight - Infrastructure Roadmap

**David Giaretta**



wePreserve

# Alliance for Permanent Access

- The Alliance aims to develop a shared vision and framework for a sustainable organisational infrastructure for permanent access to scientific information



- 1 November 2004 the Koninklijke Bibliotheek, National Library of the Netherlands, organized the international conference Permanent Access to the Records of Science within the framework of the Netherlands presidency of the EU.
- Mid 2005 the Task Force produced two documents:
  - a proposal for a European R&D programme in the field of long term preservation and
  - an overall Strategic Action Programme 2006-2010.
- In 2006 a small working group was formed to prepare the establishment of the formal Alliance for Permanent Access.
- Late 2007 – members sign up and pay membership
  - PARSE.Insight project formed from sub-group of the Alliance

The following organisations have so far joined the Alliance for Permanent Access:

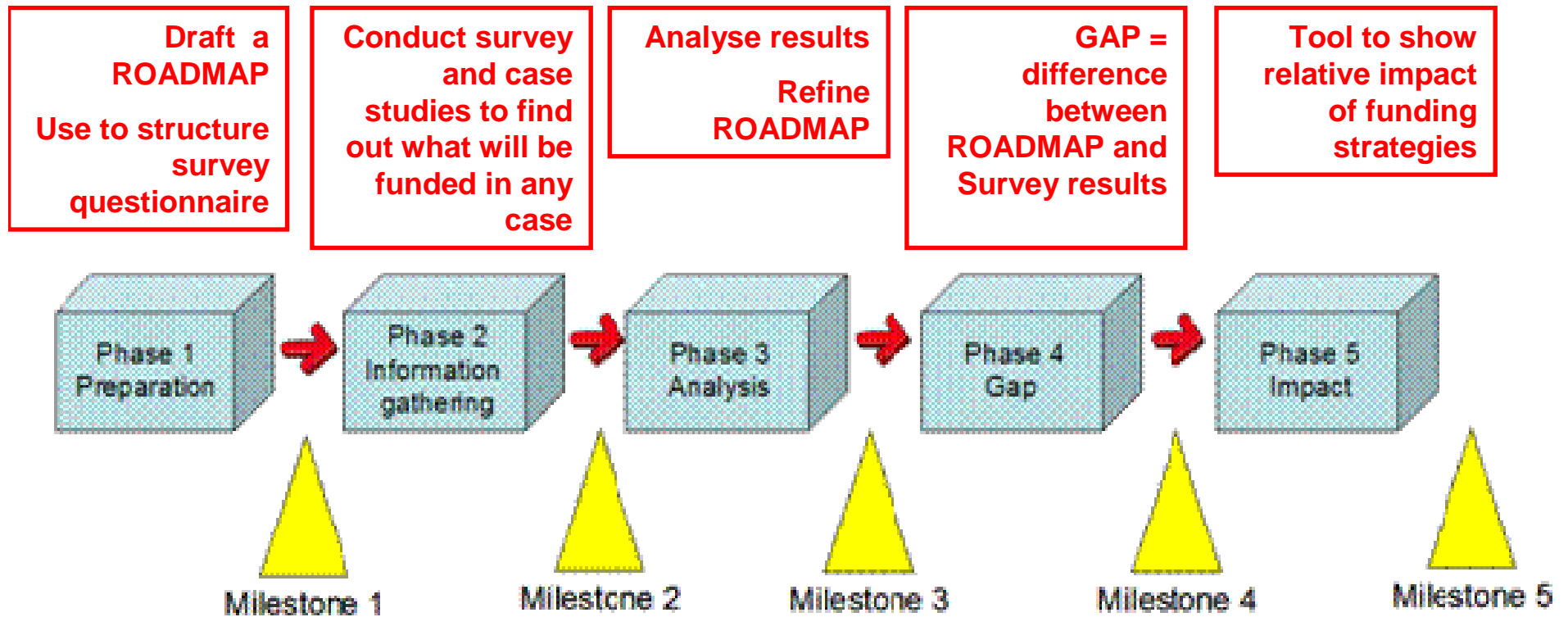
- The European Science Foundation
- European Space Agency
- CERN
- Max Planck Gesellschaft
- Science and Technology Facilities Council
- The British Library
- Koninklijke Bibliotheek
- Deutsche Nationalbibliothek
- Joint Information Systems Committee
- International Association of Scientific, Technical and Medical Publishers
- Digital Preservation Coalition
- NESTOR
- Netherlands Coalition for Digital Preservation
- Portico

Membership will be opened up to commercial organisations



- Funded by EU
  - 1.25 MEuro, 9 partners, 2 years
- Designed to help EU direct e-Sci funding in 2009/2010
- Officially started 1 March 2008
  - **Science and Technology Facilities Council** **UK**
  - **Koninklijke Bibliotheek** **NL**
  - **Deutsche Nationalbibliothek** **DE**
  - **Max Planck Gesellschaft** **DE**
  - **International Association of Scientific, Technical and Medical Publishers** **NL**
  - **European Space Agency, ESRIN** **FR**
  - **Fernuniversität in Hagen** **DE**
  - **European Organization for Nuclear Research** **CH**
  - **Georg-August-Universität Göttingen Stiftung Öffentlichen Rechts** **DE**

# Phases



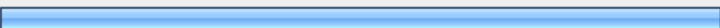


- On-line survey
  - Very broad distribution
    - Researchers – wide range of disciplines
    - Funders
    - Data managers/preservers
    - Publishers
  - NOT JUST FROM PARSE.Insight
    - From the Alliance members
- Follow up by interviews



# Initial results (after 20 hours – researchers survey)

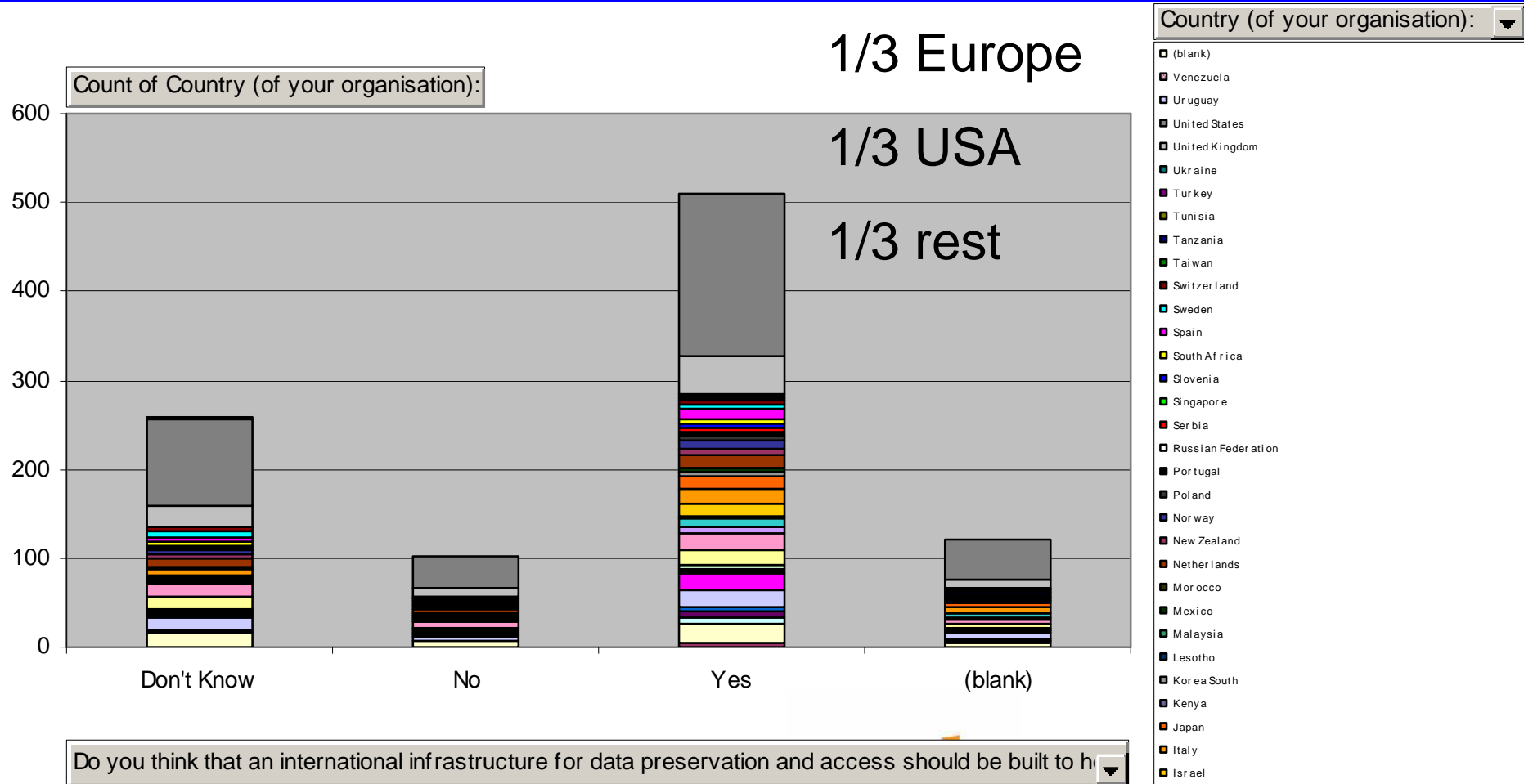
23. How important do you regard the following threats over the next 10 years?						
	Very Important	Important	Slightly Important	Not Important	Don't Know	Response Count
Users may be unable to understand or use the data e.g. the semantics, format or algorithms involved	32.9% (187)	<b>40.3% (229)</b>	16.5% (94)	7.4% (42)	2.8% (16)	568
Lack of sustainable hardware, software or support of computer environment may make the information inaccessible	<b>43.0% (244)</b>	37.5% (213)	11.6% (66)	6.2% (35)	1.8% (10)	568
Evidence may be lost because the origin and authenticity of the data may be uncertain	33.4% (190)	<b>43.1% (245)</b>	17.6% (100)	3.5% (20)	2.5% (14)	569
Access and use restrictions (e.g. Digital Rights Management) may not be respected in the future	19.2% (108)	<b>36.1% (203)</b>	25.0% (141)	12.8% (72)	6.9% (39)	563
Loss of ability to identify the location of data	24.5% (138)	<b>42.8% (241)</b>	24.2% (136)	4.1% (23)	4.4% (25)	563
The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future	37.3% (212)	<b>39.4% (224)</b>	16.7% (95)	3.7% (21)	2.8% (16)	568
The ones we trust to look after the digital holdings may let us down	20.8% (118)	<b>36.1% (205)</b>	25.4% (144)	11.6% (66)	6.2% (35)	568
	<i>answered question</i>					<b>569</b>
	<i>skipped question</i>					<b>79</b>

24. Do you think that an international infrastructure for data preservation and access should be built to help guard against some of these threats?

		Response Percent	Response Count
Yes		59.2%	337
No		13.0%	74
Don't Know		27.8%	158
		<i>answered question</i>	<b>569</b>
		<i>skipped question</i>	<b>79</b>

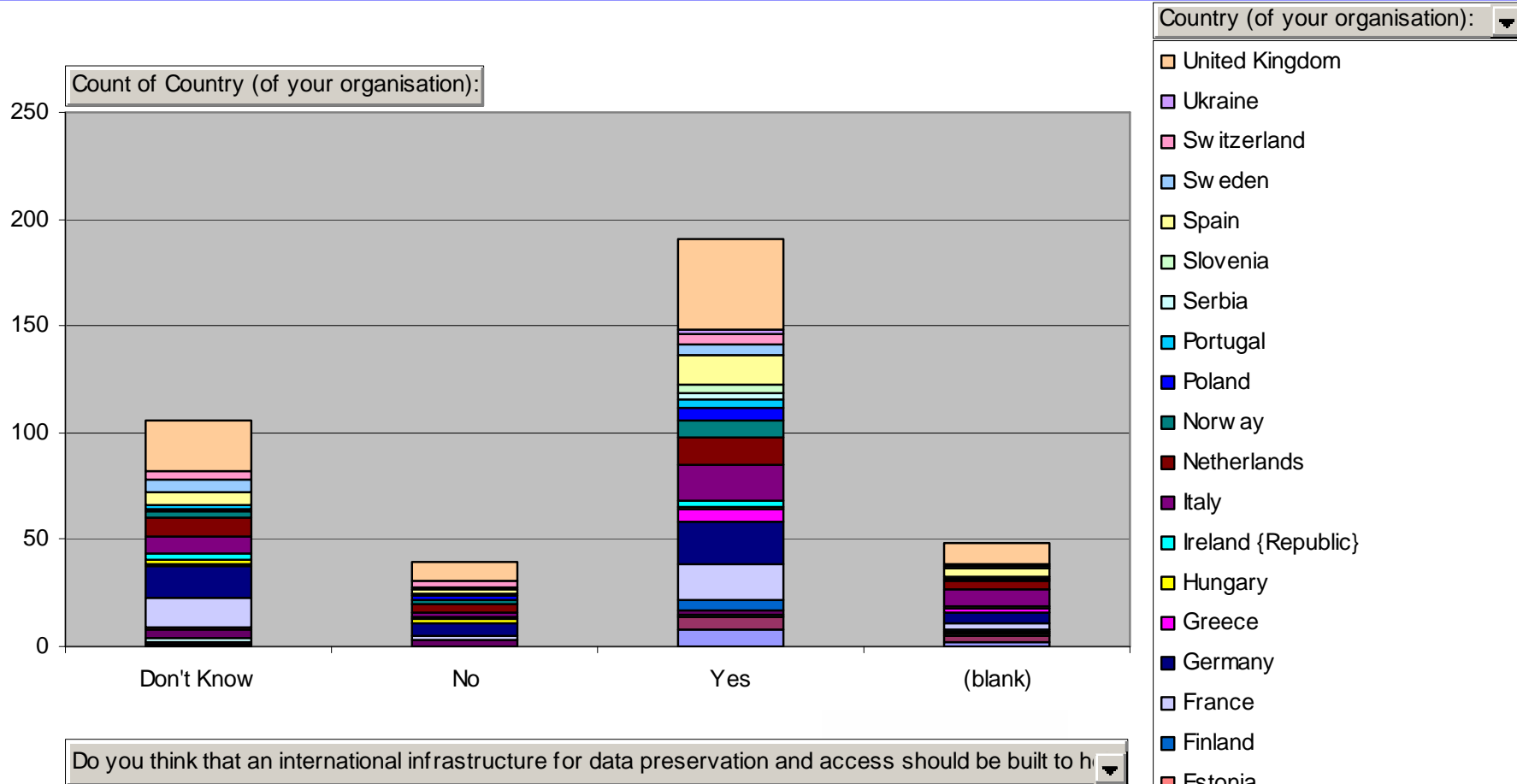
# All responses

Drop Page Fields Here



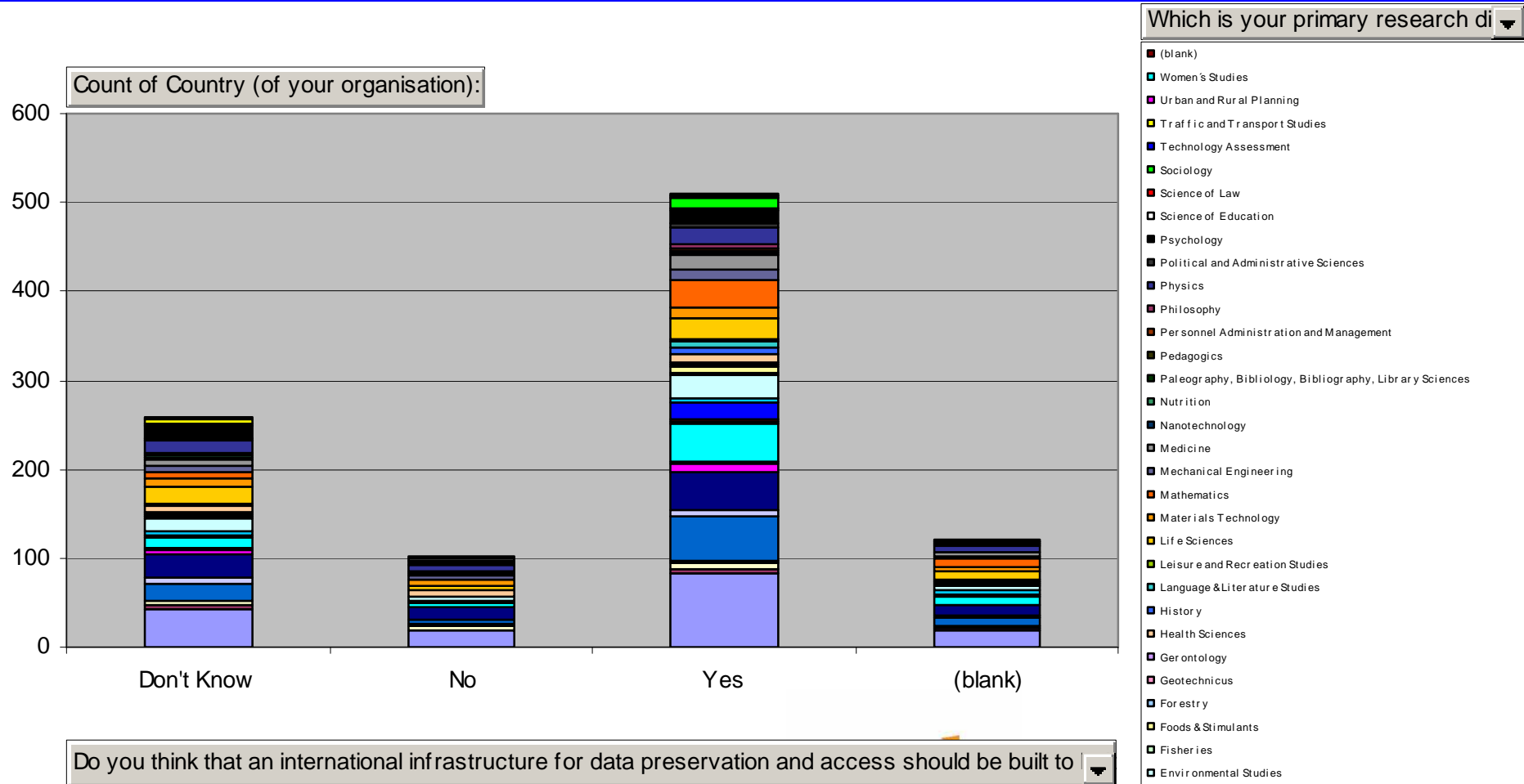
# European responses

Drop Page Fields Here



# Disciplines

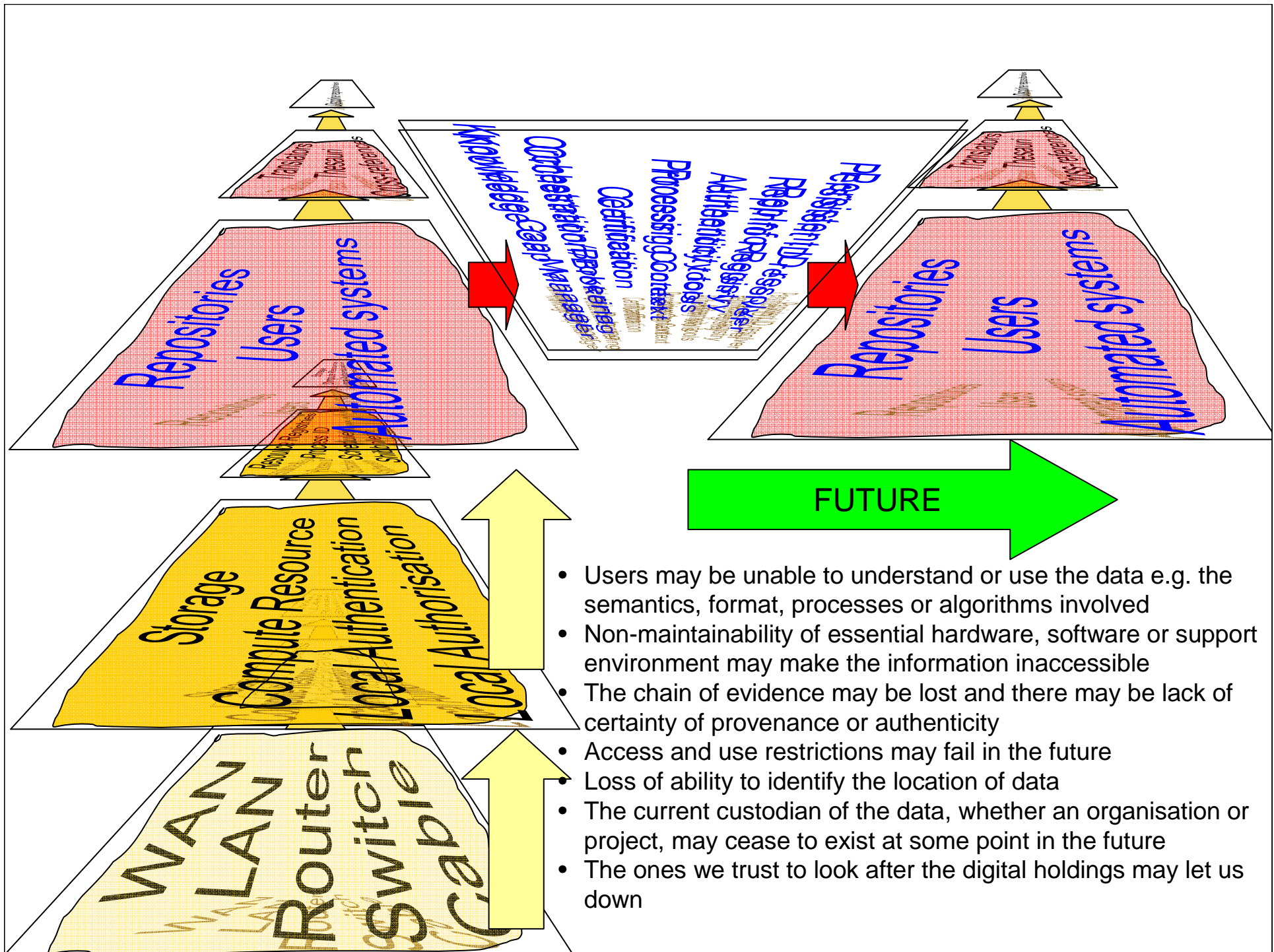
Drop Page Fields Here





# Possible European Preservation Infrastructure





- Users may be unable to understand or use the data e.g. the semantics, format, processes or algorithms involved
- Non-maintainability of essential hardware, software or support environment may make the information inaccessible
- The chain of evidence may be lost and there may be lack of certainty of provenance or authenticity
- Access and use restrictions may not be respected in the future
- Loss of ability to identify the location of data
- The current custodian of the data, whether an organisation or project, may cease to exist at some point in the future
- The ones we trust to look after the digital holdings may let us down

RepInfo Registry,  
Orchestration/Brokering,  
Knowledge Gap Manager,  
Processing Context

Authenticity tools

Digital Rights Mgt tools

Persistent ID resolver

AIP creation tool

Certification

