Digital Preservation:
The Future of our Collective Memory

Holger Brocks
Matthias Hemmje

FernUniversität in Hagen

LEHRGEBIET
MULTIMEDIA UND INTERNETANWENDUNGEN
Prof. Dr.-Ing. Matthias L. Hemmje

Agenda

- **Collective Memory**
  - EC View
  - ICT Context
- **EC Research Perspective**
- **Technology Trends**
- **Preservation Research Roadmap**
  - Analysis of previous research agendas
  - DPE recommended research
  - Online Survey
- **Conclusions**

Collective Memory: European Commission View

“Without a collective memory, we are nothing, and can achieve nothing. It defines our identity and we use it continuously for education, work and leisure”

“The Internet is the most powerful new tool we have had for storing and sharing information since the Gutenberg press, so let’s use it to make the material in Europe’s libraries and archives accessible to all”

Viviane Reding

“European cooperation is an obvious necessity in this field: it is about ensuring preservation and access to our common cultural heritage for the future generations”

Jan Figel

Archivierungsangaben

Collective Memory (II): ICT Context

http://www.ieee-tcdl.org/posstatement.html

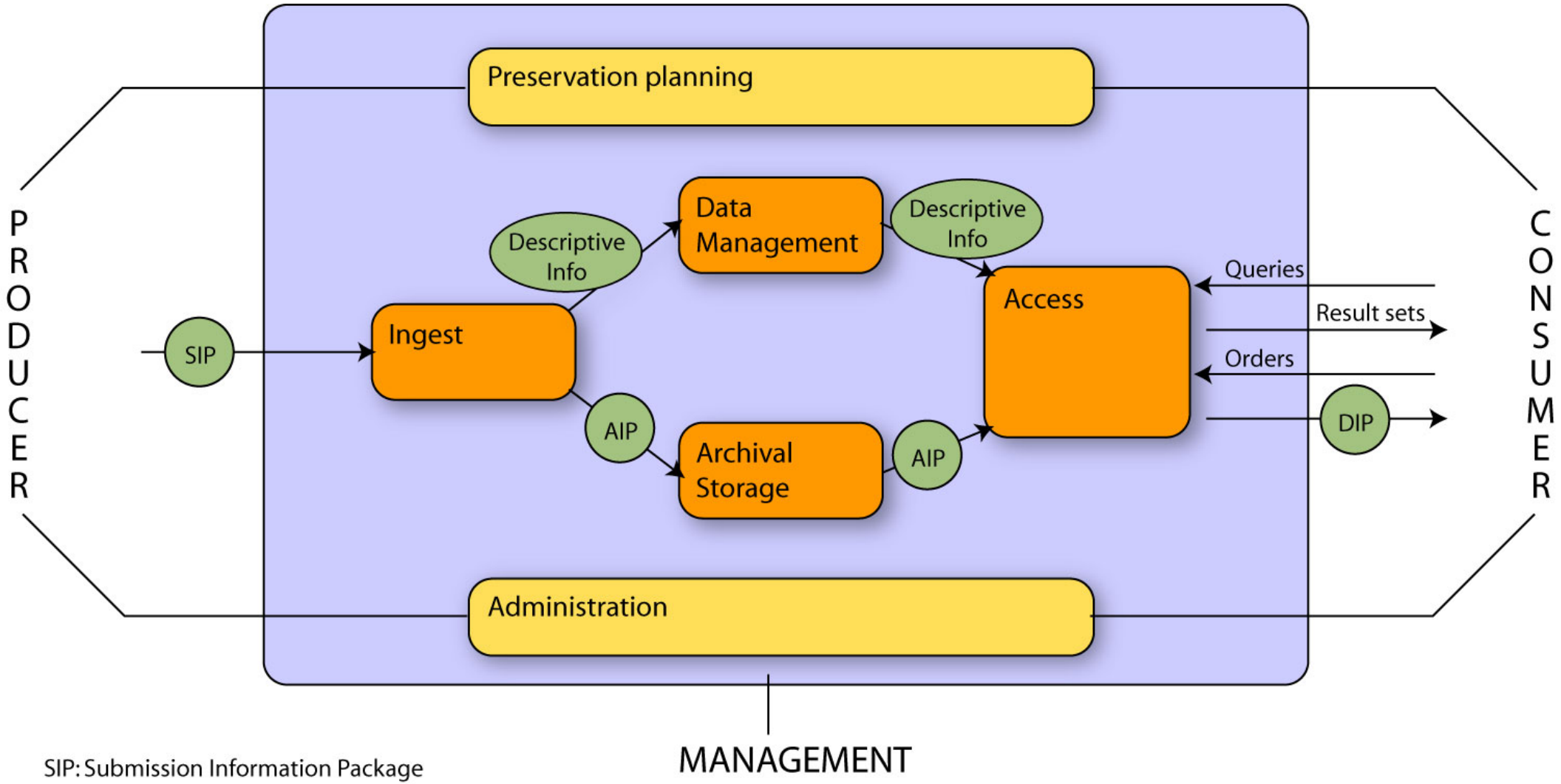
"(Digital) Collective Memory" has been used to describe the convergence of libraries, museums, archives and collections of all kinds including those of private citizens. Especially in Europe the connection of Collective Memory with its Cultural Heritage has been recognized and plays an important role in interdisciplinary research.

- **Technical Challenges**
  - Storage
  - User Interface
  - Classification and Indexing
  - Information Retrieval
  - Content Delivery
  - Presentation
  - Administration and Preservation

Policy Context i2010: Digital Libraries

- **Digital Preservation R&D Context**
  - Test bed in libraries and archives – OAIS system integrating preservation into organizational workflows and processes
  - Generic OAIS-based system for scientific data, multimedia, art and other cultural heritage material
  - Coordination of national activities mobilizing centers or networks of competence, structuring the research space for FP7

OAIS the only way?



SIP: Submission Information Package
AIP: Archival Information Package
DIP: Dissemination Information Package

OAIS as a Starting Point

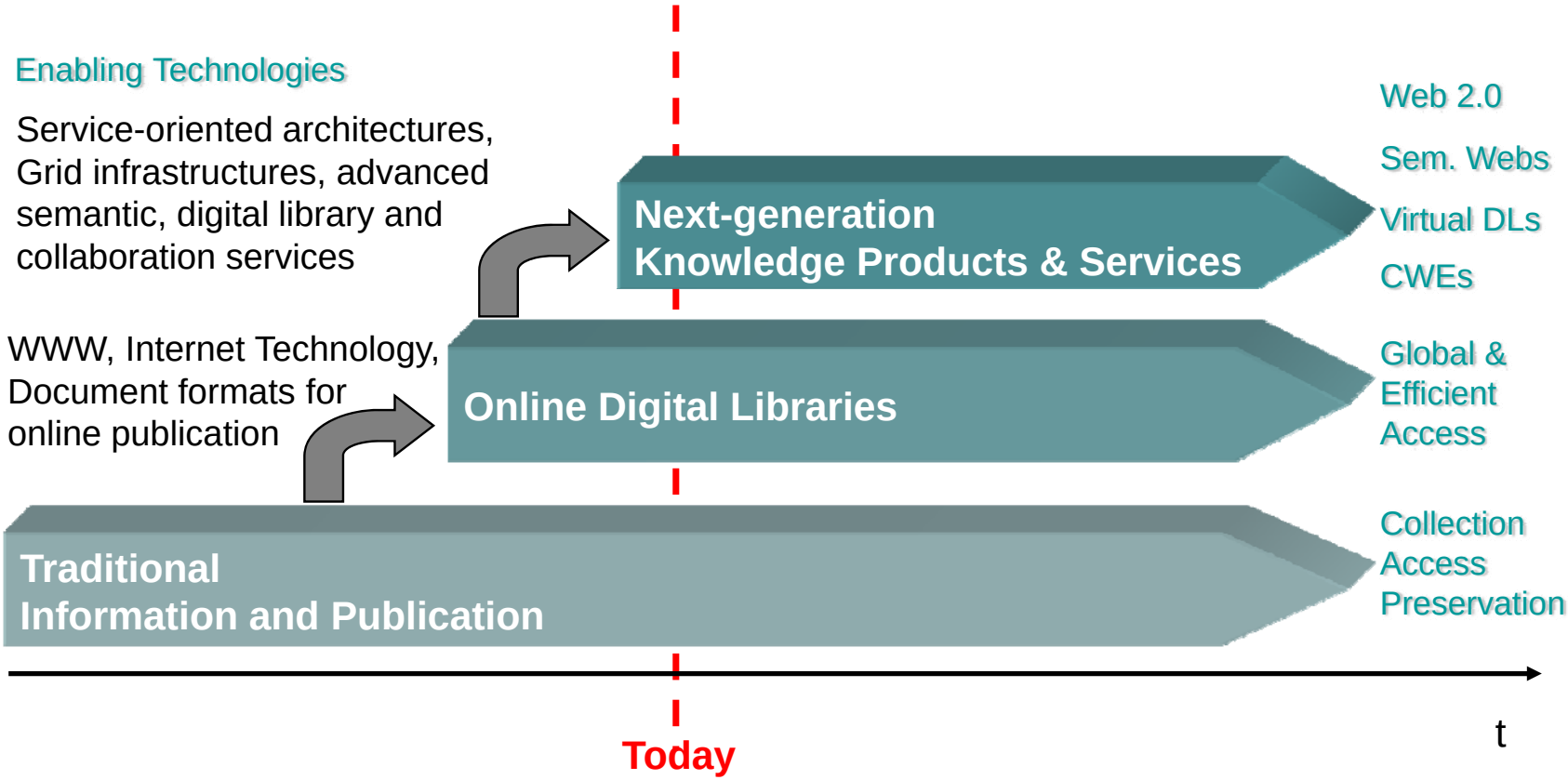
- Looking at OAIS we have…
  - a very high level conceptual model of the DP domain
  - started work on just a few early DP testbeds

- Currently support for the explicit management of
  - Static DP processes
  - Static, homogeneous information object, media and information package types

Research Objectives

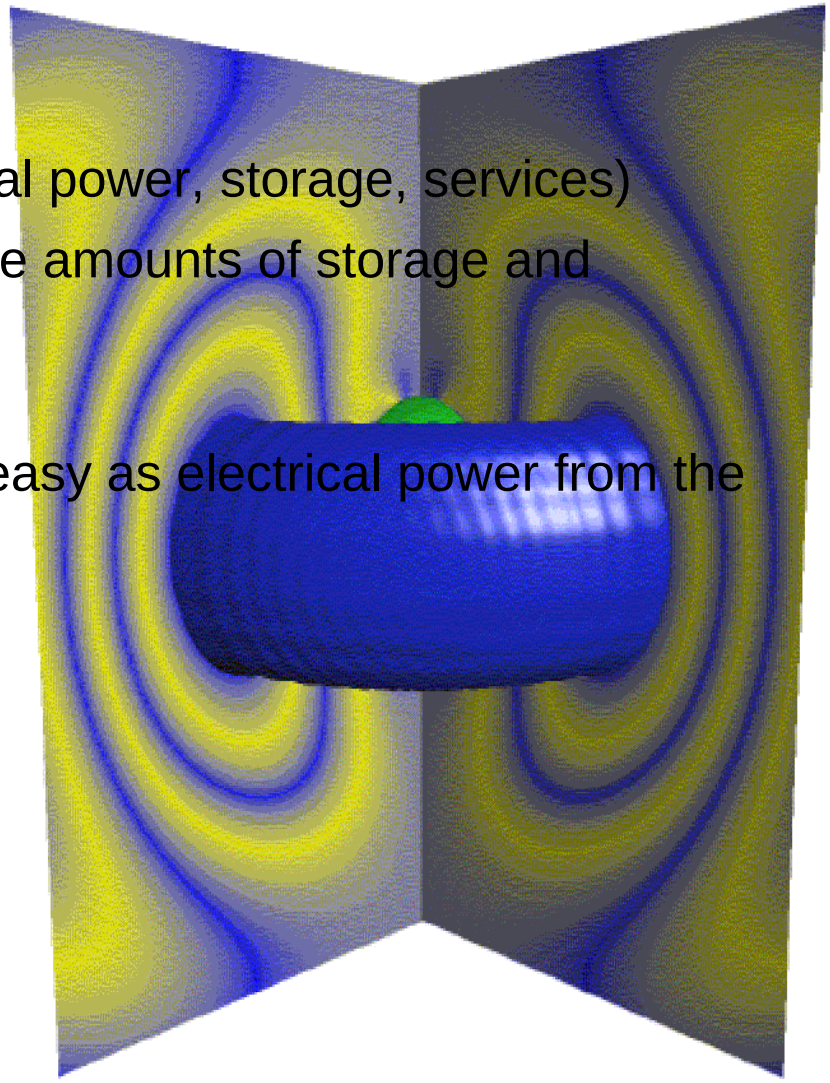
- **b) Radically new approaches to digital preservation**
  - high volume
  - dynamic and volatile digital content (notably web)
  - keep track of evolving meaning and usage context of digital content
  - safeguarding integrity, authenticity and accessibility over time
  - models enabling automatic and self-organizing approaches to preservation

Technology Trend (I): Web 2.0 and Semantics

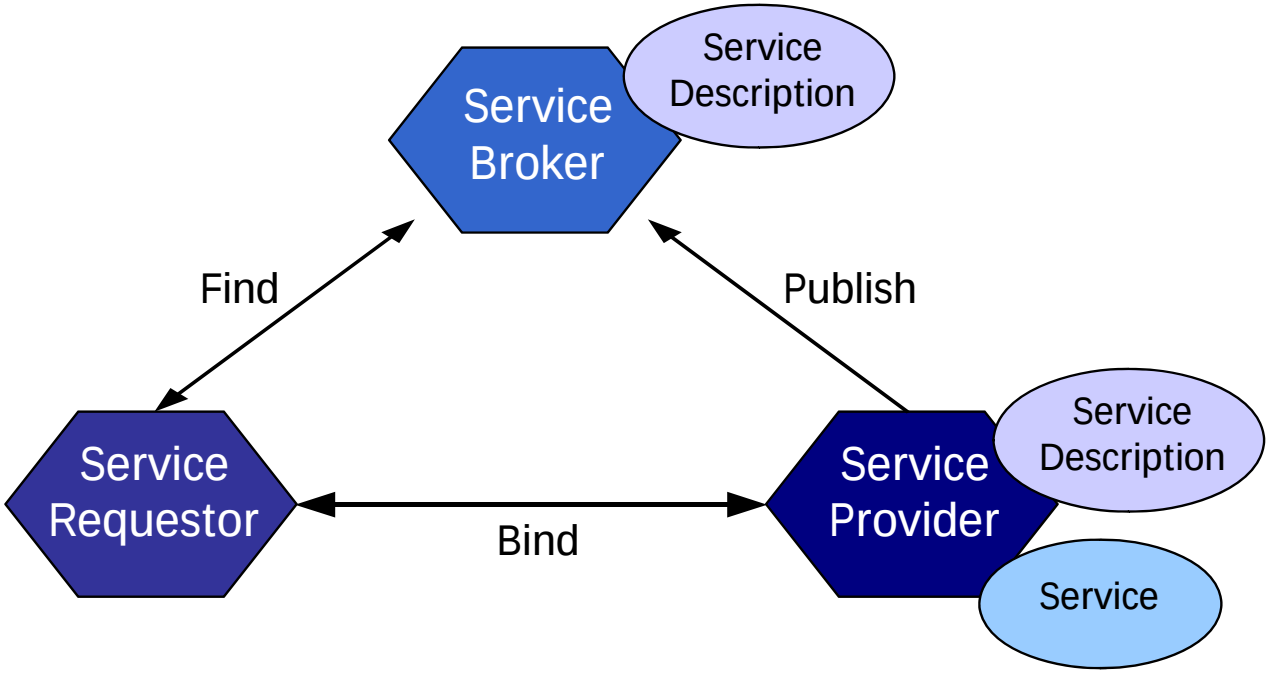


Technology Trend (II): Resource Virtualization

- **Idea**
  - Worldwide usage of resources (computational power, storage, services)
  - Born from the scientific requirements on huge amounts of storage and computational power
- **Vision**
  - Consume IT resources from the internet as easy as electrical power from the electricity grid
- **Advantages**
  - Dynamic allocation of resources
  - Cross-organizational resource sharing
  - Resource owner still have full control
  - Security infrastructure



Technology Trend (III): Service-Oriented Architectures



Technology Trends (IV)

- **Computer Science**
  - Computing architectures and resource management technologies,
  - Information extraction, e.g., linguistic and mm analysis technologies
  - Semantic representation, annotation and processing technologies
  - Effective context and workflow management technologies
  - Efficient collaboration support technologies
  - …
- **Other Disciplines**
  - Cognition
  - Cultural Heritage
  - …

Preservation Research Roadmap

- **Aims**
  - Identify the most important research issues for digital preservation
  - Specify desired capabilities and technological requirements
  - Define a time frame for addressing those issues
- **Procedure**
  - Assessment of SotA
  - Review existing DP research agendas
  - Overview of DPE core research issues
  - Expert interviews, online survey (http://www.digitalpreservationeurope.eu/delphi-survey/)
- **Acknowledgements**
  - Contributions by all DPE partners
  - Additional support by Nestor

Analysis of Previous Research Agendas

- **Research Agendas**
  - UEI, PDI, DPNU, SoDP, IAT, I2S, eScience
  - Cyber, DigiCult, Erpanet, Warwick, DRR

- **Methodology**
  - Extraction of core arguments
  - Each agenda examined by at least 2 DPE members
  - Clustering into 5 categories (Digital Object Level, Collection Level, Repository Level, Process Level, Organizational Environment)

Crosswalk Analysis Matrix

	UEI	PDI	DPNU	SoDP	IAT	I2S	eScience	Cyber	DigiCult	Erpanet	Warwick	DRR
	1991	1996	1998	2002	2003	2003	2003	2003	2004	2001-04	2005	2006
Digital Object Level												
Migration		+		+	++	+						
Emulation				++	+						+	
Experimentation		+				+						
Registries and repositories					+++	++++				+	++	
Complex Objects	+			+	++	+		+	++	++	++	
Significant properties			+	+	++	+	+					
Authenticity				++	+++++					+	+	
Acceptable loss					+			+				
Collection Level												
Interoperability				+	+++	+					+	
Metadata		+			++++							
Management					+++	+						
Standardisation			+		++		+	+		+	++	+
Media Types					+		+		+			

Archivierungsangaben

Crosswalk Analysis Matrix (cont'd)

	UEI	PDI	DPNU	SoDP	IAT	I2S	eScience	Cyber	DigiCult	Erpanet	Warwick	DRR
	1991	1996	1998	2002	2003	2003	2003	2003	2004	2001-04	2005	2006
Repository Level												
Tools and architectures		+			+		++					
Benchmarks		+			++++		+					
Hardware Issues								++++			++	
Storage		+		+	+	+++		+		+	++	
Trust		++	++	+++	+	+++			++	++++		
Scalability					+++			++				
Sustainability				+	+++					+	++++	+
Planning		+		+			++			+	+	+
Repository Management							+			++	+	
Cost					+		+			+	+	
Process Level												
Access				+		+		+	++		++	
Automation				++	+++++	++++	+		+++++		+++++	
Monitoring				+								
Organisational Environment												
Creation and use	++	+					+					
Legal Issues		+	+	+++			+	+		+		
Collaboration		+++		+++	+			++				

Archivierungsangaben

Lack of Progress…

- **Potential Explanations**
  - Lack of common understanding
  - Loss of focus
  - Lack of practical experience
  - Fragmentation
  - Frictional losses
  - IPR impediments
  - Lack of training

- **Regain Focus on fundamental DP issues!!**

DPE Recommended Research

- Restoration
- Conservation
- Management
- Risk
- Significant Properties
- Interoperability
- Automation
- Context
- Storage
- Experimentation

Online Survey

- Drawing on professional expertise to
  - Assess the current state-of-the-art in DP
  - Identify challenges that need to be met in near in medium term

- Input will be reflected in roadmap update, having a direct impact on its recommendations

- Opportunity to answer five questions about five themes…

Theme 1: Digital Objects and Collections

- Concerned with the longevity of digital objects and collections
  - Reflection on preservation in practice - strategies that work and strategies that do not
  - Handling and standardization of diverse media types
  - Handling and standardization of metadata
  - Interoperability of repositories and services

Theme 2: Management of Digital Repositories

• Concerned with the development of our understanding of the organizational requirements for long-term preservation of digital material

- The role of conceptual frameworks and standards (e.g. OAIS [ISO14721] and Records Management [ISO15489] and other emerging alternatives)?
- Defining effective preservation planning
- Training
- Repository organization and federation
- Trust and trustworthiness as institutions

Theme 3: Key Characteristics and Long-term Management of Digital Objects

• Deals with fundamental issues in digital curation such as authenticity, context and acceptable loss, but also with the automatic extraction of information from the digital objects themselves which enable their management and use

- methods for establishing integrity, authenticity and trust in digital object, such as digital signatures, persistent identifiers, audit trails for digital objects, or other alternatives
- management of intellectual property rights and privacy
- methods for identifying and maintaining the essential characteristics of digital objects that need to be preserved through time
- specification of metrics and criteria for defining audit and certification services
- judging the continuing value and significance of digital objects, balancing costs and benefits of dealing with technological obsolescence automated metadata generation by means of linguistic and image analysis

Theme 4: Emerging Research Domains

• Aims at eliciting new innovative methodologies for digital preservation and curation, drawing on experiences from related domains such as computer science and engineering, cultural heritage or cognition

- evaluating the role of developments in science fields (e.g. eScience and GRID computing) as future building blocks for digital preservation
- assessing the requirements for handling dynamic objects of increasing complexity, such as multimedia data
- identification of adequate methods for access and retrieval, linked to disciplines like information retrieval and extraction
- judging the impact of semantic technologies and knowledge representation schemes

Theme 5: Measurement and Experimentation

• Focuses on ensuring future accessibility of digital objects, especially on the role of experiments for objective evaluation of preservation strategies

- characterization of digital objects and association to appropriate preservation plans
- deployment of testbeds to enable the development of evidence-based methods for preservation
- evaluation the impact of information visualization as support medium for collection and access management
- the role of experimentation in understanding how users will interact with digital repositories

Conclusions

- **Contextualization of Digital Preservation**

- **DPE Preservation Research Roadmap**
  - Regained Focus on Core DP Issues
  - Community Validation
  - Continuously Evolving

- **Online Survey**
  - Your input is explicitly encouraged!!
  - http://www.digitalpreservationeurope.eu/delphi-survey/

- Final DPE Roadmap will be made available to public