Appendix 3

Inventory of additional ideas from the DELOS brainstorming session on an EDL

This appendix contains ideas from the 2005 DELOS Brainstorming meeting that are not included in the final report to the European Commission. It is organized using the faceted classification given in Appendix 2. The entries are sorted using columns 1 - 4 as sort keys 1 - 4. They can be sorted into a different sequence using Table Sort.

This could be the basis for an edited complete report from the meeting but is useful in itself.

A1				Horst Forster DELOS meeting on 5/6 December 2005, Nice - A roadmap towards a European Digital Library **When can European digital library become a reality?** • By 2006 there will be a full EU-wide collaboration between national libraries • By 2008 we shall have multilingual access to their collections. They must be searchable and usable • By 2010 collections need to have expanded to include archives, museums and other libraries It is a joint effort by Member States and the European Commission, requiring: • Digitisation of content • Online accessibility of this content • Digital preservation – making sure that the digital information will also be accessible for future generations **Which items will it consist of?** The European digital library contains all types of material: books, audiovisual materials, photographs, documents in archives etc. The number of collections and items that will be accessible to the public will depend on the digitisation efforts by the Member States. The Commission will help to accelerate and coordinate efforts by Member States. **What will it cost for "ordinary" people to use?** It will be free as far as public domain material is concerned. Copyrighted material may not be available free of charge. **Will it be a central or a decentrally run library?** It will be a virtual library, that is a network of libraries whose catalogues and content are accessible through a single access point (portal). Such a portal already exists. With the help of the European Commission, the portal "The European Library" was set up in 2005 by the national libraries in Europe. **What are the research issues to be addressed?** The European Commission invested in the past 4 years more than 100 Meuro in research funding in areas related to digital libraries. A substantial increase is foreseen for the years to come. We shall address: • More sophisticated treatment of digitized materials, automated indexing of texts, sound and image • Improved multilingual search engines • Services supporting annotations, collaborative working • Improve the process of digitisation • Provide systems and services that guarantee digital content will be accessible in the long term – preservation of digital content requires urgently viable solutions

A1				**Opening speech from Horst Forster** 3 challenges: economic, legal and technical In the technical challenge, there are 5 major points: 1) access and accessibility, indexing A/V; EU wants to quantify progress 2) multilingual, again to be measurable 3) services (annotation, collaboration, ecc.) 4) digitization of "everything" 5) preservation (at least 10 years from now) Competence centers on digitization More involvement of the cultural institutions in the research activity Road map 2006 full EU-wide collaboration between National Libraries 2008 must have multilingual access to library collections (searchable and usable) 2010 connections expanded to include archive and museums 2015 European cultural space, i.e. integrated access to European culture A technical agenda to go with the above has to come IN PART from the research community Provide software (multilingual, etc.) Strong technical office for The European Library DELOS must refocus its objectives in order to meet the target and make sure that the research is driven by the targets and timing suggested by the Commission Horst Forster (EC): A DL - copyright issues - fees for copyrighted material - federated system w/ central access point - EDL vs. Google: example of PPP - economic and legal challenges - competence centers

A1				Using digitized objects (incl. access): other players, scenarios, the gaps. As always in an informational and educational context the usage aspects and the users need to play a central role in our development of collections and services. We need to analyze the activities and strategies of other players, study service and usage scenarios and the gaps in what we see develop. We have to look at the integration between digitized collections and other components of DL's (or rather of eResearch, DL's, eLearning systems) and between these documents and non-traditional publication materials. This is the primary area where our research and development efforts are needed (reg. discovery, search/browse, negotiation of meaning, use in diff. contexts etc.) regardless of what the EU initiative covers. These efforts can to my mind not be delayed until after a huge Digitization effort and they can not be entirely expected from Google, Yahoo, Microsoft and similar
A1				Concept of Digital Library: definitions, different views The EU initiative is labeled "Digital Library" Programme. As so often, we need to clarify if a pile of more or less well managed and preserved digital files makes up a "Digital Library". My perception of a digital library focuses more on services than on the creation of the collection components. It involves the sustainable institutional commitment to these collections and services as well. Activities regarding documents and collection creation are a dependent prerequisite. Otherwise we would be talking about an isolated preservation effort.
A1				In Digital libraries we can look at i2010 "A European Information Society for growth and employment" (June 1 2005) (not all of it is relevant for DL research), and at eContentPlus. FP7 focus on "research on Digitization, digital preservation and access to cultural content" and much more
A1				What we need is to have the three fundamental sectors around a table to agree in a set of short/medium term PRAGMATIC targets and help the EC, our governments and our national institutions to define the correct strategies to reach them EFFECTIVELY, so we can be back and play the worldwide game (competing and cooperating, whatever it fits better in each moment and scenario)!!! Illustrated by two stories: Story 1. EU-funded TEL project lasted for three years and still no actual practical results. After two phone calls, Google included the catalog of the Portuguese National Library into Google Scholar. Story 2. The National Library of Portugal has a large amount of digital content but lack a networked storage solution.
A1				I think we should focus on streamlining and bettering what we have rather than inventing more wheels to pull an ever increasing cart
A1				Advise to Commission on system building based on DELOS research results from the research Nice results from research making possible things not possible before DELOS is missing central repository of all its papers.

A1TEL				Example: The European Library (TEL) general_views_Jill_TheEuropeanLibrary Vision.ppt
A1TEL				**TEL Vision** "Provision of equal access to promote world-wide understanding of the richness and diversity of European learning and culture." This vision is very wrapped up in the concept of access, with a strong emphasis on the richness and diversity of European Learning and Culture. DS: This could be the vision for the larger European Digital Library (EDL) **Mission** "The European Library exists to open up the universe of knowledge, information and culture of all **Europe's national libraries**". but this could easily be widened to provide access to other forms of cultural heritage, either by literally encompassing other areas or encouraging them to use the technology and standards so that any parallel system is truly interoperable DS: Actual focus more limited, component of EDL
A1TEL				TEL needs better software and functions, major upgrade Search 12 collections at one time (of 130), does not integrate results, both because of time Does not search on full text even if available in collection being searched Jill Cousins (TEL) TEL as a model for management and technology 149 collections from 12 national libraries uses inexpensive client side architecture TEL accesses 80% of available digital content - mainly metadata, only 10% digitized issues: multilinguality, thesauri, schemas vision: access - problem must be approached on a larger-scale than what has been done up to now: TEL provides an excellent test bed for this (also because it is development and implementation in context) No synonym or translation expansion TEL survey: speed and cross-browser compatibility and full-text content DELOS TEL meeting to discuss best software for TEL to make sure it is extendable and scalable.
A1				work on "perceived" differences between Google and DL. we must consider that the users of DL should be professors, researchers and students. The winning point of Google is "richness of content" as opposed to technology

A1A2				**One proposal: Two phases** **Phase 1** **Lay the groundwork based on full exploitation of present proven technology with focused research to fill gaps** **Phase 2** **More emphasis on research to add higher-level functionality to the system** The challenges that come out from the EC vision are mainly: A- 3/6 engineering (TEL) B- 2/6 management (business models, legal and political issues ...) (MINERVA...) C- 1/6 research (new issues...) (DELOS.) Same for Google, with research including maybe new features in OCR, search and access control, multilingual issues... but nothing really astonishing...) Same for present status of work in Europe We need to assure that all the European libraries, archives, museums, etc., willing to promote projects to share and put their potentially rich and very interesting contents online in the next 2 to 4 years, can have access to the required knowledge, advise and technical and economic support to do it openly and easily. Not fancy! Just easily, like it is and should be!!! If we succeed with that, in 3 to 5 years from now we'll have one HUGE achievement: a process in place toward a uniform European common space and activities of digitized cultural and scientific resources managed (described, preserved, and available for discovery and access) in uniform ways, trough uniform functional frameworks. What do we need for this? A- 3/6 engineering - but now TEL+DELOS=top engineering B- 2/6 management - but now MINERVA+TEL=effective political pushing (at the EC level, but also at the national levels, to assure that the governments provide the needed local resources, fundamental to assure the maintenance of the services) C- 1/6 research - DELOS=filling in the gaps (and keeping the practice in the front end of key research areas) Than, only after this we'll be ready to come back again to the "old speech" on smart ontologies, complex formal verifications, fancy knowledge extraction, peer-to-peer, grids, etc., etc. What the EC needs now is advising on how to focus and prioritize, and not how to disperse the efforts again! We should not repeat the mistakes of the past! We did a lot of wonderful fancy research and other things in the past, but something didn't work! My government stills not knowing about the strategic MINERVA's recommendations, so we are still not having a national Digitization programme, TEL stills lacking the engineering support that DELOS could easily provide, so TEL will have to work hard to
A1A2				general_views_Yannis_niceVision051205.ppt Individual points extracted, but hard ot separate out. Many of the points included in main report, not repeated here, see slides

A2				Yannis Delos Vision no 2 much better Yannis Expanded definition of DL much better, must be involved in daily problem solving	
A2				**Present digital libraries**	**Future digital libraries**
				Content-centric	Person-centric
				Targeted for static storage	Targeted for active communication/collaboration
				Isolated systems	Global distributed interacting systems
				Environment-specific/Isolated and repeated efforts	Generic DLMS technology to build on
				Myths about restricted applications	"All" applications
				users and providers are mutually exclusive sets	Future: `user' and `provider' are roles • Same actor may play both roles at different times • DLs at the center of scientific activity (collaboration & communication tool) • Increases set of actors and set of applications

A2				**i2010 in parallel with DELOS vision**
				• **Integration of content (e-Content+)**
				  • Creation of vertical digital repositories
				• **Management of content (e-Infrastructures)**
				  • Development and deployment of a pan-European infrastructure
				• **Addition of value to content through services (e-Learning)**
				  • Devising sophisticated functionality
				  • Serving exciting scenarios
				  • Fertilizing "all" applications with this technology	

A2				**Traditional digital library definition**
				• What: specialized content and services
				  • comprehensive
				  • rich forms and kinds
				  • read-and-expand-mostly
				• Why: learning and research
				• When: value at depth of time
				• How: measurable quality

A2				**Grand 10-Year Vision #1** Digital libraries should enable any citizen to access all human knowledge anytime and anywhere, in a friendly, multi-modal, efficient, and effective way, by overcoming barriers of distance, language, and culture and by using multiple Internet-connected devices
A2				**DLs as parts of larger organizations and application contexts:** • Health, e-health • Inclusion • Science • Government • Culture • Learning ..., e.g. virtual university • Libraries, Museums, Archives, Hospitals,
A2				From Digital Libraries to Knowledge Commons
A2	z	C4.3		If we rethink old positions, we will find surprising new answers to “..an information model for digital libraries that intentionally moves 'beyond search and access’, without ignoring these functions, and facilitates the creation of collaborative and contextual knowledge environments.” (C.Lagoze, D-Lib Magazine 2005)
A3				Federation of repositories to reduce risk and management costs (do now)

A3.4.1	B5.1			**Detecting Trustworthiness & Quality** • Preservation processes often require transforming the original bit stream. • Transformations (e.g. migration) can diminish the authenticity, quality and trustworthiness of the preserved entity. • How is the quality of reconstruction of digital entities through migrations characterised as a function of the authenticity of the digital entity? • What level of information loss is acceptable?
A3.4.1	z	C3		**Authenticity, integrity, and reliability** • each rendition carries the same force as the initial instantiation (sometime refereed to as the original) • completeness • validation of integrity and authenticity • Digital objects are what they purport to be • that we know about the history of digital objects • that we can verify that they have not changed or been modified
A3.4.1	z	C3		**Authenticity** • Requires control of ingest and its verification • Depends on immutability of the data store • Migration may destroy original byte stream • archives and stakeholders must identify significant properties and validate their migration • Support Audit of the chain of custody, process history, and the descriptions of the migration processes • Provide mechanisms to enable use
A3.4.4	z	C3		Trust • How is it established? • How is it maintained? • How is it secured? • What happens when it is lost? • How can it be verified?

A3.5				**Division of responsibility** • Digitization and physical preservation of content ➜ member states • Legality/Accessibility of content ➜ EU • Integration of content ➜ EU • Management of content ➜ EU • Addition of value to content through services ➜ EU
A3.6.1				Example: MICHAEL approach very similar to TEL
A3.6.1				Focus DELOS research on improving systems for users, esp. on programs envisioned by the Commission
A3.6.1.1				Cooperation between library and research community
A3.6.3				(greater) Entrainment research & development cycles Prototype (various) research-led elements (plug-ins) in operational DL digitization projects Practical, incremental steps yielding results for User studies and Evaluation prototypes help shape context for the application Lessons from: Rapid Application Development and Participatory Design Evolving networks: prototypes, user expectations, requirements and working practices dstudhope@glam.ac.uk http://www.comp.glam.ac.uk/pages/staff/dstudhope Lead / Ambassador Users training, tailoring and advocacy / motivation [Tudh00]
A3.6.3	B3	C4.6.2		Pilot projects to demonstrate the use of ontology and terminology services
A3.7				Developments will be demand-driven, but it is important to take a longer term and visionary view of what the user will get from the library in the way of services

A3.7	z	z	D1.2	**Users of CLIR systems**
				• A few users are truly multilingual
				  • Can formulate searches and judge relevance in many languages
				  • Want convenience of a single query
				• Many users know more than one language
				  • Want to query in their native language
				  • Can judge relevance even if results not translated
				  • Have access to document translation
				  • Objects retrieved are language-independent (e.g. images)
				Some users are only monolingual but need access to information in other languages

A3.7.1				Be prepared for new users
A3.7.1				Focus on serving given user communities, do not loose user communities - some things cannot wait

A3.7.1	z	C		**User groups**
				• Different user groups have different skill levels and different needs / requirements. Example groups of users:
				• Define user groups for cultural heritage resources
				  cultural heritage users
				  cultural tourism users
				  members of the public
				  students (undergraduate level)
				  researchers, scholars
				  genealogists
				• Users with special needs (accessibility)
				• Consider what each of the user groups will want in the future, how their needs will change (post 2010?)
				• User - who cares? Reticence in this area, "poverty of imagination" need to take the technology to the user - users don't know what possibilities are out there!
				• Machines as users - users will be decreasingly be people? (machine to machine interaction) is this a user group that needs to be considered?

A3.7.2				find out how the users are perceiving DL

A3.7.2	z	C		**User studies methodology** • The user needs to be involved from the beginning throughout the digital library life cycle • Involve users in the development of prototypes: participatory design approach • Draw on previous research in the areas of usability and accessibility, e.g. work of Marcia Bates e.g. www.gseis.ucla.edu/faculty/bates/ • Methods for defining / understanding our users, use cases, scenarios, use of personas (www.cooper.com/content/insights/newsletters/2003_08/Origin_of_Personas .asp). Also need to be aware of designing for ill-defined user requirements, when users don't know what they are looking for. Perhaps users should be defined in terms of their roles instead of groups of users, i.e. one person may be a novice on one subject and an expert in another? Scenarios and personas ; followed by prototyping ; incl. future user • Start building to help users imagine the possibilities • Need to analyze / evaluate the user experience - not just are they able to fulfill their goals - important to think about what is going on inside the users' head - the psychology of the user experience, how users deal with information, search for information. • Evaluation methods, e.g. Scenario-based evaluation, Field Studies (e.g. asking users to record the interaction with the (digital) library in a diary over 28 days for analysis) • Need for a place to share different user-oriented methodologies, Specific EC-funded support actions on usability and accessibility. To the extent appropriate, use standard methodologies for comparability of results • % of the budget should go to user-oriented work. • Making the primary user data available so that other researchers can analyze the raw data • Usability review of each of the DELOS user-oriented prototypes and The European Library - the usability testing existing prototypes can also be generally applicable.
A3.7.2	z	C4		Ethnographic user studies focusing on user tasks. Modeling characteristic research work flows and original research questions occurring at distinct stages in the research process, in order to put information models and query paradigms on a sound empirical base. Note that user questions posed to databases or produced in interviews are not original, and that there is virtually no such work so far. Advanced access systems seem to be more or less based on intuitive assumptions about user needs..
A3.7.2.1				anthropological research into the role of information in the user community

A3.7.3				Further user surveys to make improvements according to researched user needs. (TEL) actual user surveys, such as the one we carried out last May June to tell us what users want and how they want it. This summer survey showed that user expectations were high, they expect access to all the content of all the libraries of Europe and they think that everything held in these libraries is already digitized.
A3.7.3				Sharing raw data from usability tests, user surveys etc.
A3.7.3				Future research behaviors (people haven't seen film)
A3.7.3	B4			Analyze actual use of metadata by diff. types of users and use for improvements
A3.8.1				**Expectations from DLs** • Protection and conservation of cultural memory • Global access to knowledge • Preservation accountability • Reduction of costs (e.g. info reuse) • Scalable approaches • Foundation for the knowledge and creative economy • Applicability across diverse media and dynamic objects • Responsive to multilingual and multicultural collections
A3.8.1				DL evaluation measures • availability • reliability • scalability • performance • security • extensibility
A3.8.1		C3		Evaluation and success criteria for preservation programs (do now)

A3.8.3	z	C3		**Testbeds for preservation methods** • integrate, automate, and evaluate a framework for digital entity preservation by integrating and combining the test bed framework and evaluation metrics • tools to automate selected steps of the preservation process, such as ingest validation, preservation experiment set-up and control, preservation criteria definition, and verification of formal transformation, to support semi-automatic alternative evaluation. • to investigate the potential metrics for measuring the effectiveness of different preservation strategies in the context of complex digital objects • integration of software tools to support the digital preservation test bed framework. • Facility to run tests (e.g.) 'is this the appropriate preservation pathway for this digital object or class of objects, or system'
A3.8.3	z	C4.5		Test beds for mobile applications
A3.8.3.2				Develop cultural heritage testbeds, in addition to universities.
A3.8.4				Able to offer prototypes
A3.8.4				work like google: test a simple beta version of a service, and develop it if successful. EU is working the other way round
A3.9				**Completeness and Anomaly Detection in acquired collections** • Most processes for validating the completeness and closure of collections are manual with limited ability to detect missing items, errors or other anomalies. • Is it possible to detect when collections are incomplete? • Can automated processes be developed as part of the accessioning process that would provide better detection of problems with collections? • How will users be made aware of collection incompleteness.
A3.9				Quality control in preservation (do now)
A3.9	B4	C3		Quality: in the context of preservation, inconsistent, incomplete and misleading metadata must be avoided, because it will persist for a long time
A3.9	B4			Quality control of metadata

A3.9	B4.4			**Metadata quality enhancement and enrichment** • Esp. of automatically generated and harvested metadata and for aggregator services • Systematically evaluate experiences from large digitisation efforts: missing and erroneous information • Aggregation, fusion vs. cross search • Managing identity and difference (duplicates, merged, derivative) • Add authorities (names, subjects), add vocabularies for keywords (for consistency) • Relate names and identities across multiple data streams • Tools and methods needed for metadata enhancement
A3.9	B4.4			Automatic quality control of metadata
A3.9	z	z	D	Quality control of DL content
A3.9.2	z	C2.1		Data cleaning as an integral part of the digitization process
A3.9.2	z	C4.3.4		Data cleaning not as an inter-mediate processing step but as a question of continuous improvement of collaborative information integration.
A4.1	B4.2.1	C2		Use of metadata for collection analysis and to inform strategic decisions reg. multi-institution mass digitisation programmes: coverage, language, copyright, bibliographic units, convergence (adding new, other collections)
A4.2 A4.3.2		C4.3.4		Collaborative information integration
A4.2	B1.7B3			Solving the scalability bottle-neck of authority files (gazetteers, person lists, object lists), thesauri and ontologies by models of collaboration of many individuals, which converges to better stages of overall consistency of knowledge.
A4.3	B2			Interoperability cannot be based only common formats but on the capability to make the transition from one system to another (mediation, transformation, integration).
A4.3	B3			Combine and link Different types of terminology system Informal … formal folksonomy – classification – thesaurus – ontology Semi-automated methods for matching KOS Semiautomatic linking of large clusters of KOS – global duplicate detection between KOS (see LEAF Project), notably gazetteers and person authorities.
A4.3 A4.4	B3			Standards and interoperability of KOS formats (core ontology)
A4.3	B3			Interoperability of subject classifications and other KOS within member states and across member states
A4.3	B3			ontology and thesaurus merging and mapping

A4.3	B3			tools with intuitive use for domain experts to define mappings
A4.3	B3	B4.1.1		Core ontology as global model to integrate or mediate between different metadata schemes, in particular between general library schemes such as MARC, Dublin Core, FRBR based systems, and museums and archives (CIDOC CRM, TEI) and other scientific metadata schemes.
A4.3.1	B4.1.1			Harmonization of data models for interoperability between diff. metadata-schemas, as started between the Dublin Core Abstract Model and IEEE LOM/IMS
A4.3.1				Interoperability of heterogeneous resources
A4.3.1	B1.2.1			Applications typically have their own standards. Interoperability between the digital library and the applications as well as across applications is important
A4.3.1.1	B4.1.1			Multiple cludgy crosswalks tend to break the semantics
A4.3.2	z	C		**Integration of Digital Libraries** Putting together Libraries that consider the same content, but may belong to different ontologies (set of metadata). Answering questions such as: Other news about the same object Similar (in what sense?) objects Related (in what sense?) objects Prearranged pathways …. The example of DICE
A4.4	z	z	D1.2	• Unicode (http://www.unicode.org/) • Multilingual Dublin Core http://dublincore.org/groups/languages/ • RDF Encoding of Multilingual Thesauri http://www.w3.org/2001/sw/Europe/reports/thes/8.3 • OWL (Web Ontology Language) http://www.w3.org/TR/2004/REC-owl-features-20040210/

A5				**DELOS Agenda**
				• Foundational research
				  • Reference Model for Digital Library Systems
				• Systems-related research
				  • Digital Library System Architectures
				  • Information Access to Digital Libraries
				  • Audio/Visual Digital Libraries
				  • Semantic Interoperability in Digital Libraries
				• User-related research
				  • User Interfaces for Digital Libraries
				  • Digital Library Visualization
				  • Personalization in Digital Libraries
				• Horizontal issues
				  • Digital Library Curation and Preservation
				  • Digital Repositories
				  • Digital Library Evaluation Methodologies
				  • Digital Library Evaluation Infrastructures
				• Applications
				  • E-Health Digital Libraries
				  • . . .

A6				Awareness. Information to the citizens on the use and the quality aspects of content on the internet
A6				**Education programs accompanying the initiative**
A6				Education. Curriculum. On MSc and PhD level
z	B1	C3		Repository models
z	B1	C3		**Modelling Preservation Processes** • Improve preservation by building preservation functionality into systems used to create and manage digital objects. • This requires improving our knowledge about what preservation functionality really is and ensuring that this functionality can be effectively communicated to system developers.

z	B1.1			**Self-describing & self-monitoring entities** • Digital entities that know what they are • Digital entities that know something about their semantics • Digital entities that can observe the state of other objects (e.g. observe decline in numbers of similar classes of objects) • Digital entities that know where they are • Digital entities that know where their metadata are • Digital entities that can notify their originator/manager if they need to be protected, migrated, secured
z	B1.1			Move from DL as an integrated, centrally controlled system to dynamic, configurable federation of DL services and information collections. hot_tech_Erich_DLVisionV3-NEU-stick.ppt
z	B1.1.2	C4.2.1 C4.3		Peer-to-peer search engines with heavy involvement of users as information creators hot_tech_Gerhard_search_engines.ppt Individual points still to do
z	B1.1.2	C4.2.1 C4.3		---------- Weikum what Google can't do: - ontology - semi-structured and semantic data - context - humans in the loop: collaboration, recommendation P2P systems - scalability & self-organization - better search result quality: -- powerful search methods --leverage inte;;. Input -- collaboration among peers - small-world phenomenon
z	B1.2.1			Architectures for supporting important applications on top of digital libraries should be investigated
z	B1.2.1	C4.3		Models and architectures that integrate user societies with functions and workflows that generate knowledge (like with distributed annotations) is important

z	B1.4			• Current models for repositories provide a useful starting point
				• Further research is needed to develop technical specifications and standards to build persistent DLs.
				  • Definition of a service layer that would allow distributed repositories to share content, tools and services (e.g. security, user profiling, management, privacy)
				  • Models and specifications for discovery, access, security and retrieval across diverse repositories and collections
				  • models need to be tested for scalability
				• Change will be a feature of repositories
				  • Storage technologies
				  • Services, close down of some and initiation of others
				  • Workflows
				  • Verification mechanisms
				  • Migration, refreshing, emulation—and ….

z	B1.5			Tools for transforming data from one format to another (do now)
z	B1.6	C3		• Current digital preservation processes require extensive human intervention for selection, validation, description, assigning unique identifiers, data management, migration, and selection and appraisal. • The degree of human labour currently required does not scale to the size or complexity of the digital content that needs to be preserved.
z	B1.7			Make sure applications and prototypes are scalable
z	B1.7	C3		• preservation research to date has examined either large sets of homogeneous data or small collections of heterogeneous material. • This raises a series of issues concerning the scalability of current models and methods, the ingest rate, and the rate at which digital materials can be normalized or migrated. • Is it possible to develop metrics to assess the scalability of preservation strategies and methods?

z	B2.1	C		**Managing Complex and Dynamic Digital Entities** There has been little research to address how interrelationships between the components of compound documents/digital entities might be maintained. • How can complex and dynamic entities be authenticated and their integrity verified? • How can dynamic entities be accessioned and managed in an archive? • To what aspects of a dynamic document should metadata be attached and what metadata would be required? • How do we ensure dynamic qualities across time? • At what level is loss of dynamic qualities acceptable? What measurement metrics?
z	B2.1.2	C4		access to structured documents
z	B2B5.3			Standard formats to assists generic information extraction tools
z	B3			Standard tagging of places, persons, events, concepts early in the digitization process and support of harvesting to support semiautomatic creation of KOS from sources.
z	B3			hot_tech_Martin_KnowledgeOntologies_2.ppt Individual points still to do
z	B3	C		Role of ontology in preservation (longer term research issue)
z	B3	C3		change of ontologies/vocabularies over time
z	B3	C4.6		Ontology / terminology services to support Vocabulary creation and maintenance Mapping, merging vocabularies Document creation and maintenance Indexing, classification, annotation intellectual, automatic Discovery of services and databases/collections Searching for concepts --> controlled terminology, auto-disambiguation Querying and result display Cross-searching, cross-browsing, mapping services Browsing and user interface/visualization Query expansion Extraction/mining of terms Translation support using vocabularies Content integration and mediation
z	B3	z	D1.2	translation resources (dictionaries, corpora, MT systems)
z	B3.3			Increased use of Subject Thesauri - Ontologies (mentioned by TEL as a need)

z	B3.5B4	C3		**Long-term Metadata Viability** • meaning of metadata itself changes over time, what we might describe as 'metadata drift'. • For purposes of interpretation and authenticity, users will need access to the metadata schema used at the time the digital entity was created. • research needed into metadata schema and ontology evolution mapping to ensure that, over time, metadata and underlying ontologies do not lose their meaning. • Tools to track provenance, version control
z	B3B6			Modeling core information structures of widely used discourse classes, such as description of past events (archeology, history, paleology, geology, food safety, epidemiology etc.), scientific observation in descriptive sciences (biodiversity, geo-sciences, meteorology, medicine), theory and evidence in science, in order to support content-linking models
z	B4			hot_tech_Traugott_metadata_MDc.html Individual points from this extracted
z	B4			**Metadata general** • List of important research and development requirements focused upon digitisation of documents/objects and their subsequent use • Metadata is orthogonal to the other aspects reported today: needed to make the other efforts possible • Metadata approaches are highly heterogeneous for different purposes like digitisation, stewardship, preservation, multimedia or multilingual objects, eScience or eLearning. Metadata for interactions between entities, for usage data, user data, choices and behaviours Metadata for services, collections, institutions, people etc., for other than solid information objects Widely different types of metadata: descriptive (discovery); administrative (incl. rights, access, technical, management); and structural (incl. context, presentation) • Research and development needs are tightly linked and hard to prioritise, since we are preparing a comprehensive practical task, a mass digitisation effort and global access to the documents
z	B4			Recommendations: a) minimal level of metadata required, based on OAI b)develop clear guidance reg. level of metadata and subject indexing
z	B4			Research in metadata propagation and inheritance between related objects
z	B4			Deeper conceptual content description

z	B4			A recent evaluation of a German national digitisation programme during the last couple of years reveals clearly insufficient metadata practice, endangering the usage of the digital documents, not to speak of their preservation: 33% of the objects had no metadata at all, 33% bibliographic metadata only, 10% had both bibliographic and subject metadata (rest: no information) Less than a third of the metadata was digital
z	B4	C4		Technical issues of exposing large quantities of metadata
z	B4	C4		metadata needs for information access
z	B4	C4.3.4		• Social use, annotation, (collaborative web) tagging, rating, rankings, reviews/recommendations • End-user creation (social tagging)
z	B4	C4.6		Making services available from (metadata) records rather than vice versa
z	B4	C4.6.3		Metadata services as web services m2m, e.g.: generation, augmentation, transformation, equivalence, crosswalking schemas and vocabularies, archiving/persistence, annotation, metadata improvement and rating services
z	B4	C4.7		Leverage metadata functionality outside the systems they reside in today and make them useful inside new applications (web services approach), in multiple repositories, creation environments and discovery mechanisms. How to tie together (service orchestration); infrastructure for such services
z	B4	C4.7		Address problems when reusing metadata created for multiple diff. purposes and contexts
z	B4	C4.7		Metadata methodologies allowing packaging and repurposing, derivative and aggregated works, recombinable content
z	B4	C4.8.1		Use structure of the data in interfaces Use metadata to create adaptive user interfaces for diff. groups of users
z	B4.1.1			Generic profiles for discipline/purpose (e.g. eScience, preservation)
z	B4.1.1			Application Profiles for consistency and interoperability
z	B4.1.1			Schema creation tool
z	B4.1.1			• Common principles for XML-binding and other encodings • Role of syntax and vocabulary encoding schemes and terminologies/authorities
z	B4.1.1	z	D5.2	New schemas needed for increasingly interactive contents

z	B4.1.2	C3		Representation Information Registries
				• Representation Information registries provide keys to understanding the nature of digital objects
				  • to identify the format of unknown files,
				  • to verify whether a file is the format that it purports to be,
				  • to assess the viability and implications of transforming from one file format to another,
				  • to provide an information resource to support the investigations of file format risk, and
				  • to store information about how to render an object from a particular format.
				• guide the managing of their transition from one state to another

z	B4.2			Metadata for entire life cycle of objects
z	B4.2			Enhance object metadata with author profiles and vice versa
z	B4.2 B4.4.1			Identify genres/document types, formats
z	B4.2			Relationship between metadata for actual discovery and access and metadata for long term preservation
z	B4.2			Specific importance of contextual metadata to archivists and records managers (draft ISO standard Records management metadata): people, policies, processes and systems and the records themselves
z	B4.2			**Documentation of Functionality and Behaviour** • Need formal ways to express the functionality and behaviour of digital entities. • These are needed to establish benchmarks and measure consistency of performance across migrations or emulations. • Approaches to functionality and behaviour abstraction and representation are also needed to enable us to reconstruct applications and systems. (e.g. Culture?)
z	B4.2	C2		metadata needs for organizing the digitization effort

z	B4.2	C2.1		role of metadata in digitization: determine from two metadata records whether they refer to the same document, which is not always easy • Once a book has been digitized, metadata can be enhanced. • Often it will be advisable to digitize the front matter and then based on the information obtained do a more reliable check to see whether this edition has been digitized already • Especially for older books, the metadata should contain a note on the condition of the copy the library owns (such as found in catalogs of used books for sale). That way it can be arranged that the best available copy is digitized. This also may give an indication of the quality of the image and of the OCR that can be expected.
z	B4.2	C3		• Preservation metadata is an essential part of the information infrastructure necessary to support all the processes in digital preservation. • automatic or semi-automated creation and authoring of the technical, descriptive, and structural metadata are a crucial issue. • Need for creation of metadata supporting the discover, use and understandability of digital objects.
z	B4.2	C3		• Scope and depth of information needed to support digital preservation: processes not known yet • Better understanding needed of the role of metadata in supporting preservation and data curation
z	B4.2	C3		Different kinds of metadata will be needed to support different digital preservation strategies
z	B4.2	C3.8		• Digital entities need to be characterized independently of underlying software and hardware infrastructure to reconstruction in newer environments • Machine interpretable expression of the significant properties of digital assets • Mechanisms for identifying and representing these 'significant properties' • Registries store expressions and as a source of generic expressions (e.g. .xls, .sxw)
z	B4.2.1			Provenance; legal, administrative, procedural, documentary and technological context; use history, integrity and authenticity of materials more important
z	B4.2.1			Data provenance important for some applications on top of DL (e.g. e-science)
z	B4.2.3 B4.2.2. 3			Include content standards (vocabularies, name authorities etc.) into metadata generation applications: tools supporting indexing, classification, KOS, ontologies
z	B4.3			Bibliographic relationship control: Work-expression-manifestation

z	B4.3.2			Collection description are available in native language and English, guidelines for preparing collection description
z	B4.4			Metadata for entire life cycle of objects
z	B4.4			Support diff. metadata profiles for diff. collections
z	B4.4			Handle multiple metadata creation and repository environments; multiple metadata formats
z	B4.4			Exploit existing sources of metadata fully; exploit implicit metadata; extract from document, context, related content, author information, usage context and feedback
z	B4.4			Role of manually created metadata; integrate human and automatic processes
z	B4.4.1			Automatic capturing of metadata about the (complex) objects, the actions undertaken on them and about people, organizations or software controlling these actions
z	B4.4.1	z	D3.5	Multimedia indexing / Feature extraction on audio-visual objects (generic or application-dependent)
z	B4.4.1 B5.3	z	D3.5	Multimedia Knowledge Extraction and Representation • Currently mostly low level visual or audio features are extracted • Recent approaches started looking at higher level concept extraction (like events) supported by the MPEG-7 standard • Models integrating domain ontologies with the content description standards to represent video content should be investigated • Additional knowledge representation structures should be investigated to capture more knowledge on the video content (activities, states, facts, opinions, etc) • Models that also represent the context at the time of capture related to the video objects should facilitate retrieval and data mining in the long run • The video knowledge extraction process should investigate methods to organize systematically the software to exploit domain knowledge, multimodal extraction clues, and previous knowledge on the type of video

z	B4.4.1.4			**Automated Appraisal & Description**
				• Automating the process of selecting material
				• Annotation and Provenance
				  • What about structural differences—say between radiological and linguistic data sets and their annotations
				• Summarization technologies (point of view)
				  • What about composite documents
				  • What about databases, images,
				• Knowledge representation developments & tools
				• What if collection development testing

z	B4.4.1 A5.4.1.2	z		**Automated extraction of metadata** • Understandability of objects • Scalability and cross-media applicability • Responsiveness to semantic layering • Applicability to unstructured or semi-structured materials Example: Bridging the semantic gap – contrast in imaging searches on color, shape, or texture with those for objects (e.g slippers) or concepts (success) or in the case of audio emotion, decision points, interaction patterns.
z	B4.4.2.1	C3		• Automated preservation metadata tools needed
z	B4.4.2.3	C4..3		Metadata creation embedded in routine workflows (repository ingestion) and organizational patterns
z	B4.4.2.3	C4.3		Embedded/improved tools for metadata creation (into authoring, content managing, learning object managing tools, scientific machinery etc.) Cooperation with content generation software vendors
z	B4.4.3			Sharing metadata (but: trust and policy patterns)
z	B4.4.3			Link metadata from different sources
z	B4.4.3	C4.2.1		Expose metadata through a variety of interfaces and protocols for searching, harvesting and search engine reuse
z	B4.4.3	C4.3		Exploit work spaces of many users for metadata generation by harvesting metadata assigned by users and by mining usage history of documents / resources
z	B4.4.5			• Assure metadata update • Tool for detecting and reporting changes to resources • Digital curation of metadata itself

z	B4.4.5	C3		Preservation metadata must be preserved itself, migrated and described and upgraded to evolving new metadata standards. This might be more difficult when it is packaged together with the objects
z	B5	z	D3.5	Methodologies for semantic compression and delivery of multimedia documents are needed
z	z	C2.1	D	Issues in digitization that require decisions • Is it important to retain the appearance of a book, or is it sufficient to get the text? This may need to be decided case by case using guidelines. May need to store both an image and the OCR text (see multivalent documents, http://multivalent.sourceforge.net/) • Should different editions of the same work be digitized? This question is related to item 1. It may be sufficient to digitize front matter, some sample pages, and any full-page images of the edition of a work for which another edition has already been digitized. (If there are partial-page illustrations in the text, the whole book needs to be digitized.) • A priority scheme for digitization should be established using these and other criteria.
z	z	C2.1		libraries should collect more material which is already in digital form
z	z	C2.1	D3.6	Digitizing solid objects as 3-D images that can be further processed and analyzed (see spotlight_Griffin.ppt)
z	z	C3		Seamus_delos_dpc_cluster_nice_brainstorming_mtg.ppt
z	z	C3		preservation orthogonal to all other issues

z	z	C3		**Preservation and curation challenges**
				• Medium
				  • storage media naturally decay or become obsolete
				• Technological (e.g. hardware/software)
				  • hardware and software obsolescence makes data/information inaccessible
				• Context
				  • avoid loss of meaning with metadata
				• Process & dynamic nature
				• Legal Impediments
				• The organisation and its staff
				  • Lack of organizational will – visible benefits
				  • Decentralized and node-based organisation

z	z	C3.7		Digital curation: “maintaining and adding value to a trusted body of digital information for current and future use”

z	z	C3.7	D3	**Approaches to Curation** • Different formats require different kinds of strategic approaches to ensure that they can be accessed in the future. • Problems with formats are exacerbated by the fact that archival collections, which need to be managed as a whole, generally contain entities in multiple formats; these formats have different rates of obsolescence. • E.g. we need predictive measures to enable developers to assess the preservation impact of attributes of formats in advance of their completed development or use.
z	z	C3.8.2		**Salvage and Rescue** • little work has been done on developing techniques to enable raw data streams to be analyzed and the original meaningful (e.g. logical) units they represent reconstructed (e.g. crypto-analysis methods) • generic devices for reading media • Could be classed as digital archaeology
z	z	C3.8.3		**Repositories of obsolete software** • Emulation and salvage and rescue techniques may depend on software that may no longer be available • A small number of software repositories to collect, maintain, and provide access to obsolete software. • Examples: it might hold a characterization of the capabilities of prior systems, which can be implemented using modern technology, or it may hold routines that can migrate obsolete encoding formats to contemporary encoding formats. DS: Can be incorporated in general software repositories
z	z	C4		**P2P search engines have great potential:** • harness local resources for power search engine • rich models for content extraction, annotation, summarization, and indexing of text, images, speech, audio&video, feeds, portals • customization and personalization • collaboration & recommendation networks with other peers • naturally fits with mobile clients and context awareness • naturally gears for rich cognitive model of user behavior • no monopoly, no central profiling or bias • great benefit for European society, economy, science • business applications in intranets, communities, web archives, search embedded in business intelligence, mobile apps, etc. See hot_tech_Gerhard_search_engines.ppt

z	z	C4		Generalized “search+” services knowledge-based integrated into applications triggered by context search plus analysis and further steps
z	z	C4	D1.2	Interactive CLIR systems can help users locate and identify relevant foreign-language documents • Formulate and translate the query (e.g. entering diacritics, selecting translation alternatives) • Query re-formulation (e.g. selecting query expansion terms) • Browsing/navigating results (e.g. translating metadata) • Identifying relevant documents (e.g. summarizing and translating results)
z	z	C4	D1.2	Multilingual information access, cross-language retrieval
z	z	C4	D3.1	access to audio materials through technical advances in speech recognition.
z	z	C4	D3.2	access to audio materials through technical advances in speech recognition.
z	z	C4.2		multiple information access methods
z	z	C4.2		have an excellent search system that works from the full text of documents. There should be a separate call for the development of such systems, especially in the social sciences and the humanities
z	z	C4.2		Search aids for people who don't know what they are looking for or what is there already
z	z	C4.2 C4.8.1		Multi-modal specification of queries (navigation, interactive, symbolic vs. visual, …) Spoken queries
z	z	C4.2	D1.2	Non-language access to information (images and A/V)
z	z	C4.2		Both browsing and searching of digital library content required
z	z	C4.2	D1.2	More difficult in open document space as opposed to controlled collection Multilingual search of metadata about all kinds of documents (books, multimedia) - short term Multilingual search of spoken text - medium term
z	z	C4.2	D1.2	User interface: disambiguate terms interactively, more of a problem in multilingual search Assistance in selecting terms in another language
z	z	C4.2	D3	Cross media discovery and retrieval (e.g. across audio, images, text, and three dimensional models)

z	z	C4.2	D3.4	Music search The access and search of musical documents is an emerging research area that focuses on the content-based access and retrieval of musical documents against musical queries. Approaches and paradigms: content-based approach to indexing query-by-example, or query-by-humming, paradigm user-friendly interfaces for evaluation.
z	z	C4.2.1		Why not use the FAST search engine on top of Tel
z	z	C4.2.1		**Recommendation** **Produce a full-functioned search/access engine specifically geared towards cultural heritage materials and users** **Underlying Methodology Principles** ■ Schedule: Incremental delivery of functionality, w/ explicit time line (to be determined) ■ Requirements: Advancing hand-in-hand with the users (cooperation with projects – TEL, EPOCH) ■ Maturity: Everything DELOS has produced ■ Advertisement: Demonstrator Workshop ■ Evaluation: Creation or formation of testbeds, user-oriented evaluation throughout life-cycle **R&D Dimensions** ■ Indexing of objects and structure at digitization time ■ Multimedia data (text, image, video, film, …) ■ Query formulation ❑ User aides ❑ Multi-modal specification (navigation, interactive, symbolic vs. visual, …) ■ Query-less user notification ■ Multi-modal similarity (pre-computed and ad hoc) ■ Multi-modal result delivery ❑ Relevance-based ❑ Visualized Underlying R & D Principles ■ High scale ■ Personalized / Communalized / Contextualized

z	z	C4.2.3 C4.3.1		Annotations can contribute to a novel search model supporting contextual navigation A collection of multimedia documents and user's annotations over them constitute a hypertext If there are no constraints on the number of annotations that connect a pair of documents, different information of the same contents given by independent users can be represented. Methods for automatically processing the annotation-based hypertext can be studied to: automatically extract information about the relationships among digital contents provide advanced search functionalities by exploiting the relationships among annotations and annotated documents give the user the possibility of navigating the hypertext maintaining contextual information.
z	z	C4.2.5		Navigational Interfaces • Search is a limited, narrow, and unsatisfactory (if alone) paradigm • Links, pathways, grouping, indexes, directories, guided tours, can do a better jobs • (several) Navigational interfaces can be built on top of the same library (or a set of libraries) • The example of VICE
z	z	C4.2.6 C4.4		Presentation models that take into account the user context to structure the results and present them with appropriate personalized visualization metaphors should be investigated
z	z	C4.2.6		Principles for result presentation: • Ranked by relevance. • Related items together. Relationship established by similarity or by links. • Same use ('supermarket principle") Multimodal result presentation
z	z	C4.2.6	D1.2	Result presentation in multilingual systems Results must be presented in manner appropriate for user (e.g. languages he can understand) (personalization) With respect to translation, there are three levels: - Translation of metadata, perhaps just subject terms, perhaps the title, perhaps the abstract (possibly an automatically produced abstract), use the metadata as context for disambiguation - Automatic "gisting" into user's language from full text (if no abstract is available) - Draft translations of the full text

z	z	C4.2.6	D1.2	results presentation (extraction and merging from multilingual collections/summarization/translation ….)
z	z	C4.3		**User-oriented functions for searching and working with retrieved material** Digital library life cycle Collation: create personal or group digital library Interpretation Authoring Move from metadata level to content level! Collaboration throughout the whole DL life cycle! **www.daffodil.de**
z	z	C4.3		provide tools that allow scholars to really work in the digital medium
z	z	C4.3	D3.6	Analyzing / processing 3-D images of solid objects (see spotlight_Griffin.ppt)
z	z	C4.3.1		Services for the production of knowledge, such as creating annotations etc., the ability to be able to make annotations independent of the physical location of the user is important
z	z	C4.3.2		"We see in the future that users can and will be much more actively involved in contributing to their cultural heritage online (e.g. the Wikipedia model)." (European digital libraries: MEMO/05/347) This requires good, easy-to-use authoring and contributing tools. Need to decide on the degree of control, somewhere in between a DL crated simply by digitizing the contents of existing libraries and incorporating borne-digital document from publishers on the one hand and the open Web on the other.
z	z	C4.3.4		Public domain information and other material for general use is already made available online in a large scale and widely used in many users' daily business (e.g., the free encyclopedia wikipedia). It is important to stimulate the large community of users hat provides valuable information and documents to the community to further invest in their voluntary and unpaid work. A major concern must be to seamlessly integrate this information and content into the European Digital Library. [HOW could an award model for volunteer contributors look like, will this be feasible at all? In general, I think it is a very good idea when individual unpaid efforts are rewarded in some way just to foster voluntary contributions. But in the context of the future European DL, I do not see right now, how this could be done]
z	z	C4.3.4		provide tools for collaboration among users For example www.collate.de/Collate-System-Public-Version-Download.htm European Digital Library as an environment that facilitates collaboration across borders Remember the user to user communication
z	z	C4.3.4		Services to encourage collaboration (u2u) user to user services.
z	z	C4.4		Personalization responsive to e.g. cognitive abilities or work tasks or environments what data can and should be collected, how should it be collected, how should it be packaged

z	z	C4.4		Present search models are unaware of context since they were defined by assuming that there is one user, one information need for each search, one location where the user is, and no temporal dimensions. Contextual information includes both explicit and implicit knowledge about end users, systems and their environment. Such factors constrain the search without forcing the user to re-express his own information need explicitly and frequently. Context modelling and engineering is necessary to have effective searches. Investigation is necessary to: define new models to search in context, implement efficiently search in context, define new evaluation frameworks.
z	z	C4.4		Methodologies for building semantic user profiles based on his interactions with the multimedia data on different device types should be investigated
z	z	C4.4	D3.5	Personalization models should utilize multimedia content and context standards, as well as domain ontologies
z	z	C4.4		Awareness of location is important for cultural heritage (I am in Florence and …)
z	z	C4.4.3		but doing GPS and personalizing is future stuff (post 2010
z	z	C4.5		DL access through mobile devices: it is already here (within 2010 time frame);
z	z	C4.6		Nice example in Yannis_niceIAP051205.ppt
z	z	C4.6		Services can be built on an infrastructure of coordinated collections
z	z	C4.6		DLs have to provide core functionality (by means of services) to • access content in different quality • consider context and location • deliver content in different formats • make the results usable on different (mobile) devices
z	z	C4.6		Infrastructure to support access and re-use of content

z	z	C4.6		• DLs have to support the generation of applications on the basis of functionality
				  • definition of new services
				  • incorporation into the DL

z	z	C4.6		• DL Applications can make use of this functionality/services, thereby producing new content.
				• DLs must provide functionality/services to
				  • store and to incorporate this new content
				  • Keep track of provenance

z	z	C4.6		Different user groups will have different requirements for services and functionality
z	z	C4.6.2		Combined toolkits selecting appropriate tool for the job or merging results statistical, network analysis, query-expansion, logic-based Balance automatic and user control
z	z	C4.6.2		Advanced users want to save time, reduce the number of navigational clicks, more complex search functionality, more functionality within the documents themselves, e.g. access points within the documents themselves, not just top-level access, theme collections for the researchers? Personalization, contextual, visualization technologies)
z	z	C4.8		The User Interface • Different user groups will have different requirements for the user interface • Simplify, lower the learning curve, especially for new users, make using a new interface more intuitive • Creation of help texts / user documentation should be a core on ongoing part of the digital library development process • User Interface (more intuitive, richer, multi-lingual, non-textual)
z	z	C4.8.1		Visual interfaces
z	z	C4.8.1		Natural language and speech interfaces t should be investigated
z	z	C4.8.1 C4.4		Interfaces that exploit context and ontologies should be investigated
z	z	C4.8.1		importance of user-system interaction (e.g. search request refinement)
z	z	C4.8.1		General public - simple interface
z	z	C4.8.1	D1.2	interfaces (studied to facilitate interaction with user according to linguistic and cultural diversities)

z	z	C4.8.2		**Enhancing Usability for DL's** • Disseminate a "usability culture" in DL's communities • Usability evaluation as "pervasive" activity along the life cycle (not only at the end) • Methods, tools and processes for making usability evaluation: • Effective (making a difference in quality) • Efficient (within project budget)
z	z	C4.8.2		Add usability studies in existing projects (DELOS tasks + TEL + …) Do usability evaluation of the existing systems and prototypes
z	z	C4.8.3		**Accessibility** • Accessibility is a necessity for "public" applications • W3C guidelines are not sufficient • DL's in general do not even apply them • Design, operations and typical interfaces of DL's must be considered • Specific evaluation techniques, design methods, interfaces techniques are needed
z	z	z	D	consider in our comments the e-content Plus program, as it is a big part of i2010.
z	z	z	D1.2	hot_tech_Carol_CLEF1.ppt
z	z	z	D1.2	Increasing pressure for access to information without language or cultural barriers: • Find information in foreign languages • Read and interpret that information • Merge with information in other languages Need for Multilingual Information Access

z	z	z	D1.2	**What is Multilingual Information Access (MLIA)?**
				• MLIA related research regards the storage, access, retrieval and presentation of information in any of the world's languages.
				• Two main areas of interest:
				  • multiple language access, browsing, display
				  • cross-language information discovery and retrieval

z	z	z	D1.2	**Cross-Language Information Retrieval (CLIR)** Crossing the language barrier… • querying of multilingual collection in one language against documents in many other languages… • filtering, selecting, ranking retrieved documents • presenting retrieved information in an interpretable and exploitable fashion

z	z	z	D1.2	**What could we have now?**
				Note: individual items appear elsewhere
				• Multilingual centralized portals
				• Support monolingual search in multiple languages
				  • Character encoding issues / stopword lists / stemmers / morphological analyzers
				• Support simple cross-language search
				  • querying on metadata (central metadata registry) and keywords
				  • dictionary-based search / interlingua or pivot language
				  • thesauri for domain-specific search
				  • interactive search / browsing functionality
				• Present results in a simple fashion

z	z	z	D1.2	**CLIR methods**
				• How is it done?
				  • Translate: search requests, documents (or both)
				• Translation resources
				  • Machine Translation (MT)
				  • Parallel/comparable corpora
				  • Bilingual Dictionaries
				• Example problems
				  • Handling non-ASCII character sets
				  • Morphology: inflection, derivation, compounding, …
				  • OOV terms, e.g. proper names
				  • Multi-word concepts, e.g. phrases and idioms
				  • Ambiguity, e.g. polysemy

z	z	z	D1.2	Multilingual Portals • How many languages • how many levels should be multilingual • how to handle updates • linguistic and cultural dependent issues
z	z	z	D1.2	Europe's Cultural Heritage • Europe's collective memory is multilingual • Making our historic and cultural heritage available to all citizens for studies, work, leisure via the Internet implies efficient functionality to represent, store, access, interpret and reuse this material whatever the form, media, and language • Impact is social, cultural and economic
z	z	z	D1.2	In the EDL, include in each entry in a search results pointers to translations of the work, where available, or simply a "find translations" button.
z	z	z	D1.2	Multilingual component should support both information providers and information seekers.
z	z	z	D1.2	One goal: preservation of national languages

z	z	z	D1.2D 2	For multilingual services, deal first with collections and techniques in this priority order: • controlled collections, metadata only • controlled collections, full text • uncontrolled collections Three Basic Levels for Multilingual Search - search on catalog data (short-term) o (metadata / controlled vocabularies- with and without abstract) - search on full content (for all types of media - text, image, speech, …) (medium-term) - long-term (uncontrolled collections) Multilingual search of abstracts, notes Multilingual search of free text in libraries medium term open space, Web (requires cleaning) - long term Three Basic Levels for Multilingual Search search on catalog data (short-term) (metadata / controlled vocabularies – with and without abstract) search on full content (for all types of media – text, image, speech, …) (medium-term) uncontrolled collections (long-term)
z	z	z	D3.5	See hot_tech_Stavros_Multimedia DL.ppt
ZX				adjust role of libraries
ZX				US National Science DL

ZX				**Some references** A systematic compilation of ideas on how to transform a digital library from a repository into a power tool that supports entirely new ways of intellectual work, see DLib Magazine article A Framework for Digital Library Research: Broadening the Vision, http://www.dlib.org/dlib/december02/soergel/12soergel.html] Fourth, it is time to remember the main references for all of this: "Commission unveils plans for European digital libraries" http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/05/1202&format=HTML&aged=0&language=en&guiLanguage=en There are many references in the end of this page, but I'd like to point you especially this fundamental one (please read it): http://europa.eu.int/rapid/pressReleasesAction.do?reference=MEMO/05/347&form at=HTML&aged=0&language=EN&guiLanguage=en We know also that all of this "vision" of the EC is, not only, but in large part, a reaction to Google and especially to Google Print! To make it "worse", we have now also this vision for the "World Digital Library": http://www.loc.gov/today/pr/2005/05-250.html
ZX				Mathematical models for algorithms controlling global consistency of references to the same things and concepts in different contexts. Integration of automated methods with large-scale manual, collaborative control and correction
ZX				No one technique fits all kinds of data
ZX				Example of advanced cultural heritage archives from the US: Steve Griffin slides Delos Dec05a.ppt

ZX	B4			Points from hot_tech_Traugott_metadata_MDc.html not specifically classified " Creation o Differentiate provenance of metadata, trust, validation o Judge degree of format adherence o Granularity problem, e.g. for datasets o Support for multiple controlled vocabularies o Experiments with semi-structured metadata o Versioning " Packaging, complex objects o Need to manage complex digital assets/objects o Preserve context o Manage huge repositories of objects and metadata o METS, MPEG-DIDL approaches; IMS Content Packaging Spec.; SCORM. Some are not useful for frequently changing (meta)data o More programmatic approaches of working with multiple simple and complex objects o Packaging issues need more investigation from a preservation perspective " Use, reuse o Seamless connection between different types of collections: books/journals; special collections and archives; research and learning materials; freely accessible web resources o Merger between metadata and KO services o Knowledge extraction from metadata o Link between document metadata and citation indexing (ISI Web Citation Index) o Investigate how formal registries and informal social tagging might eventually overlap or converge o Use metadata programmatically: to 'FRBRize', to do collection analysis, to generate interesting displays _____ 0 III Specific metadata requirements for digitized documents and preservation " Needs to allow re-creation and interpretation of the structure and content of digital data over time: discovery, technical rendering of objects, recording of contexts and provenance, documentation of repository actions and policies " PREMIS Data Dictionary format 2005, based upon the OAIS reference model; about objects at different levels of aggregation, events, agents and rights; technical elements are still missing " Explore implications of exchanging metadata through heterogeneous digital archiving systems used for collaborative metadata management " Research data and multimedia products have had specific problems with insufficient incentives for their creators to do metadata " Hidden subjectivity and cultural bias is potentially more damaging for metadata to be used under long time. Data and its organisation must be historicized with as rich semantics and representation information as possible " In Cultural Heritage sector: object description turns into valuable work itself " Provide for long-term access and management