



**Counting the pace and cost
of conserving and accessing
our cultural heritage**

Cultural statistics in Europe

Digitisation in the US

**Framework for capturing
digitisation in Europe**

European digital library



Counting the cost of conserving and accessing our cultural heritage

One of Europe's key Information Society policy objectives is to make the content and digitally preserved materials of archives, museums and libraries more widely available. Across Europe these institutions are converting their 'analogue' collections into digital form thereby creating new opportunities for all interests to benefit from improved and convenient access to these resources. Numeric is a study financed by the European Commission that will define the empirical measures for digitisation activities and establish the current investment in digitisation and the progress being made by Europe's cultural institutions.

What is Numeric?

Across Europe, a quiet revolution is taking place as cultural institutions employ digitisation and cultural internet services to safeguard and widen access to their materials, but who knows at what cost? And how can we measure the relative progress that has been made? To answer these questions, the Institute of Public Finance¹ has been contracted by the European Commission² to undertake Numeric, a two-year study running until April 2009 to measure the investment made in the EU, and to catalogue the achievements of digitising Europe's cultural heritage and information materials.

Numeric will cover all 27 member states of the European Union. It will further anticipate the future take-up of the framework and survey methodology in EU accession/candidate countries.

International comparisons will be made with relevant other countries, including EFTA members, the United States, Canada, Australia, China and India.

What is the Numeric approach?

The most important precedent will be to establish appropriate and consistent definitions of 'digitisation' activities. This is a classic example of the problem of finding appropriate indicators that describe value rather than simply volume. There is such a diversity of materials to consider: books, journals, newspapers, photographs, antiquities, other objets d'art and documents, as well as the audio-visual materials of academic and scientific institutions, publishers, industry libraries, museums and archives. A census approach is clearly impractical, and has not proved successful for previous researchers, so Numeric intends to spend the first year of the study testing simpler statistical methods such as establishing the proportion of analogue materials (e.g. the physical documents) that have been digitised.

During the second year the digitisation activities will be measured, in order to complete the assessment and establish an ongoing survey methodology for the future.

1. The Institute of Public Finance Limited (IPF) is a company that is wholly owned by the Chartered Institute of Public Finance and Accountancy .

2. Information Society and Media Directorate General of the European Commission.



Who is Numeric?

The Numeric team consists of experts in the field of cultural digitisation, international public finance, libraries, museums, archives and statistics.

The team will collaborate with those already working in the diverse areas, including Eurostat and UNESCO, national ministries and statistical agencies, and cultural institutions. The Numeric study, is setting out to engage the interest and cooperation of partners working on other projects across the cultural sector in Europe and beyond. It is intended to make use of infrastructures already in place, rather than set-up duplicate or parallel systems and networks.

How to find us?

Progress can be followed through the Numeric website:

<http://www.numeric.ws>
numeric@ipf.co.uk

The study’s day-to-day work is managed by Phillip Ramsdale. If you believe that you have the experience and inclination to lend this study your advice and assistance, or in case you have any questions please contact him at:

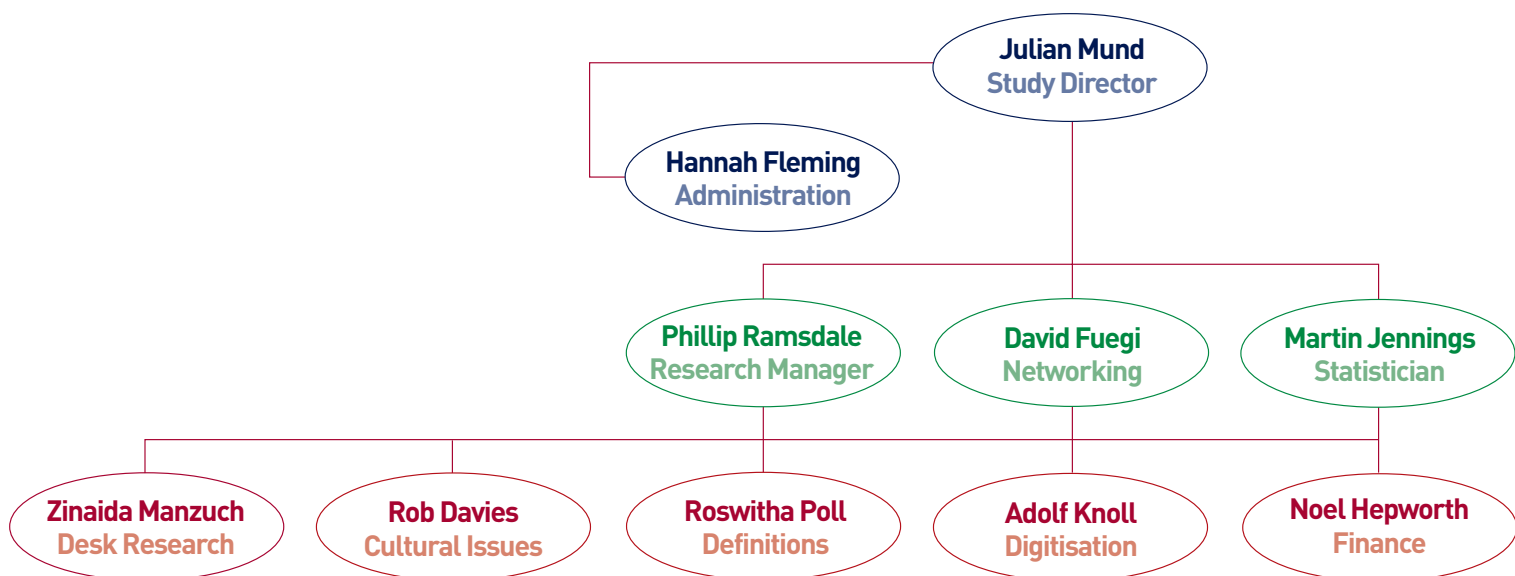
Phillip.Ramsdale@ipf.co.uk
 +44 (0) 20 8677 8508

In particular, if you are designing or launching a survey to measure digitisation activities, or you know of such an initiative, please let us know.

“NUMERIC will cover 27 member states of the European Union.”

“NUMERIC intends to establish appropriate and consistent definitions of digitisation activities.”

The Numeric team



Other Research

The experience in the United States shows that it is possible to gather statistical measures of digitisation of the cultural heritage in archives, libraries and museums.

One of the most comprehensive surveys of digitisation activities of cultural heritage materials is collated by the United States Institute of Museums and Library Services. The Institute is currently conducting its third survey, but results for those earlier samples gathered in 2002 and 2004 have been published. These show an inevitable difference in the certainties of planning for digitisation programmes, between the larger national institutions and the local public libraries. This arises as a result of the funding programmes made available to different organisations. Although it also follows that resourcing of digitisation efforts will be directed towards those institutions with the appropriate objects and content.

For instance, the response from **public libraries** demonstrates most uncertainty [Table 1].

By way of contrast, nearly six out of ten **archive services** were able to report the availability of resources for digitisation work, and the proportion increased to nearly three-quarters amongst the major State Library Administrative Agencies. (73% had earmarked funds for digitisation work in the previous year) [Table 2].

Table 1: Public libraries in the United States

In the past year did you have funding for...?			
	Yes	No	?
Technology	81	17	2
Digitization	12	71	17
Next year do you plan to have...?			
	Yes [%]	No [%]	? [%]
Technology	75	9	17
Digitization	20	52	29

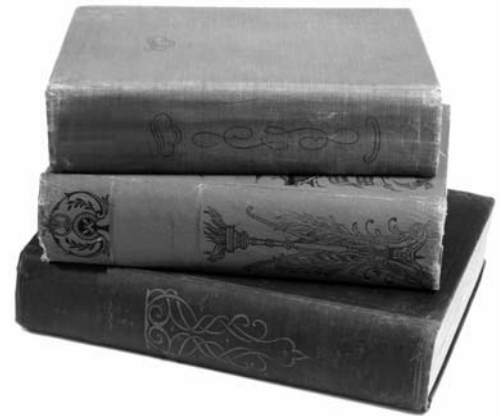


Table 2: Archive Services in the United States

In the past year did you have funding for...?			
	Yes [%]	No [%]	? [%]
Technology	76	20	4
Digitization	57	38	5
Next year do you plan to have...?			
	Yes	No	?
Technology	67	13	20
Digitization	59	19	22

When each institution was asked whether they were currently digitising or had digitised certain materials in the previous 12 months, the response was very varied. Different types of institution represent different analogue collections, and this is reflected in the type of material that most institutions reported. The following table ranks these materials by those most frequently quoted, and the top five are highlighted by each type of institution [Table 3].

In Europe there have been several surveys of which the most notable are:

- European Board of National Archivists (EBNA), 'Report on Digital Material in European National Archives' – by Pertti Hakala in November 2006 provides an overview of current activities
- The EDLproject survey of CENL members measuring digitisation efforts in National Libraries – conducted by Max Kaiser
- Also status reports arising from Minerva Europe and Michael projects

“Funding impacts on the certainty attaching to digitisation plans.”

Table 3: Institutions in the United States reporting digitisation of the following in the past year:

	Public Libraries	Archives	State Lib. Admin. Agencies
Photographs	4.8	17.5	2.7
Correspondence, etc.	2.4	6.5	12.8
Historical doc/archives	3.3	11.6	5.1
Maps	1.9	6.6	8.1
Government publications	0.0	1.1	15.4
Info re institution	4.8	5.4	5.3
Films, videotapes	1.0	6.5	7.9
Other items	0.0	5.0	10.0
Manuscripts	1.0	7.4	2.6
Images of items	1.5	6.5	0.0
Records re collection	1.4	3.2	2.8
Special exhibits	1.0	6.4	0.0
Newspapers	0.5	1.1	5.3
Education re collections	1.4	2.2	2.7
Journals/serials	1.0	0.0	2.7
Rare books	1.0	0.0	2.6
Music/rec sound	0.5	2.2	0.0
Course material	1.4	1.1	0.0
Sheet music	0.0	2.2	0.0
Theses/dissertations	0.5	0.0	0.0



Source for tables 1,2 and 3: 'Status of Technology and Digitization in the Nation's Museums and Libraries', US Institute of Museum and Library Services, Jan 2006.

Cultural Statistics in Europe

– An overview

Eurostat continues to work on the harmonisation of statistics on culture in close collaboration with Member States of the European Union and other international agencies such as UNESCO and the OECD.

The current sources of data draw on the EU Labour Force Survey, Structural Business Statistics, Eurobarometer, Community Survey [Table 5] on ICT usage in households and the Households Budget Survey. Drawing from these sources, the Département des études de la prospective et des statistiques (DEPS) of the French Ministry of Culture and Communications have produced 'Cultural Statistics in Europe – An Overview' for Eurostat. <http://epp.eurostat.ec.europa.eu/>

The ambition of Numeric is to provide information for this series in future. In the meantime, the rates of internet connection at home for households in each country. This provides an interesting context to the digitisation efforts designed to widen access of cultural heritage materials to wider audiences.

The Netherlands, Denmark, Sweden and Luxembourg (Grand-Duché) lead the rest of the European Union in the take-up of home connections, with more than 70% of households reporting internet links.

Table 4: Internet connection and visits to libraries and museums

Country	% of population making at least an annual visit to a:		% of households connected at home to the internet
	Library	Museum	
Netherlands	44	32	80
Denmark	61	46	79
Sweden	65	52	77
Luxembourg	19	32	70
Germany	29	33	67
Finland	68	38	65
United Kingdom	49	42	63
Slovenia	47	27	54
Belgium	27	23	54
Malta	20	21	53
Austria	18	30	52
Ireland	32	25	50
Estonia	50	31	46
Latvia	38	32	42
France	22	24	41
Italy	-	-	40
Spain	23	22	39
Cyprus	14	17	37
Poland	30	21	36
Lithuania	36	22	35
Portugal	17	16	35
Hungary	24	33	32
Czech Republic	33	39	29
Slovakia	33	26	27
Greece	9	14	23
Bulgaria	17	11	17
Romania	14	10	14

The colour shows the quintile in which the country lies for the distribution of indicators represented in the column.

Table 5: Level of internet access of households %

	2002	2003	2004	2005	2006
European Union (27 countries)	–	–	–	–	49
European Union (25 countries)	–	–	42	48	51
European Union (15 countries)	39	43	45	53	54
Euro area (EA11 – 2000, EA12 – 2006, EA13)	36	40	43	50	51
Netherlands	58	61	–	78	80
Denmark	56	64	69	75	79
Sweden	–	–	–	73	77
Luxembourg (Grand-Duche)	40	45	59	65	70
Germany	46	54	60	62	67
Finland	44	47	51	54	65
United Kingdom	50	55	56	60	63
Belgium	–	–	–	50	54
Slovenia	–	–	47	48	54
Malta	–	–	–	–	53
Austria	33	37	45	47	52
Ireland	–	36	40	47	50
Estonia	–	–	31	39	46
Latvia	3	–	15	31	42
France	23	31	34	–	41
Italy	34	32	34	39	40
Spain	–	28	34	36	39
Cyprus	24	29	53	32	37
Poland	11	14	26	30	36
Portugal	15	22	26	31	35
Lithuania	4	6	12	16	35
Hungary	–	–	14	22	32
Czech Republic	–	15	19	19	29
Slovakia	–	–	23	23	27
Greece	12	16	17	22	23
Bulgaria	–	–	10	–	17
Romania	–	–	6	–	14
United States	–	55	–	–	–
Australia	46	53	56	–	–
Korea (Rep of South)	70	69	86	92	–
Canada	51	55	60	61	–
Japan	49	54	56	57	–
Turkey	–	–	7	8	–
Iceland	–	–	81	84	83
Norway	–	60	60	64	69
Macedonia, (prev Yugoslavia)	–	–	11	–	14

“The opportunity to widen access to cultural heritage is growing as households connect to the internet.”

Emerging statistical framework – Defining the measures

It is, as yet, too early in the study to provide definitive guidance on survey schemes that should be adopted to ensure international comparability. However, the main elements of data that will be required are likely to be higher level summaries of the following conceptual framework [Table 6].

If you have any comments or queries about this framework, then please send your feedback to numeric@ipf.co.uk

The domains to be covered by this framework include:

- Audio-Visual Collections
- Archives
- Libraries
- Galleries and Museums

Once an agreed and practicable framework is established, national statistics will be collated on a consistent base across Europe.

Table 6: Draft survey main data elements

Heritage materials	Analogue collection			Digital output			
	Measure:	Total units:	'Priority units:'	Achieved ²		Planned	
				Units	Cost €	Units	Cost €
Rare books	Volumes						
Other books	Volumes						
Newspapers	Volumes						
Journals and other serials	Volumes						
Government publications	Volumes						
Other type printed material not classified above	Number						
Manuscripts	Number						
Maps	Number						
Photographs	Number						
Engravings	Number						
Drawings	Number						
Posters	Number						
Postcards	Number						
Sheet music	Scores						
Other images not classified above	Number						
Archived records of government/administration	Metres						
Archived records of historic importance	Metres						
Other archived records	Metres						
Man-made artefacts in museums	Artefacts						
Natural world specimens in museums	Objects						
Works of art – 2 dimensions	Exhibits						
Works of art – 3 dimensions	Exhibits						
Other objects in museum collections	Exhibits						
Film and video recordings	Hours						
Music and other recorded sound	Hours						
Other items not classified above	Number						

1. 'Priority' units are those items in analogue collections that need to be digitised, either for conservation/preservation purposes or to provide wider access. Music 'scores' are otherwise measured in physical units.

2. Not all of the 'total analogue collection' needs to be digitised. The overall progress of digitisation has to be measured by the sum of 'achieved' and 'planned' digital output as a proportion of the 'priority' analogue collection that has been identified as necessary to digitise.

Context for our advice – Initial scoping work

At present, the study team are devoted to networking, desk research, specifying the statistical framework, assembling data leading to a report in October/November 2007, on the way ahead. This will define the target population of the study for each country and explain the measurement concepts, definitions and classifications to be used, with a justification for the sampling frame and sources for the subsequent data collection exercises. These research findings, will also form a basis for checking the intended statistical framework with Eurostat and UNESCO prior to firming-up on the methodology advocated alongside the fuller analysis of data reported in January/February 2008. This later report is intended to outline the tested approach to measuring activity during 2008 and beyond. With this timetable in mind, it is intended to organise the initial research phase under four main headings:

- Taking note of the experience and understanding built-up in previous research and deriving from other associated projects – i.e. taking a systematic approach to the desk research
- Clarifying the framework for measuring digitisation activity – i.e. getting the definitions and classifications right
- Building a better understanding of the universe within which the digitisation activity is taking place – i.e. establishing the foundation database for scaling statistical estimates
- Reviewing the possibilities for robotic approaches to collecting data and seeking the assistance of industry experts in locating and costing digitisation projects – i.e. investigating the potential to gain statistical information from search engine providers and suppliers of digitisation services

1. Evaluation of prior surveys and desk top research

This research will aim to outline the 'State-of-the-art' of measuring digitisation activity and consider the usefulness of the previous survey questions. The analysis will seek to build on good practice/definitions and identify current initiatives or potential opportunities for exploiting the collection of data in future. This will inevitably involve consultation and dialogue with other projects and/or relevant institutions involved in these other studies.

2. Definitions/Classification of data

For the reasons described above, an appropriate guideline to settle what statistics are most required and what measures are feasible is required. This will be evaluated by examining previous surveys and associated research findings and consulting the interested contacts in each country on the suitability of various frameworks. The relationship between the cost and technical complexity of different digitisation procedures and the quality of the output will account for considerable variation in the statistics. Therefore, it will be necessary to consider a number of different classifications under which to group the data to best account for these 'drivers of difference'.

In the following list are examples of test categories:

[i] **Objective:** Preserve a true representation of the original, mainly audio-visual material, distribute access to either the most needed part or all the original content.

[ii] **Input/output:** Audio-visual –
 (a) 2 dimensions
 (b) 3 dimensions e.g. static objects, manuscript requiring careful handling/colour management, normal printed books, newspapers and other periodicals, normal one-sheet materials (postcards, photographs, etc.), difficult flat materials (plans, art objects, maps, etc.).
 For example '**Heritage materials**' framework described in the table [Table 6].

[iii] **Quality:** Still images –

- (a) Text
- (b) Graphics
- (c) Mixed, granularity (sound/picture definition), other standards e.g. with transcription of written text, attached metadata.

[iv] **Heritage type:** Museum, archive, library, audio-visual institution, other, architectural/archaeological/natural/scientific/industrial etc, academic/private/public collection/copyright.

[v] **Usage/access:** Limited to physical storage device, closed/limited access, e.g. at the institution, internet.

[vi] **Type of end user/exploitation:** Scientific/academic research, educational, commercial, professional/administrative, general open access, local history/community memory.

Needless to say, there are always 'Other' categories, and it will be necessary to balance what is useful for local classification purposes against a summarised classification that for the sake of consistency is feasible for higher level aggregation and international comparison. There are also reconciliations to consider in the established international classifications. The activities associated with 'Digitisation' and 'Culture' are often perceived differently and there are distinct developments surrounding the methods of digitisation and the policy priorities of different institutions. Differences in headline terms adopted for accounting purposes and those adopted by UNESCO/ISO and the more specific industry related definitions will also be reviewed. The Council on Library and Information Resources (CLIR) developed some terms in 1990 relating to digitisation and ISO have considered various indicator definitions more recently. However, it is now essential to define a practical framework for the mutual benefit of establishing measures as soon as

possible.

3. Foundation Database

This database is needed to establish the analogue universe, against which the digitisation activity can be scaled. Taking into account the necessity to refine and settle appropriate definitions and classifications (as described above), the basic data will include:

- Relevant cultural institutions by type; Analogue collections; Staffing (to help reference costs); Expenditure; Users/visitors (actual and virtual)
- Demographic/economic data for the sector/country

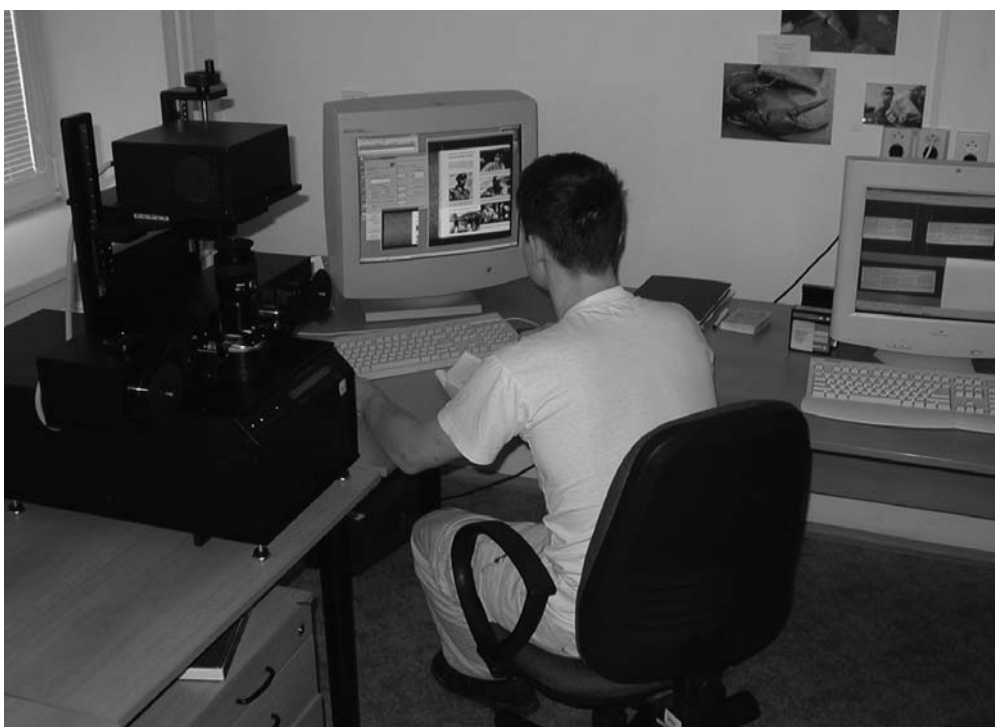
The data will be compiled from available international sources and will inevitably contain many missing values, and approximations will be made for such estimates to be scrutinised by suitable representatives active in the area so described. In this way, the scope and coverage of the data will be extended and improved. The foundation database will also be published to the numeric website and the scrutiny it receives from other researchers will also provide a mechanism for improving the accuracy

of the estimates and statistics.

4. Web-crawling/industry support

Whilst the concept of gathering statistics relating to the scale of materials available on-line by 'crawling' the known URL listing of cultural institutions seems attractive, there are limitations to such data. The main difficulty relates to the proportion of material that is on-line that originated from analogue sources and that which was 'born-digital'. Unless there are associated tags ('persistent identifiers') denoting whether the materials encountered during the robotic search were derived from original analogue sources, the data arising from a web-crawl will be unsuitable.

Nevertheless the concept needs to be explored with the Internet Archive. They are currently using Heritix to crawl the URL listings supplied by invited cultural institutions. The assistance of industry suppliers will also be sought to help locate and estimate the cost of digitisation on a range of different projects. Such assistance will be sought to better inform the range of expected cost benchmarks, and these can help in checking the plausibility of other survey data and making estimates of the cost of digitisation where only the volume and type of digitisation activity is known.



Picture: Digitisation of materials onto computer. **Source:** National Library of the Czech Republic.

CENL/EDLproject digitisation questionnaire

CENL is currently collecting information from its members on digitisation. CENL [Conference of European National Librarians] <http://www.cenl.org/> has more than 40 European national libraries as members. CENL is actively supporting the EU's efforts in creating a European Digital Library and its members are involved in the EDLproject <http://www.edlproject.eu/> which is taking forward the expansion of The European Library <http://www.theeuropeanlibrary.org/portal/index.html> by supporting the addition of nine more national libraries to it as full members.

EDLproject (www.edlproject.eu/) is co-funded by the European Commission under the eContent^{plus} Programme and coordinated by the German National Library. The project, which started in September 2006, is a direct response to the request of Viviane Reding, Commissioner for Information Society and Media, made at the CENL conference in Luxembourg on 29 September 2005 that national libraries should use their influence in the debate on the digitisation of European content for access through the web. One of the objectives of EDLproject is to leverage the influence and resources of CENL as a key player and stakeholder in the content field to work towards a consensual resolution of certain issues raised by the Commission's *i2010: Digital Libraries* initiative, such as potential availability of digital content from national libraries.

Recognising the current political importance of digitisation, CENL aims to establish an even better strategic overview derived from good information on what its members have done and are doing in the field of digitisation of their collections. Within the context of EDLproject, Dr Hans Petschar of the Austrian National Library, Chair of the CENL Content Working Group, is leading the effort to collect extensive and up to date information by means of a series of questionnaires. The intention is to continue to collect the data on a regular basis if this first round succeeds.

The questionnaires cover:

- Institutional digitisation programmes and strategies for the next 3-5 years
- Digitised collections and ongoing digitisation projects in national libraries

The results of the survey will form the baseline of a roadmap to be produced by the CENL Content Working Group and EDLproject to describe the issues in digitisation in Europe and possible solutions. The results of the survey should give an overview on institutional digitisation programmes and policies and provide measurable data for content digitisation. The report is scheduled for completion in the autumn.



“CENL recognises the current political importance of digitisation.”

IPF

IPF is a support services company wholly owned by CIPFA. We specialise in financial advice and governance, property and asset management solutions, the supply of information, expertise and people with the skills to help you at the highest level.

info@ipf.co.uk
www.ipf.com

CIPFA

The Chartered Institute of Public Finance and Accountancy (CIPFA) is the leading professional body for public services, whether in the public or private sectors. It provides education and training in accountancy and financial management, and sets and monitors professional standards. Its professional qualification is high quality, relevant and practical, and is supported by a range of other products and services.

Numeric

Numeric is a study funded entirely by the Information Society and Media Directorate General of the European Commission.

If you require further information or in any other format, please contact:

Phillip Ramsdale
Research Manager
Phillip.Ramsdale@ipf.co.uk

Martin Jennings
Statistician
Martin.Jennings@ipf.co.uk

numeric@ipf.co.uk
http://www.numeric.ws



European Commission
Information Society and Media

Cover picture: Centre Pompidou, Paris

IPF is a company owned by the Chartered Institute of Public Finance and Accountancy

Registered Office: 3 Robert Street, London, WC2N 6RL