



Users and Uses of Online Digital Libraries in France

ECDL 2003

Houssein Assadi, Thomas Beauvisage (France Telecom R&D)

Catherine Lupovici, Thierry Cloarec (Bibliothèque Nationale de France)

Present document contains information proprietary to France Telecom. Accepting this document means for its recipient he or she recognizes the confidential nature of its content and his or her engagement not to reproduce it, not to transmit it to a third party, not to reveal its content and not to use it for commercial purposes without previous FTR&D written consent.

(confidential)

BibUsages Project: Presentation and Objectives



▶ **Partners: France Telecom R&D, Bibliothèque Nationale de France**

▶ **Objectives:**

- ▶ Describe the uses made of online digital libraries, crossing them with the characteristics of the user base.
- ▶ Detailed contextual view of the uses of digital libraries, through a finely-tuned analysis methodology.
- ▶ Reveal how the developing DL uses are modifying long-standing practices.



Introduction : scope and methodology

- ▶ DL users
- ▶ DL uses in context
- ▶ Typical uses and users profiles
- ▶ Conclusion

Digital Libraries : the scope of our study



▶ Online "digital libraries"

- ▶ Institutional collections, e.g. [Gallica](#): the digital collection of the Bibliothèque Nationale de France
- ▶ Initiatives sponsored by associations, such as [ABU](#) (Association des Bibliophiles Universels)
- ▶ Online newspapers and magazines
- ▶ etc.

→ Online text collections in text or image formats, including specific formats for PDAs

Project Phases / Methodology



▶ Phase 1: preliminary survey

- ▶ Online questionnaire on the Gallica site, March 2002: 2 340 replies

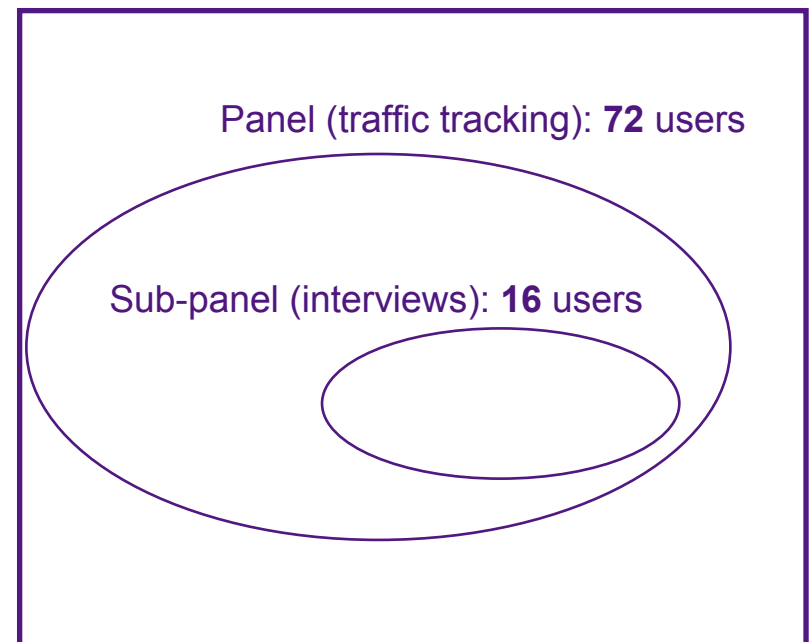
▶ Phase 2: set up a panel and track traffic

- ▶ Set up a panel based on replies to the questionnaire: 72 volunteers were selected, May 2002
- ▶ User-centric Web traffic tracking (a software installed on the user's PC allows Web usage tracking): July – December 2002

▶ Phase 3: qualitative survey of a sub-panel

- ▶ 16 interviews, October 2002, 3-part structure: use of the Internet, use of digital libraries and Gallica, relationship with "off-line" practices

Preliminary survey: **2 340** replies



Outline



▶ Introduction : scope and methodology

DL users

▶ DL uses in context

▶ Typical uses and users profiles

▶ Conclusion



Digital Library Users

From a socio-demographic point of view

- ▶ **Rather masculine population** (69,3% men vs. 58% for the French Internet users population at the same period) , **people over 45 years-old are majoritary in our panel** (43.7% vs. 15.8% for the French Internet users population at the same period)
- ▶ **Strong representation of academic users, as expected: postgraduate (teaching staff and students) plus secondary school levels.**
 - ▶ Occupation (for the 82% who have a professional activity): **45%** of them hold a **public service executive position** in the education or research sector.
 - ▶ 23% of those not in the workforce are **retired** (they are more highly represented here in the panel). Moreover, it should be noted that 20% of them are graduate students.

Digital Library Users as Internet users



▶ They are **advanced and active Web users**

- ▶ Experienced Internet users: 70.9% connected since 1999 and earlier, 47% since before 1997
- ▶ Over-representation of broadband connections: 39% vs. 9% in France at the same period
- ▶ Generally speaking, these are very heavy Web users : 7.6 sessions a week on average (vs. 2.4 for the French Internet users population at the same period)

Outline



▶ Introduction : scope and methodology

▶ DL users

DL uses in context

▶ Typical uses and users profiles

▶ Conclusion



The main categories of visited web sites

- ▶ Frequent use of search engines and general-interest portals
- ▶ Very high traffic to "cultural" portals: digital libraries, e-commerce sites selling cultural goods (Alapage, Fnac, Amazon, etc.) and media sites are very popular
- ▶ Sites devoted to genealogy (denoting a very strong personal centre of interest) are highly ranked

Type of site	Number of sessions	Number of panellists	% of sessions	% of panellists
General-interest portal	8005	71	51,9%	98,6%
Search engine	4183	67	27,1%	93,1%
Personal website	3142	70	20,4%	97,2%
Digital library	1195	57	7,8%	79,2%
Media / Press	1144	49	7,4%	68,1%
e-commerce / cultural goods	833	63	5,4%	87,5%
Genealogy	684	31	4,4%	43,1%
e-commerce / Bank, Stock exchange	443	30	2,9%	41,7%
WebMail	439	14	2,8%	19,4%
Media / Radio	305	34	2,0%	47,2%
Media / TV	219	38	1,4%	52,8%
e-commerce / tourism	141	46	0,9%	63,9%
e-commerce / shopping	9	4	0,1%	5,6%



Search Engines

- ▶ Search engines are used in 32% of sessions, as against 20% of sessions for the NetValue panel (representative of French Internet Users) in 2000.
- ▶ 15 search engines used, with Google far ahead (present in 79% of sessions using a search engine)
- ▶ Very specific search topics according to the user, including a large number of names of people or places

TOP 20 queries, by number of sessions		
Query	Number of sessions	Number of panellists
villemoisson sécheltes	41	1
casanova à l'opéra	38	1
casanova saint-savony	29	1
villemoisson-sur-orge	25	1
NOGAREDE	23	1
sncf	22	14
saint-savony	22	1
fnac	19	8
caramail	18	2
TF1	17	3
bnf	16	6
internet explorer	14	5
villemoisson sur orge	14	1
casanova opéra	13	1
montpellier	13	1
genealogie	12	5
sète	12	1
Télégraphe Chappe	11	1
tfou	11	1
google	10	6

TOP 20 queries, by number of panellists		
Query	Number of sessions	Number of panellists
sncf	22	14
fnac	19	8
bnf	16	6
google	10	6
internet explorer	14	5
genealogie	12	5
Michelin	7	5
gallica	6	5
darty	8	4
hp	8	4
anpe	6	4
camif	6	4
pages jaunes	6	4
ratp	6	4
mappy	5	4
paris	5	4
sexe	5	4
annuaire inversé	4	4
la poste	4	4
le monde	4	4



Digital Library Audience

- ▶ 17 sites identified as "digital libraries", with a rather wide definition (online text collections)

- ▶ Gallica is the leading site in terms of number of sessions and panellists, but the other sites of this category are also regularly visited
 - ▶ The interviews suggest that Internet users regard Gallica as one of a number of text sources.
 - ▶ The richness (volume and quality) of the collection accounts for the higher traffic than on other sites.

Portal	Number of sessions	Number of panellists	Average time per session
BNF-Gallica	822	54	21.4
BNF-Autres	577	51	6.5
ABU	31	13	3.0
Revue.org	22	12	1.8
Bibliothèque de Lisieux	20	11	3.9
Bibliopolis	15	6	8.0
Athena	14	10	1.1
ClicNet	14	8	0.6
Online Books Page	10	4	1.7
Electronic Text Center	8	5	0.7
BN Canada - Numérique	5	3	1.0
American Memory	4	2	6.3
Arob@ase	4	3	0.4
Berkeley DL	3	1	7.3
eLibrary	3	3	0.5
Gutenberg project	2	2	0.8
Alex Catalogue	1	1	0.8
Bibelec	1	1	0.0

Media sites Audience



- ▶ Media sites, in particular reference newspapers and magazines, show high audience scores

Portal	No. panellists	No. sessions	No. sessions per panellist	Average time per session (min.)
Le Monde	37	719	19.4	8.1
Libération	24	86	3.6	3.9
Les Echos	17	118	6.9	5.3
Nouvel Obs	14	131	9.4	4.9
Telerama	14	107	7.6	6.3
Le Figaro	13	148	11.4	14.2
New York Times	6	81	13.5	7.7
Le Point	4	12	3.0	3.4
Paris Match	3	31	10.3	5.8

E-commerce sites Audience



▶ The use of e-commerce sites selling "cultural goods" (eg. Amazon) reveals a significant number of book-lovers, with high traffic on Chapitre.com and Livre-rare-book.com

▶ Web sites of publishers are also visited

Portal	No. panellists	No. sessions	No. sessions per panellist	Average time per session (min.)
Amazon	55	386	7.0	4.1
Fnac	42	299	7.1	2.4
Alapage	20	107	3.7	2.2
Chapitre.com	26	160	6.2	9.3
Livre-rare-book	17	92	5.4	5.0
Galaxidion	6	73	12.2	6.4
Librissimo	6	7	1.2	4.5
Numilog	5	14	2.8	4.7
Eyrolles	3	4	1.3	1.0
CNRS Editions	2	8	4.0	8.6
Barnes & Noble	1	1	1.0	0.4
Cylibris	1	1	1.0	0.2
Edibook	1	1	1.0	0.1
Les Introuvables	1	1	1.0	0.3
Litraweb	1	1	1.0	1.3



Navigation on the Gallica site

- ▶ **The 1 063 sessions including access to Gallica are on average twice as long as the others:**
 - ▶ 1 h 01 min. on average for sessions with Gallica
 - ▶ 28 min. on average for all sessions combined
 - ▶ In a session with access to Gallica, the total time spent on Gallica averages 24 mins.
- ▶ **Multiple activities and alternating navigation:**
 - ▶ **In 52% of cases (558 sessions), the time spent on Gallica occupies a single sequence and is not alternated with visits to another site**
 - ▶ Only 22% of sessions comprise 2 distinct sequences on Gallica.
 - ▶ Only 10% of cases comprise 5 or more sequences on Gallica.

"Multi-tasking" is rare: a visit to Gallica is seldom alternated with visits to other sites (in 75% of cases, there is no alternation) .
- ▶ **Services used**
 - ▶ The majority of instances concern services used on Gallica *via* the internal search engine and are often followed by downloads
 - ▶ Consultation of the files remains marginal (cf. interviews) in session numbers but concerns half of the panel, who visit "out of curiosity" (confirmed in the interviews).

Outline



▶ Introduction : scope and methodology

▶ DL users

▶ DL uses in context

Typical uses and users profiles

▶ Conclusion



Typical user profiles

▶ Several typical user profiles have been identified (but this is not a general typology of DL users!):

- ▶ The **information seeker**:
 - Interest in Gallica as "primary information source"
 - Gallica is one of a number of sources, used in conjunction with search engines
 - Research in a professional context or in connection with a hobby or personal interest (genealogy and personal history of prime importance)
- ▶ The **book-lover**: Gallica is used to prepare a purchase (or a loan); the goal remains the book as an object.
- ▶ The **"on-screen reader"** (atypical profile): shift in keen readers' habits from printed media to the screen (PC and PDA)

Outline



▶ Introduction : scope and methodology

▶ DL users

▶ DL uses in context

▶ Typical uses and users profiles

Conclusion

Conclusion



- ▶ **DL attract a new public, not only the regular users of classical libraries.**
- ▶ **User profiles:**
 - ▶ Students and teachers: the expected profile, resources for courses and "professional" research
 - ▶ But also the less familiar profile of the senior-advanced-web-user, and the non-professional researcher with very specific topics (often dealing with her/his own family or region history)
- ▶ **Reading online is rare. Users are more likely to "consult" or "surf". The focus is more on reference books than great works, for which users often turn to paper.**
- ▶ **Digital libraries hold a very specific position in the Internet world, from the usage point of view**
 - ▶ Connecting with the private sphere: personal web pages of "enthusiasts and specialists" (eg. Genealogy and local/family history)
 - ▶ Connecting with the commercial sphere (second-hand booksellers and e-commerce sites)