

Bibliothèque nationale de France (BnF)

BnF mission summary:

- Build up and enrich its collections
- Preserve its collections, developing suitable techniques
- Communicate and promote its collections to the public
- Provide its readers with a complete reference catalogue, with coherent, multiple and easy to use access to its collections
- Cooperate in scientific activities at a national and international level, primarily within Europe
- Participate in research activities
- The BnF is actively involved in producing and making available digital documents and data.

At a national and international level, it is involved in three main public digital libraries projects: Gallica, BnF Digital library, “Europeana” (European Digital Library) project, French-speaking Digital Library project.

Gallica

Gallica is the online public access digital library of the Bibliothèque nationale de France: the project started in 1997 and the digital library was released in 2000. It is always been conceived as the BnF digital library on the Web, and intended to be a tool for researchers and teachers. Gallica 2 (<http://gallica2.bnf.fr/>) is the new enhanced version of the original BnF digital library (<http://gallica.bnf.fr/>): version beta was released in October 2007 and it is currently under implementation and will be progressively enriched over the coming months. The interface Gallica 2 will coexist a few months with the previous version of Gallica, before taking its place in 2008 definitively when all the functionalities will be operational and that the whole of the documents of the old version will have been transferred to Gallica 2. It was conceived as a digital encyclopaedia (copyright-free and available free of charge) and consists of printed materials (books, journals, newspapers, printed music, and other documents), graphic materials (engravings, maps, photographs, and others), and sound recordings. Gallica makes it possible to find and access rare, unusual, out-of-print, or difficult to find documents. From Antiquity to the First World War, it currently includes 100,000 digitized documents and monographs, 2,600 manually transcribed text documents, 250,000 fixed images, 500 audio documents. This digital library is partially accessible onsite the BnF and on the Web. At the end of 2010, the BnF intends to make available via Gallica about 400,000 digitized documents, available in image and text mode, for a total of more than 45 million pages.

Gallica preservation and planning activities

Acceleration of the increase in the numerical collections (currently higher than 100 Tb per year) and diversity of their digital formats at BnF. Currently raw master files and post-processed files are being preserved in Gallica. Derivates are being preserved because it is not possible to separate the dissemination part from the archived part. Backups are on CDs as transportation medium and magnetoptical disks. Two storage areas are available, one within the BnF and another in a site outside Paris. They are synchronized and backed up every hour, preserving everything within Gallica: XML, identifier, digital TIFF documents, directories for resolution (dissemination copies) compressed file formats. In 2005, the BnF announced a project for designing and developing a Distributed Archiving & Preservation System (SPAR) to store digital objects in a protected and long-lasting environment. SPAR will be operational from the beginning of 2009.

Technical infrastructure

Documents to be included in Gallica are selected by the BnF and digitized by external companies contracted by BnF. The formats of digitization required by the BnF are specified in its technical digitization manuals, with a resolution never lower than 300 dpi. These formats are then transformed for the consultation in Gallica 2. Digitized documents are ingested into the Gallica system, and the BN-OPALE PLUS catalogue – which is completely separated from Gallica - is updated. The catalogue lists all the digital collections of BnF and gives direct access to Gallica materials. A software takes care of security issues, converts on the fly TIFF in PDF, and also handles the rights with Java Advance Imaging for handling TIFF natively. Digital objects are described according to simple Dublin Core Metadata Element Schema. For the works digitized by the BnF, these descriptive data can be downloaded in XML according to OAI-PMH. Gallica 2 is developed with J2EE (Java 2 Enterprise Edition) and free software for some modules. The Apache Lucene search engine full-text provides a search inside the OCRed documents and results classification. Users have the possibility of creating a personal space where they can store their documents and bookmark pages. Other search functionalities, such as advanced search, Boolean operators and semantic tools will be gradually implemented in the next versions of Gallica 2.