

# Projekt Gallica

Anežka Koublová, UČO 373772, 5. semestr B-IN-UMI

5. prosince 2012

Esej pro předmět PV070 - Digitální knihovny  
Gallica, projekt Francouzské národní knihovny, online na <http://gallica.bnf.fr>



## 1 Základní informace

Gallica je digitální knihovna Francouzské národní knihovny (Bibliothèque nationale de France, dále jen BnF) založená v roce 1997. Obsahuje vybrané důležité historické dokumenty z různých oborů – zejména historie, literatury, přírodních věd, filosofie, práva, ekonomie, politologie. Ačkoli se zaměřuje na původní frankofonní dokumenty, zpřístupňuje i některá cizojazyčná díla, a to buď v originále, nebo francouzském překladu. Zpracovává různé druhy médií – tisk (knihy, periodika, noviny; v textovém i obrazovém formátu), dále rukopisy, zvukové nahrávky, ikonografické dokumenty a mapy a plány.



Areál BnF

## 2 Historie a současný stav

Za první pokus o vytvoření digitální knihovny lze považovat *Project Guttenberg*, který se už v roce 1971 pokusil o zpřístupnění textů se statusem volného díla. Tato dobrovolnická organizace stále funguje a v současnosti nabízí k volnému stažení přibližně 40 tisíc elektronických knih [4].

Digitální knihovny tak, jak je vnímáme v současnosti, začaly vznikat v průběhu devadesátých let. První reálnou snahou o vytvoření digitální knihovny „dnešního typu“ byl projekt *DLI-1* (Digital Library Initiative Phase 1) v roce 1994 [5]. Další organizace pak v digitalizaci textů a dalších kulturních památek, jejich shromažďování a zpřístupňování pokračují dodnes.

### 2.1 Podrobná historie projektu Gallica

**První verze (1997)** Gallica je jedním z nejstarších a největších evropských projektů svého druhu. Byl odstartován v roce 1997 s cílem stát se „virtuální knihovnou“. Během prvních tří let prvotního provozu byl nasbírán dsotatek podnětů k tomu, aby byla v roce 2000 spuštěna její druhá verze. Největším pokrokem oproti dřívějšímu bylo z hlediska uživatele rozhodně mnohem intuitivnější a detailnější vyhledávání v poskytovaných záznamech, které byly lépe označeny metadaty a nabídly tak nové možnosti a značné zrychlení jejich třídění [1].

**Gallica2 (2000)** Po spuštění nové verze byla započata velká akce, která měla za cíl digitalizovat celostátní francouzský denní tisk a publikace francouzské Akademie věd (Société savante).

**Soupeření s Google Books (2005)** Dalším milníkem ve vývoji Gallicy byl v roce 2005 rychlý vzestup iniciativy *Google Books*, který způsobil snahu o zvýšení počtu digitálně přístupných médií ve všech knihovnách, tedy i v BnF. V

následujících měsících začalo růst množství digitalizovaných dokumentů mnohem rychleji než dosud. Ve velkém se používaly a nadále používají OCR (Optical Character Recognition, optické rozpoznávání znaků) aplikace, které umožňují převod skenovaných dokumentů do textové podoby, která má mnoho výhod: menší velikost, možnost vyhledávání a úprav.

**Europeana (2007)** V roce 2007 se Gallica stala partnerem *Europeany*, iniciativy, jejímž cílem je sjednotit a podpořit národní digitální knihovny z celé Evropy. Prvními členskými zeměmi se stalo Maďarsko a Portugalsko. I přes snahu všech zúčastněných není toto sjednocení jednoduchým cílem. Je pro něj třeba velká finanční i legislativní podpora, která je problémová i v západních zemích, proto bude zapojení většího množství zemí ještě léta trvat [3].

**Překročení hranice 1,5 milionu dokumentů (2011)**  
V druhé polovině roku 2011 překročil počet dokumentů v databázi 1,5 milionu a jejich počet stále rychleji roste.



Logo Europeany

## 2.2 Současný stav

Nejnovější funkcí knihovny je *Gallica intra muros* (Gallica uvnitř hradeb), která zpřístupňuje dokumenty chráněné autorským zákonem a to pouze z domény výzkumné části BnF. Služba byla spuštěna v březnu 2012 a poskytuje do té doby postrádanou možnost přistupovat jednoduchým a efektivním způsobem i k dokumentům „sous droits d’auteur“ – „pod autorským právem“. Touto funkcionalitou napodobuje klasické knihovny a knihy s možností prezenční výpůjčky do studovny a znemožňuje jejich kopírování a další distribuci, která by v případě těchto děl byla protizákonná. Zároveň možnost „prezenční výpůjčky“ rozšiřuje o možnost pracovat s některými dokumenty jako s digitálním textem, který lze mnohem rychleji zpracovávat než klasickou papírovou knihu.

Počet nabízených dokumentů ke konci roku 2012 je přibližně 2 miliony. Je jedním z nejdůležitějších zdrojů například pro *Wikisource*, jednu ze součástí nejpoužívanější veřejné internetové encyklopedie Wikipedia [7].

## 3 Cíle a výsledky projektu

Gallica je zaštitěna Francouzskou národní knihovnou, z jejíž sbírek také pochází podstatná část digitalizovaných zdrojů. Francie je s Gallicou a dalšími národními digitálními knihovnami spolu s Německem největším průkopníkem evropské digitalizace.

**Finanční zdroje** Projekty zabývající se digitalizací se narozdíl od jiných nemohou spoléhat na zdroje z fondů Evropské Unie, která dotaci tohoto odvětví

odmítá. Gallica je tedy financována zejména francouzským ministerstvem kultury a soukromými sponzory.

**Obsah** Gallica vznikla ke konci devadesátých let s vizí zpřístupnění francouzského kulturního a vědeckého dědictví široké veřejnosti. Knihovna se snaží zveřejňovat dokumenty ze všech vědních oborů, historické i současné.

V poslední době se zaměřuje zejména na digitalizaci periodik minulých století a na masovou produkci textů vzniklých pomocí OCR. Jedná se totiž o jedno z provozně nejlevnějších odvětví digitalizace, mj. proto, že výtisky těchto časopisů má BnF trvale k dispozici ve svých sbírkách.

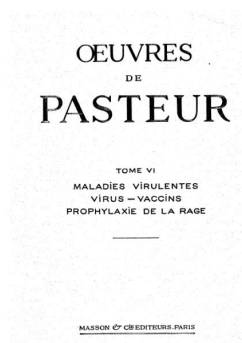
**Forma** Dokumenty jsou k dispozici v různých formátech. Historické tiskoviny jsou zpravidla naskenované, některé z nich převedené pomocí OCR do digitálního textu. Současné *born-digital* (vzniklé přímo v digitální podobě) publikace jsou přebírány od tvůrců a distributorů přímo v elektronické podobě. Jedná se zejména o širokou škálu periodik – od všeobecných přes zaměřená až po vědecké časopisy.

**Cílová skupina a obory** Účelem digitální knihovny Gallica je vedle zachování dokumentů také jejich distribuce co nejširšímu okruhu zájemců. Tuto možnost mohou využít jak laicové, kteří se o danou oblast nebo konkrétní dokument zajímají, tak studenti nebo akademičtí pracovníci a další odborníci, pro které by jiný způsob získávání záznamu byl složitější a zdlouhavý. Gallica není zaměřena na konkrétní vědecké obory, naopak se snaží o co nejširší záběr dostupných publikací [2].

**Spolupráce** Gallica spolupracuje s týmem specialistů, kteří obnovují nebo vytvářejí obsahy, rejstříky a další užitečná metadata k dokumentům. Tato spolupráce umožňuje uživatelům efektivněji vyhledávat nejkvalitnější a nejrelevantnější výsledky.

Dále nabízí spolupráci dalším digitálním knihovnám (zejména specializovaným), které mohou sdílet zdejší digitální materiál. Tato spolupráce je zjednodušena zejména možností *metadata harvesting*, tedy sběru metadat ve velkém. Z těchto dat si potom knihovny vyberou záznamy, které jsou pro ně relevantní a mohou je zpřístupnit uživatelům spolu s odkazem na samotný materiál v Gallice.

Projekt je také otevřen příspěvkům z dalších zdrojů, největším „outsourcovaným“ příspěvkem jsou Rousseauovy rukopisy, které digitalizovala poslanecká Knihovna Národního shromáždění a poté je dala k dispozici BnF, respektive Gallice. Knihovna Národního shromáždění je zároveň majitelem mnohých unikátních dokumentů z francouzské historie, proto probíhají další jednání o jejich digitalizaci a zpřístupňování [6].



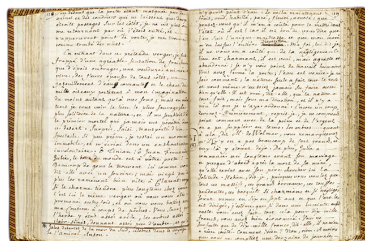
Titulní strana  
Pasteurových spisů

## 4 Hlavní přínosy projektu

### 4.1 Vytvoření kolekce historických dokumentů

*Gallica* je starší tvar slova Gaulois – tedy Gal, galský. Projekt se tak již svým názvem hrdě hlásí k historii Francie a dalších frankofonních zemí (v kontextu Galského národa pak zejména k Belgii).

Podobně jako například projekt *American Memory* Kongresové knihovny se



Rukopis Nové Heloisy

snaží shromažďovat, uchovávat a dávat k dispozici ke studiu co největší část historických dokumentů. To bylo zejména cílem první etapy digitalizace fondů BnF. Bylo třeba vybrat důležité a zajímavé dokumenty, ať už se jednalo o dokumenty z oblasti práva, krásné literatury nebo například soukromé rukopisy osobností, které hrály důležité role ve francouzské historii. I v současnosti však pokračují

jednání a realizace získávání přístupu k dokumentům v soukromých nebo jinak chráněných sbírkách. Jedná se například o již zmíněnou Knihovnu Národního shromáždění a o různé soukromé sběratele a historiky.

**Výhody** Takto vytvořená sbírka má mnohé nesporné výhody, které se většínou překrývají s obecnými důvody pro vytváření digitálních knihoven.

- V případě vhodného využití fyzických úložných médií a dostatečného počtu záloh mají digitálně zálohovaná data větší trvanlivost. Jedná se o „rezervu“ pro případ živelných nebo jiných pohrom, které mohou originály nenávratně poničit.
- Jak již bylo několikrát zmíněno, manipulace s digitálními daty je mnohem rychlejší a efektivnější než s originálními dokumenty.
- Digitální verze podporují ohleduplnost vůči originálům.

Pokud studujeme jejich obsah a využíváme naskenovanou nebo rozpoznanou verzi, nemá ani tisíceré čtení a listování dokumentem žádný vliv na stav originálu. Díky tomu je možné zachovávat tyto památky co nejméně poškozené. Do své kopie dokumentu si dokonce můžeme zapisovat poznámky, což by nám knihovna v případě zapůjčení originálu rozhodně neodpustila.

- Odborníci i laikové získávají přístup ke spoustě dokumentů, které by jim byly bez digitalizace zcela „zapovězeny“ – ať už se jedná o cenné památky, které by nadměrně častá manipulace poškodila, nebo dokumenty, které se fyzicky nacházejí na druhém konci státu nebo dokonce Země.

**Úskalí** Na druhou stranu však existují i jistá rizika a problémy, způsobené buď nedokonalostí člověka a jím vytvořené technologie, nebo jeho přístupem.

- Skenované dokumenty zpracovávané pomocí OCR stále ještě nemají sto-procentní úspěšnost s převodem obrazu do prostého textu.

S každým rokem toto procento roste, stále však není možné se na tento převod naprosto spolehnout. Někdy je tedy třeba konzultovat i původní verzi, stále však zůstáváme v rovině digitálních dat – pouze musíme část textu nahradit obrazem. U velmi důležitých nebo často využívaných dokumentů je možné provést korekci, jedná se ale bohužel o časově (a tedy i finančně) náročný proces.

- Existence digitální kopie textových a obrazových historických památek může svádět (zejména v blízké budoucnosti) ke snížení respektu a péče o jejich originály.

V současné době většina lidí chápe, že se jedná o důležité kulturní dědictví národa a často i celého lidstva, další generace by však postupně mohly toto povědomí ztrácet. Je proto třeba zdůrazňovat význam těchto památek a učit další generace, že to, že máme například Bibli v digitální podobě a všichni máme po batozích čtečky, neznamená, že originál uložíme do vlhkého sklepa, kde nebude překážet, ale že mu musí být věnována adekvátní péče.

## 4.2 Digitalizace denního tisku minulých století

Poslední dobou se do popředí zájmu digitalizace fondu BnF dostal denní tisk zejména devatenáctého a počátku dvacátého století. Jedná se o přibližně tři miliony stránek těchto titulů: Le Temps, Le Figaro, L'Humanité, La Croix, Le Petit Parisien, La Presse, Le Matin, L'Action française, Le Petit Journal, L'Aurore, La Justice, Le Constitutionnel, Le Gaulois, Le Siècle, Le National, Le Rappel, L'Intransigeant, L'Univers, Gil Blas, La Lanterne. Některé z těchto časopisů a novin jsou stále vydávány a patří mezi nejčtenější a nejserióznější tiskové zpravodajství ve Francii. Plánovaný čas věnovaný této akci je 5-7 let [2].

Přístup k těmto periodikům umožní získávání lepšího přehledu o tomto období francouzské historie. Je v nich možné dohledávat i „obyčejnější“ události než ty celostátního významu.

## 4.3 Spolupráce s projektem Europeana

Jak bylo zmíněno výše, je Gallica jedním z nejvýznamnějších přispívatelů projektu Europeana, evropské iniciativy pro společné a dostupné digitální knihovny. Zároveň je Francie jedním z průkopníků vytváření digitálních knihoven v Evropě, takže určitým způsobem udává trendy v tomto oboru.

Ukazuje tak možnou cestu zemím, které s tvorbou těchto kolekcí nemají zkušenosti a to jak legislativní, tak organizační a další. Díky „průkopnictví“

bohatších zemí je tak tato možnost dostupná i pro státy, které nemají dostatečné prostředky k provádění výzkumů a mají možnost čerpat z již dosažených výsledků.

## 5 Závěr

Gallica je jedním z „tradičních“ projektů: národní knihovna digitalizuje svou sbírku a posléze si vybírá užší specializaci, ve které bude pokračovat, v případě BnF se jedná o periodika 18. a 19. století.

Knihovna obsahuje dokumenty z prakticky všech vědních oborů, v přehledu jsou vypsány následující: slovníky, encyklopedi, bibliografie, životopisy; filosofie, religionistika, historie, geografie, etnologie, sociologie a vzdělání, právo, ekonomie, politologie, přírodní vědy a technika, jazyky a lingvistika, literatura, umění a architektura, hudba a divadlo. Vyčerpává tak všechny podstatné oblasti lidské činnosti.

Projekt se strážlivě zaměřuje na levnější odvětví digitalizace – zmíněná periodika. Uznává, že dokud nezíská širší finanční podporu mezi komerčními podniky, nemá dostatek prostředků k náročnějším akcím.

Považuji tento přístup za velmi rozumný – Gallica zpřístupňuje důležité historické dokumenty a další nekomerční zdroje tak, aby se k nim mohl dostat takřka kdokoli. Buduje si tak dobrou reputaci, se kterou je cesta k soukromým investorům jednodušší a rychlejší. Věřím, že pokud bude projekt pokračovat stejným způsobem jako dosud, zdroje rozhodně najde a bude moci pokračovat v ambicióznějších odvětvích digitalizace a opět pomůže posunout tuto disciplínu kupředu.

## Reference

- [1] *Analyse et Traitement Informatique de la Langue Française* [online]. 2012. Dostupné z: <http://www.atilf.fr/>.
- [2] *Charte documentaire de Gallica : 1997 - 2007* [online]. 2007. Dostupné z: [http://www.bnf.fr/fr/professionnels/anx\\_pol\\_num/a.Charte\\_documentaire\\_de\\_Gallica\\_1997\\_2007.html](http://www.bnf.fr/fr/professionnels/anx_pol_num/a.Charte_documentaire_de_Gallica_1997_2007.html).
- [3] *Europeana, Gallica et la Bnue* [online]. 2007. Bulletin des Bibliothèques de France. Dostupné z: <http://bbf.enssib.fr/blog/2007/03/23/europeana-gallica-et-bnue>.
- [4] *Free eBooks by Project Gutenberg* [online]. 2012. Dostupné z: <http://www.gutenberg.org/ebooks/>.
- [5] *International Conference on Asian Digital Libraries* [online]. 2010. Dostupné z: <http://www.icadl.org/programs.htm>.
- [6] *Gallica: A propos* [online]. 2011. Dostupné z: <http://gallica.bnf.fr/html/editorial/propos>.

[7] *Wikisource:Gallica* [online]. 2012. Dostupné z: <http://fr.wikisource.org/wiki/Wikisource:Gallica>.