

Plán správy dat

Administrativní informace

Číslo projektu: LM2023043

Název projektu: Česká literární bibliografie 2023–2026

Datum řešení projektu: 1. 1. 2023 – 31. 12. 2026

Verze DMP: 2.0

Datum zveřejnění DMP: 14. 5. 2025

Licence DMP: [Uved'te původ 4.0 Mezinárodní \(CC BY 4.0\)](#)

Verze	Datum	Odpovědná osoba	Popis změn
Verze 1.0	říjen 2023	Veronika Zemanová	První verze DMP
Verze 2.0	duben 2025	Veronika Zemanová	Rozšíření první verze o aktuální informace ke správě dat v ČLB

Přehled změn

Kapitola	Bližší určení změny proti verzi 1	Odůvodnění revize
1. Popis dat a jejich zpracování	Tabulka 1 rozšířena o další databázi	Vytvoření nové databáze a zveřejnění příslušného datasetu v dubnu 2025
1. Popis dat a jejich zpracování: Zpracování dat – Metodické příručky	Odkaz na nově zpřístupněné příručky	Nově zveřejněné materiály
1. Popis dat a jejich zpracování: Kontinuální a převzatá data	Autorství datasetů	Vysvětlení přiznání autorství jednotlivých verzí datasetů
1. Popis dat a jejich zpracování: Opatření pro zajištění kvality dat	Doplněn popis redakce databáze a aktualizace URL linků	Rozhodnutí podrobněji popsat postup zpracování dat
2. FAIR data: Zpřístupnění dat	V části Metadata doplněna informace o ORCID	ORCID součástí metadat
3. Ukládání a zálohování dat	Vytvoření nové kapitoly Datové záznamy v repozitáři ASEP	Doplněny informace k jednotlivým verzím datasetů
Přílohy	Vytvořena kapitola přílohy	Vytvořen seznam všech datových sad ČLB uložených v repozitáři ASEP

Toto dílo podléhá licenci Creative Commons Uveďte původ 4.0 Mezinárodní License.

Licenční podmínky jsou dostupné zde:

<https://creativecommons.org/licenses/by/4.0/deed.cs>

Obsah

Úvod	5
1. Popis dat a jejich zpracování.....	6
Vytváření a shromažďování dat.....	6
Zpracování dat.....	8
Kontinuální a převzatá data	11
Opatření pro zajištění kvality dat	11
2. FAIR data	13
Zajištění dohledatelnosti dat	13
Zpřístupnění dat	13
Interoperabilita dat	14
Znovupoužití dat.....	15
3. Ukládání a zálohování dat	16
Datové záznamy v repozitáři ASEP	17
4. Etické a právní otázky.....	18
5. Další výstupy ČLB.....	19
6. Odpovědnost za správu dat.....	20
Seznam tabulek	21
Přílohy	22

Úvod

Tento plán správy dat (Data Management Plan – DMP) byl vytvořen pro potřeby činnosti výzkumné infrastruktury Česká literární bibliografie (ČLB), především pro potřeby projektu LM2023043, který je řešen v letech 2023–2026. Východiskem činnosti ČLB je zpracovávání a průběžné doplňování analytické bibliografie české literatury, která souvisle mapuje reflexi české a výběrově též světové literatury a kultury v periodickém tisku a knižní produkci vydávaných na území českých zemí ve všech historických komunikačních jazycích a dále exilovou produkci po roce 1948. Databáze ČLB chronologicky pokrývají období od poslední třetiny 18. století až po současnost. **ČLB produkuje data dlouhodobě a kontinuálně. Data ČLB nejsou statická (neměnná), ale jsou průběžně aktualizována a doplňována o nově zjištěné skutečnosti a další údaje** (linky na plné texty, perzistentní identifikátory atp.). Pro tuto skutečnost dochází v ČLB k pravidelnému verzování datových sad, které jsou archivovány a sdíleny se zájemci o jejich další využívání.

ČLB považuje DMP za důležitou součást správné výzkumné a dokumentační praxe. Předkládaný DMP popisuje správu dat v ČLB a vysvětluje, jak se ČLB vypořádává s požadavky na otevřený přístup. Cílem DMP je nejen poskytnout informace o životním cyklu dat, která jsou generována, shromažďována, zpracovávána, uchovávána i sdílena v rámci ČLB, ale také zajistit, aby byla data zpřístupněna v souladu s FAIR principy¹. DMP má přispět ke zvýšení viditelnosti dat a jejich snadnějšímu využití třetími stranami. DMP je živý dokument, který je dle potřeby aktualizován.

¹ FAIR principy vysvětlují, jak by data měla být zpracována, aby byla lépe dohledatelná (Findable), dostupná (Accessible), interoperabilní (Interoperable) a opětovně využitelná (Reusable).

1. Popis dat a jejich zpracování

Vytváření a shromažďování dat

ČLB shromažďuje, zpracovává, uchovává a zpřístupňuje metadata, tj. strukturovaný popis tištěných i digitálních dokumentů, které mapují literární dění a literární vědu v českých zemích či cirkulaci české literatury v zahraničí od konce 18. století až do současnosti bez ohledu na jazyk, formát či médium, a to zejména v podobě bibliografických, popř. dalších znalostníchází: taková metadata jsou považována za skutečná data, která jsou v ČLB spravována. Tato data jsou vytvářena prostřednictvím manuální excerptce, která probíhá jako zaznamenávání excerpt z primárních i sekundárních literárních zdrojů. Data jsou vytvářena pracovníky ČLB především v softwarech Aleph² a Koha³, popř. historicky i v proprietárních systémech vlastních (Retrobi), ve kterých jsou i dále spravována. Z těchto softwarů jsou data pravidelně exportována v otevřených formátech a ukládána do institucionálního repozitáře ASEP⁴, kde jsou dlouhodobě archivována a zároveň k dispozici ke stažení.⁵

ČLB dosud shromáždila více než 2,3 mil. záznamů a v současnosti činí roční přírůstek asi 20 tisíc záznamů.

² Aleph je elektronický knihovní systém firmy Ex Libris.

³ Koha je open source informační knihovní systém.

⁴ ASEP je institucionální repozitář Akademie věd České republiky, který slouží k evidenci všech relevantních výstupů pracovišť AV ČR.

⁵ Do repozitáře ASEP jsou od října 2023 ukládány datové sady všech jednotlivých databází uvedených v tab. č. 1 s výjimkou databáze České literární osobnosti, které budou z důvodu přípravy a čištění dat vloženy do repozitáře později v průběhu trvání projektu.

Typ dat	Název databáze	Způsob získání dat	Zpracování dat – vstupy	Zpracování dat – výstupy	
Bibliografická a biografická data	Retrospektivní bibliografie české literatury 1770–1945	Digitalizované excerpční listky	RETROBI	RETROBI	VUFIND
		(Polo)automaticky upravené databázové záznamy	RETROBI	RETROBI	
		Ručně přepsané databázové záznamy	ALEPH	ALEPH	
	ALKARO – Almanachy, kalendáře, ročenky v českých zemích 1801–1945	Manuální excerpce veřejných zdrojů	ALEPH	ALEPH	VUFIND
	Současná bibliografie (od roku 1945)	Manuální excerpce veřejných zdrojů	ALEPH	ALEPH	VUFIND
	Bibliografie českého literárního internetu				
	Bibliografie českého literárního samizdatu				
	Bibliografie českého literárního exilu (1948–1989)				
	Český literární exil 1948-1989				
	Bohemistické konsorcium	Manuální excerpce veřejných zdrojů			
		Přebírání a úprava dříve vytvořených záznamů v ČLB			
	Česká literatura v překladu	Vytváří odborník na dané téma + manuální excerpce veřejných zdrojů	Excel soubor	VUFIND	
	České literární osobnosti	Manuální excerpce veřejných zdrojů a dotazníky	Koha		

Tabulka 1: Zpracování dat

V tabulce č. 1 jsou uvedeny způsoby získávání dat a systémy, ve kterých jsou bibliografická i biografická data zpracovávána a prezentována dle jednotlivých databází. Z hlediska struktury je ČLB rozdělena na 10 databází, současně ale platí, že některé databáze jsou součástí většího celku (např. Současná bibliografie obsahuje Bibliografii současného českého literárního internetu, samizdatu, exilu a Bohemistické konsorcium, také se plánuje v budoucnu zahrnout ALKARO pod Retrospektivní bibliografii české literatury). V další tabulce č. 2 je přehled všech formátů, ve kterých je možné tato data v ČLB spravovat. Velikost dat jednotlivých databází se pohybuje v desítkách až v nižších stovkách megabytů v závislosti na daném formátu. Softwarová data a jejich formáty jsou popsány v kap. 5.

Cílová skupina	Systém	Formát dat
Veřejně dostupné formáty	VUFIND	JSON
		MARCXML
		MARC21
		DOCX
		PDF
	ALEPH OPAC	SAV
		END
Formáty pro interní potřebu	ALEPH klient	MARC21
		MARCXML
		ALEPH sekvenční
		TXT
	Interní skripty (python)	CSV
		JSON
		XML
	KOHA	MARC
		XML
	EXCEL	XLSX

Tabulka 2: Formáty bibliografických a biografických dat v jednotlivých systémech

Zpracování dat

ČLB zpracovává soubor deseti samostatných databází, které jsou sjednoceny na společný metadatový standard MARC21 (bibliografické a autoritní záznamy)⁶. Datová struktura databází je harmonizována včetně systémů věcného popisu (soubor národních autorit a jeho lokální doplňky, MDT⁷, skupina konspektu) a doplněna o perzistentní identifikátory (identifikační číslo autority, ISBN⁸, ISSN⁹, DOI¹⁰, číslo ČNB¹¹, lokální identifikátory ČLB¹²).

⁶ MARC21 je mezinárodně uznávaný bibliografický výměnný formát.

⁷ MDT – Mezinárodní desetinné třídění je univerzální klasifikační jazyk, který kombinuje třídění hierarchického a fasetového typu.

⁸ ISBN je mezinárodní identifikátor knih a neperiodických publikací.

Databáze jsou technologicky vnitřně propojeny a zastřešeny jednotným rešeršním systémem na platformě softwaru VuFind¹³, který slouží jako primární přístupový bod i pro koncové uživatele. Standardizovaný mezinárodní výměnný formát a sjednocená katalogizace umožňuje technologické propojení dat ČLB s dalšími systémy výměny vědeckých informací.

Knihovnické standardy

Data jsou v co největší možné míře zpracovávána v souladu s těmito standardy a normami:

- [Katalogizační formát MARC21](#) (pro bibliografické i autoritní záznamy)
 - [Katalogizace podle RDA ve formátu MARC21](#)¹⁴
- Pravidla mezinárodního standardizovaného bibliografického popisu ISBD¹⁵
- Standardy jmenného a věcného popisu - [Národní autority ČR](#)

Výše uvedené standardy jsou základním východiskem pro zpracování jakýchkoli dat v moderních knihovnických databázových systémech a jejich dodržování výrazně usnadňuje vzájemné sdílení dat mezi jednotlivými institucemi. Zároveň je potřeba zdůraznit, že primárním cílem ČLB je vytváření komplexního bibliografického popisu, který obsahuje veškeré dostupné informace o daném dokumentu, tak, aby sloužil potřebám vědecké komunity včetně zájemců o kvantitativní výzkum bibliografických dat. Z tohoto důvodu záznamy ČLB obvykle obsahují informace nad rámec běžné knihovní katalogizační praxe (týká se zejména údajů o odpovědnosti a míry podrobnosti věcného popisu; všechny záznamy jsou důsledně anotovány) a způsob zápisu je zejména s ohledem na možnosti vyhledávání více formalizován. Protože ČLB zpracovává typy záznamů, pro něž na národní úrovni neexistují uspokojivé katalogizační manuály (články) nebo pravidla nejsou k dispozici vůbec (analytický rozpis knih, zpracování online dokumentů, samizdatových a exilových periodik, audiovizuálních materiálů, retrospektivní excerpce nepravých periodik atp.), vytváří ČLB nezbytné metodické materiály ve vlastní režii formou interních katalogizačních manuálů či veřejně dostupných certifikovaných metodik.

⁹ ISSN je mezinárodní identifikátor periodických publikací a jiných tzv. pokračujících zdrojů.

¹⁰ DOI je mezinárodní identifikátor digitálního objektu v prostředí internetu.

¹¹ Číslo ČNB je český národní identifikátor bohémikální produkce od 19. století.

¹² ČLB vytváří lokální soubory perzistentních identifikátorů pro ty hodnoty, které nejsou v obecných souborech zastoupeny.

¹³ VuFind je otevřený knihovní vyhledávací systém.

¹⁴ Praxe ČLB vychází z národní interpretace katalogizačních pravidel RDA. Při neexistenci pravidel pro popis článků na národní úrovni využívá ČLB nad rámec stávajících pravidel interpretace vlastní, popsané v manuálech pro zpracování dat uvedených níže.

¹⁵ ISBD International Standard Bibliographic Description. Consolidated ed. Berlin: De Gruyter Saur, 2011.

Metodické příručky

- MACEK, Emanuel: *Bibliografie české beletrie a literární vědy. Metodická příručka*. Ústav pro českou literaturu ČSAV, 1969.
- [MIKA, Jiří]: *B97. Databáze české literární bibliografie od roku 1997*. [Interní dokument] Praha: ÚČL AV ČR 1996.
- MALÍNEK, Vojtěch. *Oborová analytická bibliografie. Metodika zpracování*. Ústav pro českou literaturu AV ČR, 2020. ISBN 978-80-7658-008-4. Dostupné z: <https://kramerus.lib.cas.cz/uuid/uuid:2679143f-a41f-443d-be8a-34f7be85fc91>.
- *Metodika tvorby a kontroly jmenných autorit podle RDA*. Dostupné z: <https://autority.nkp.cz/jmenne-autorit/metodicke-materialy/metodika-jmena-cvicne-2>.
- *MARC 21. Pracovní manuál pro potřeby zpracování České literární bibliografie*. [Interní elektronický dokument]. Verze 6.0. Česká literární bibliografie, 2023.
- *MARC 21. Pracovní manuál pro potřeby katalogizace jmenných autorit v databázi v České literární osobnosti*. [Interní elektronický dokument]. Verze 3.0. Česká literární bibliografie, 2023.
- ŘEHÁK, Daniel. *Metodika zjišťování autorů u šifer a pseudonymů v periodickém tisku*. Ústav pro českou literaturu AV ČR, 2015. Certifikovaná metodika MK-S 13094/2015 OVV. Dostupné z: <http://hdl.handle.net/11104/0253681>.

První dvě uvedené práce byly v minulosti pořízeny za účelem standardizace zpracování oborové článkové bibliografie v Ústavu pro českou literaturu ČSAV/AV ČR¹⁶. Mackova příručka primárně řeší zpracování článkové bibliografických záznamů pro Retrospektivní bibliografii české literatury, vzhledem k době svého vzniku ještě v podobě lístkové kartotéky. Manuál Jiřího Miky pro excerpci v období 1997–2012 byl prvním metodickým materiálem pro zpracování oborové literárněvědné článkové bibliografie v databázové podobě, platným pro původní databázový systém ISIS. Obě příručky jsou dostupné v repozitáři ASEP viz kap. 3 Ukládání a zálohování dat. Malínková metodika zpracování oborové bibliografie vznikla v souvislosti s řešením projektu NAKI a popisuje zpracování oborové analytické bibliografie ve formátu MARC21 s důrazem na potřeby literárněvědné bibliografie.

Zpracování biografických dat pro potřeby databáze České literární osobnosti vychází z původních norem pro zpracování proprietární biografické databáze, které byly při konverzi do formátu MARC21 doplněny o zásady Metodiky tvorby a kontroly jmenných autorit podle RDA, vypracované v Oddělení národních jmenných autorit NK ČR (tzv. příručka ONJA). Tato pravidla byla dále doplněna a adaptována pro potřeby biografického výzkumu v oboru literární vědy. Aktuální pravidla popisu jsou fixována v podobě interního manuálu.

¹⁶ ČSAV/AV ČR – Československá akademie věd / Akademie věd ČR

Kontinuální a převzatá data

Tento DMP je vytvořen pro potřeby projektu Česká literární bibliografie, který je realizován mezi roky 2023–2026. Data, která do tohoto projektu vstoupila, byla vytvořena v ČLB v předchozích letech jejího fungování. Označujeme je za tzv. kontinuální data, neboť jejich generování probíhá nepřetržitě od počátků vzniku ČLB a v každém dalším období s nimi ČLB dále pracuje.

Autorství datasetů

Autorství by mělo odrážet skutečný přínos ke konkrétnímu dílu – v našem případě ke každé nové verzi datové sady. Z toho důvodu ty autory, kteří přispěli pouze k předchozím verzím, ale nikoli k nové verzi, automaticky nezahrnujeme jako autory nové verze. Autorství je tak navázáno na konkrétní verzi dat. Jelikož každá nová verze dat navazuje na předchozí verze, které se stávají součástí novějších verzí, jsou autoři předchozích verzí zahrnuti v metadatech datasetu v souboru README.

Kromě dat získaných vlastní excerpcí zpracovává ČLB i data, která přebírá od spolupracujících institucí (např. otevřené sady dat poskytnuté Národní knihovnou v Praze). V současnosti ČLB využívá také následující sady dat ve formátech MARC21:

- [datová sada Souborný katalog ČR](#), zveřejněna pod licencí CCo.

Pro experimentální účely v rámci vývoje portálu Literarybibliography.eu (viz kap. 5) jsou dále využívány:

- [datová sada Polské literární bibliografie](#), zveřejněna pod licencí CC BY,
- [datová sada Fennica](#) Finské národní knihovny, zveřejněna pod licencí CCo,
- [datová sada Arto](#) Finské národní knihovny, zveřejněna pod licencí CCo,
- [datová sada Španělské národní knihovny](#), zveřejněna pod licencí CCo.

Datové sady Národních knihoven Finska a Španělska byly po úpravách pro potřeby portálu Literarybibliography.eu zveřejněny v repozitáři ASEP 8. 11. 2024. Přímé odkazy jsou uvedeny v kap. 3 Ukládání a zálohování dat.

Opatření pro zajištění kvality dat

Data management ČLB odpovídá standardům oboru. Kvality dat ČLB je dosaženo zavedením systematického třífázového procesu kontroly dat včetně využití automatizovaných kontrolních skriptů. Záznamy jsou zároveň průběžně redigovány a harmonizovány na úrovni rejstříků jednotlivých polí tak, aby byly odstraněny rozdíly dané historicky (užívání odlišných pravidel popisu) i odchylkami individuálních rukopisů jednotlivých zpracovatelů.

ČLB používá pokročilý systém kontroly zpracovávaných dat. Ta probíhá na několika stupních. Datová struktura (interpunkce, formální správnost zápisu atp.) je kontrolována (a) základní automatickou kontrolou již v momentě uložení záznamu, (b) každodenně dávkově detailním kontrolním skriptem, který sleduje všechny nově založené či nově upravené záznamy, (c) tato kontrola se pak opakuje ještě jednou za týden, (d) minimálně jednou do roka pak pomocí detailních jednorázových testů. Paralelně probíhá věcná kontrola, při níž (a) každý pracovník kontroluje své vlastní záznamy vytvořené za poslední měsíc, (b) následně tytéž záznamy poté prochází jiný, náhodně zvolený zaměstnanec, (c) na závěr záznamy prověřuje redaktor báze.

Redakce databáze probíhá zejména dle rejstříků jednotlivých databázových polí. Údaje v nich jsou dle potřeb automatizovaně či manuálně opravovány, aktualizovány a harmonizovány tak, aby bylo dosaženo jednotného zápisu napříč celou databází. Systematicky je přitom ověřována zejména míra propojení záznamů na autoritní báze a perzistentní identifikátory.

Podrobný postup redakce databáze

Součástí redakce a zkvalitňování databáze je i revize odkazů odkazujících na plný text excerpovaného článku (zejména u e-zdrojů), potažmo jejich zpětné doplnění – manuálně v případě složitější strukturovaných webů či hromadně u volně dostupných časopisů v prostředí digitálních knihoven Kramerius. Katalogizační praxe je v tomto ohledu taková, že pokud je v momentu excerpce text dostupný v online verzi, odkaz na něj doplňujeme do záznamu hned. V případě existence archivní verze textu je současně do záznamu vkládán i odkaz vedoucí do Digitální knihovny českých webových zdrojů Webarchiv, zpravidla vedoucí na jeho první zarchivovanou verzi. Pokud je text obohacen o identifikátor DOI, považujeme ho za plnohodnotnou alternativu zajišťující dlouhodobý přístup k plnému textu a odkaz do Webarchivu již do záznamu nevyplňujeme.

Jednou ročně se všechny odkazy kontrolují, zda jsou stále funkční, tj. zda daný web stále existuje, zda se nepřejmenoval apod. Zároveň se vytváří seznam záznamů, které neobsahují odkaz do Webarchivu. Zásadní je tato kontrola zejména u e-zdrojů, u tištěných periodik bude v případě zániku webu stále dostupná, resp. dohledatelná verze tištěná.

Následně u nefunkčních odkazů dochází dle typu chyby k opravě. Link se buď nahradí funkčním odkazem či v případě zániku webu spolu s jeho obsahem je odkaz ze záznamu smazán a zůstává v něm tak pouze funkční link vedoucí na archivní verzi, která je buď uživateli dostupná přímo z jeho počítače nebo z Referenčního centra Národní knihovny v Praze (v souladu s autorskoprávní ochranou). Záznamy, které odkaz do Webarchivu neobsahují, jsou rozděleny do dvou skupin. V první jsou záznamy, u kterých byl v mezidobě web archivován a link je tak možné do záznamu zpětně hromadnou opravou doplnit. Ve druhé skupině jsou odkazy, které buď archivací neprošly nebo jsou dostupné pouze z Referenčního centra NK, tudíž nelze jednoduše ověřit, zda jsou skutečně zazálohovány. Tato skupina záznamů je zasílána vedoucímu pracovníkovi do Webarchivu s prosbou o přidání do jejich příští sklizně k archivaci. Tímto postupem je zajištěno, že do příští kontroly ne/funkčnosti odkazů budou tyto odkazy již archivované a bude je tak možno do záznamů

doplnit. Tímto postupem by mělo být zajištěno, že by nemělo dojít ke ztrátě obsahu, na který je v záznamu odkazováno.

2. FAIR data

Jednou z priorit ČLB je, aby výzkumná data splňovala FAIR principy, tj. že data mají být dohledatelná, dostupná, interoperabilní a opětovně použitelná.

Zajištění dohledatelnosti dat

Zpřístupňování dat a výsledků je jedním z hlavních cílů výzkumné infrastruktury. Data jsou k dispozici prostřednictvím široké škály kanálů v otevřených a standardních formátech. Primárně jsou data dostupná ve webovém vyhledávacím systému VuFind, dále také jako tabulky a grafy exportovatelné ze Statistického a analytického modulu (SAM) a prostřednictvím národních kolaborativních sítí, jako jsou Knihovny.cz¹⁷ nebo Souborný katalog Akademie věd ČR. Pro zájemce o výpočetní zpracování dat jsou data ČLB dostupná v institucionálním repozitáři ASEP ve standardizovaném formátu MARC21, který je obvyklý pro obor bibliografie a knihovnické vědy. ČLB aktivně umožňuje propojení s jinými datovými portály a uvolňuje svá data pro potřeby kooperativních sítí výměny vědeckých informací.

Jednotlivým datovým sadám z každé databáze jsou přiděleny mezinárodní identifikátory handle a DOI. Datové sady jsou opatřeny metadaty, jejichž struktura vychází z formátu UNIMARC. Indexování a vyhledatelnost metadat zajišťuje provozovatel repozitáře, ve kterém jsou data uložena, zároveň jsou dostupná na webu ČLB.

Zpřístupnění dat

Obecným principem šíření dat v ČLB je otevřený přístup. V praxi to znamená, že všechna výzkumná data a metadata, u kterých to právní předpisy umožňují, jsou zpřístupněna pod licencí Creative Commons. Osobní a citlivé údaje žijících osob jsou zveřejněny pouze pokud byla veřejně publikována nebo na základě souhlasu dotyčné osoby. Na základě přání dotyčné osoby může být část údajů pouze archivována a zpřístupněna až po naplnění smluvných podmínek.

Datové sady jako celek jsou zpřístupněny následujícími způsoby:

- 1) prostřednictvím institucionálního repozitáře ASEP,
- 2) prostřednictvím webových stránek a služeb ČLB,
 - a) centrální přístupový bod VuFind ČLB,
 - b) proprietární rozhraní jednotlivých databází (OPAC Aleph, RETROBI. CLO),
 - c) portál Literarybibliography.eu,

¹⁷ Portál Knihovny.cz poskytuje jednotný přístup ke službám českých a moravských knihoven.

- d) exportní modul SAM,
- 3) kooperační sítě a souborné portály (Knihovny.cz, Souborný katalog AV ČR, platforma INDIHU.cz).

Repozitář

Data jsou uložena v institucionálním datovém repozitáři ASEP: <https://asep-portal.lib.cas.cz/>, který je registrován v registru datových repozitářů [RE3DATA](#). Takto uložená data jsou zveřejněna v on-line katalogu a opatřena persistentními identifikátory handle a DOI. Popis jednotlivých datových záznamů vychází ze standardu MARC21. Metadata všech záznamů jsou volně přístupná.

Data

Datasety ČLB jsou pravidelně zveřejňovány 2x ročně v režimu Open Access (v dubnu a říjnu) v repozitáři ASEP. Takto veřejně dostupná data neobsahují žádné osobní, utajované či citlivé údaje. Údaje s omezeným přístupem jsou uvolňovány v souladu s příslušnými právními předpisy (viz kap. Etické a právní otázky).

Metadata

Základní metadata k datovým sadám jsou zveřejněna na webu ČLB na samostatných stránkách jednotlivých databází. Rozšířená podoba metadat k jednotlivým datasetům jsou součástí jejich metadatového záznamu v repozitáři ASEP. Struktura těchto metadat vychází z formátu UNIMARC. S Knihovnou Akademie věd ČR ČLB spolupracuje v otázkách potřeb datového repozitáře a předpokládá se dlouhodobá dostupnost metadat i dat samotných. Součástí metadat v repozitáři ASEP je pole autor, které je propojeno s ORCID autora.

Interoperabilita dat

ČLB si klade za cíl shromažďovat a dokumentovat data standardizovaným způsobem. Interoperabilita je zajištěna tím, že se data zpracovávají v metadatovém standardu MARC21. Data lze zpracovávat strojově a propojovat s dalšími existujícími datasety. ČLB přitom aktivně umožňuje propojení s jinými datovými portály. Při zpracovávání dat se ČLB řídí standardizovanými tezaury, které zahrnují jmenné (soubory personálních a korporativních autorit, autoritní data o akcích a dílech) a věcné (soubory tematických, geografických, chronologických a formálních autorit) autority, které jsou podporovány odbornou knihovnickou a bibliografickou komunitou. Díky vysokému stupni standardizace a napojení na stávající systémy perzistentních identifikátorů a také implementaci nástrojů pro sdílení a výměnu dat je velmi snadné propojit data ČLB s jinými.

Znovupoužití dat

Data jsou uživatelům nabídnuta již v průběhu činnosti výzkumné infrastruktury ČLB. Veškerá zveřejněná data jsou k dispozici v režimu Open Access a opatřeny licencemi Creative Commons, konkrétně CC BY-NC-SA 4.0, které umožňují opětovné použití v souladu s požadavky otevřené vědy.

CC BY-NC-SA 4.0 definují rozsah pravomocí při nakládání s uvedenými databázemi České literární bibliografie (dále jako Dílo) následovně:

Uživatel je oprávněn:

- Dílo sdílet = vytěžovat, zužitkovávat, rozmnožovat a sdílet celý obsah databáze nebo podstatnou část obsahu databáze.
- Dílo upravit = pozměňovat nebo na něj navazovat.

Za těchto podmínek:

- Uvádět původ – Uživatel má povinnost uvést původ Díla předem stanoveným způsobem. Citace musí obsahovat informace o názvu a pořizovateli Díla, zdroji dat a odkaz na plné znění veřejných licenčních podmínek CC BY-NC-SA 4.0 Mezinárodní.
- Neužívat Dílo komerčně – Je zakázáno Dílo užívat pro komerční účely.
- Zachovávat licenci – Bude-li uživatel toto Dílo upravovat, pozměňovat nebo na něj navazovat, musí svá odvozená Díla vystavovat pod stejnou licenci jako původní Dílo.

Užití databází je podmíněno jejich náležitým citováním a uváděním informace o využití ČLB u výstupů výzkumu a vývoje podle doporučeného vzoru:

Při vzniku práce [knihy/studie/...] byly využity zdroje výzkumné infrastruktury Česká literární bibliografie – <https://clb.ucl.cas.cz/> (kód ORJ: 90243).

3. Ukládání a zálohování dat

Veškerá data jsou zálohována přímo při Ústavu pro českou literaturu (ÚČL) minimálně ve dvou kopiích fyzicky uložených na různých místech. Hlavní část dat ČLB (bibliografická báze) je zpracovávána v systému Aleph, provozovaném Knihovnou AV ČR. Tato data jsou zálohována Knihovnou AV v pěti kopiích na různých místech ČR. Data ČLB jsou pak vystavena též v dalších sdílených katalozích (např. Knihovny.cz) s průběžnou aktualizací a s vlastním systémem zálohování. Prakticky tedy nehrozí možnost, že by na datech ČLB mohla vzniknout větší škoda či že by došlo k nějaké ztrátě.

Správa dat ČLB a IT provoz a vývoj probíhá primárně v rámci sekce digitalizace a IT při ČLB a ve spolupráci se Střediskem výpočetní techniky Ústavu pro českou literaturu (SVT ÚČL). Všechny síťové aplikace jsou provozovány na virtualizovaných serverech umístěných v zabezpečeném datacentru ve výhradním vlastnictví ÚČL, v režii a pod přímou správou zaměstnanců SVT ÚČL. Kapacita úložišť je plánována tak, aby byla zajištěna dostatečná rezerva (alespoň 30 %) po plánovanou dobu životnosti a je pravidelně obměňována a navyšována ve dvouletých cyklech. Data i aplikační servery jsou zálohovány v souladu s bezpečnostními standardy primárně na vlastních úložištích a několikaúrovňově. Zároveň je udržována jejich pohotovostní kopie v několika historických verzích v budově a archivní verze ukládaná v šifrované podobě na úložišti CESNET (e-INFRA CZ) mimo budovu ÚČL.

Dlouhodobá archivace dat je zajištěna i jejich uložením v institucionálním repozitáři ASEP provozovaném Knihovnou AV ČR (viz kap. 2 Zpřístupnění dat).

Udržitelnost dat ČLB je momentálně dána jejich velikostí a implementací mezinárodních standardizovaných formátů. Udržitelnost dat je pak značně posílena i přechodem na mezinárodní standard MARC21, který je možné v případě potřeby snadno konvertovat prakticky do libovolného jiného formátu. Pro tyto účely je k dispozici řada specializovaných volně dostupných nástrojů.

Datové záznamy v repozitáři ASEP

Seznam datových sad je součástí přílohy tohoto dokumentu.

Verze 1

Do datového repozitáře ASEP bylo 26. 10. 2023 nahráno 9 datových sad.

Verze 2

22. 4. 2024 proběhl hromadný import dat. Do datového repozitáře ASEP bylo 22. 5. 2024 nahráno 8 datových sad (nebyl nahrán dataset Databáze překladu, protože v něm nedošlo od první verze k žádné změně)

Verze 3

8. 11. 2024 bylo zveřejněno 8 aktualizovaných datových sad (nebyl nahrán dataset Databáze překladu, protože v něm nedošlo od první verze k žádné změně)

Při kontrole předchozích sad bylo zjištěno, že dataset Současné bibliografie verze 1 obsahuje konkrétně 20 185 duplikací. Rovněž verze 2 stejného datasetu obsahuje duplicity. Tyto chyby se pak promítly i do příslušných verzí datasetu celé ČLB.

Od třetí verze obsahují datasety soubory ve formátu XLSX (seznamy excerpovaných zdrojů).

Ve stejný termín byly zveřejněny dvě datové sady zpracované při projektu Literarybibliography.eu:

Národní knihovna Španělska. Národní knihovna Španělska – září 2024.	https://doi.org/10.57680/asep.0598584
Národní knihovna Finska. Národní knihovna Finska – září 2024.	https://doi.org/10.57680/asep.0598591

Verze 4

Na konci dubna 2025 bylo v rámci 4. verzí datasetů zveřejněno 9 aktualizovaných datasetů a jeden nový. Všechny datasety byly pro vložení do repozitáře exportovány 16. 4. 2025. S touto dávkou dat byl v repozitáři zveřejněn i dataset příruček a pomocných materiálů pro zpracování dat v systému ISIS.

Česká literární bibliografie v systému ISIS – interní dobové příručky a pomocné materiály.	https://doi.org/10.57680/asep.0584645
--	---

4. Etické a právní otázky

V rámci ČLB se zpracovávají a obohacují zejména veřejná data. Přesto část dat patří do kategorie osobních a citlivých dat (zejména obsah databáze CLO), ke kterým je potřeba přistupovat v souladu s nařízením GDPR. ČLB nezpřístupňuje data, u kterých by hrozilo, že jejich zveřejněním dojde k nepřiměřenému zásahu do práva na ochranu soukromí a osobních údajů. Ochrana osobních údajů je ústředním etickým požadavkem a pracovníci mají povinnost poskytovat informace o tom, jak se s těmito údaji nakládá. Na webových stránkách ČLB jsou zveřejněny [Informace o ochraně a zpracování osobních údajů](#), které obsahují přehled zpracovávaných údajů, účel a formu jejich zpracování, ale i způsob uložení, přístupnost osobních údajů, způsob jejich zabezpečení a jiné důležité informace.

Přístup pro zadavatele dat je povolen pouze z vymezených IP adres a zabezpečen uživatelským účtem pracovníků ČLB. ČLB je z hlediska zákona pracovištěm oprávněným zpracovávat osobní údaje – příslušné osvědčení vystavil Úřad na ochranu osobních údajů. Osobní data jsou zveřejňována pouze v souladu s právními normami (zákon na ochranu osobních údajů, GDPR). Osobní údaje se nezveřejňují, (a) pokud s tím dotyčná osoba nesouhlasí, (b) pokud dotyčné údaje nebyly zveřejněny jinde a ČLB nemá výslovný souhlas dotyčné osoby k jejich zveřejnění. Veškerý sběr informací, jejich zpracování i uchovávání se striktně řídí platnou legislativou se všemi ohledy na ochranu osobních údajů žijících osob (zákon na ochranu osobních údajů, GDPR).

5. Další výstupy ČLB

ČLB je aktivní též v oblasti vývoje softwaru a IT nástrojů. Kromě softwaru pro digitalizaci a online prezentaci lístkových kartoték RETROBI jde zejména o adaptaci opensourcového discovery systému VuFind, který uživatelé ČLB mohou od roku 2020 využívat jako hlavní přístupový bod k datům. Pro společné prohledávání jsou v něm přístupné všechny bibliografické báze ČLB a v únoru 2025 došlo k zapojení i biografické databáze České literární osobnosti. Jako samostatný modul pro VuFind je při ČLB vyvíjen statistický a analytický modul tzv. SAM, který umožňuje vizualizaci a statistickou analýzu bibliografických dat ČLB.

Na vývoji softwaru a IT nástrojů ČLB intenzivně kooperuje se svým nejbližším zahraničním partnerem, pracovištěm Polské literární bibliografie. Obě národní infrastruktury společně budují portál [Literarybibliography.eu](https://literarybibliography.eu) zpřístupňující bibliografická data o jednotlivých národních literaturách. Tímto způsobem je prozatím k výběru databáze české nebo polské literární bibliografie, Finské národní knihovny a Španělské národní knihovny.

ČLB vyvíjí software pod volnou licencí a dokumentaci uveřejňuje na GitHubu.

Typ dat	Popis dat	Formáty	Dostupnost
Skripty	Skripty pro správu, kontrolu a čištění a opětovné využití dat	.py, ipynb	https://github.com/CzechLitBib/UCL
Software	SAM	php	https://github.com/uclavcr/sam
	RETROBI	java	https://github.com/uclavcr/retrobi

Tabulka 3: Další výstupy ČLB

Také ve všech výstupech publikační činnosti ČLB jsou zohledňovány principy otevřené vědy. V rámci edice Bibliographica vychází specializované knižní bibliografie, jež jsou archivovány v institucionálním repozitáři ASEP a v Digitální knihovně AV ČR, kde jsou po uplynutí tříletého časového embarga zpřístupněny uživatelům. Publikacím je od r. 2022 přidělován identifikátor DOI. Všechny ostatní typy dokumentů publikují pracovníci ČLB v souladu se Strategií otevřeného přístupu k vědeckým informacím pro roky 2022–2027¹⁸. ČLB doporučuje svým pracovníkům zvážit možnost publikovat v režimu Open Access a zajišťuje podporu autorům, kteří chtějí archivovat podkladová data ke svým dílům.

¹⁸ Interní dokument ÚČL.

6.Odpovědnost za správu dat

Producentem dat je výzkumná infrastruktura Česká literární bibliografie, která je provozována při Ústavu pro českou literaturu AV ČR, v. v. i., který je vlastníkem všech dat.

Každý člen výzkumné infrastruktury je zodpovědný za správné zpracování dat, která ČLB shromažďuje, tím, že postupuje v souladu se zavedenými standardy a pracovními postupy.

Od roku 2020 má ČLB koordinátorku otevřeného přístupu, která má otevřený přístup a související agendy jako specifický cíl své práce. V průběhu sledovaného období byla na úrovni ÚČL zpracována Strategie otevřeného přístupu, která v relevantních bodech řeší i aktivity ČLB.

Seznam tabulek

Tabulka 1: Zpracování dat	7
Tabulka 2: Formáty bibliografických a biografických dat v jednotlivých systémech	8
Tabulka 3: Další výstupy ČLB	19

Přílohy

	Verze 1	
1.	Česká literární bibliografie	https://doi.org/10.57680/asep.0576768
2.	Retrospektivní bibliografie české literatury 1770–1945	https://doi.org/10.57680/asep.0576760
3.	ALKARO – Almanachy, kalendáře, ročenky v českých zemích 1801–1945	https://doi.org/10.57680/asep.0576764
4.	Současná bibliografie (od roku 1945)	https://doi.org/10.57680/asep.0576761
5.	Bibliografie českého literárního internetu	https://doi.org/10.57680/asep.0576762
6.	Bibliografie českého literárního samizdatu	https://doi.org/10.57680/asep.0586031
7.	Bibliografie českého literárního exilu (1948–1989)	https://doi.org/10.57680/asep.0576766
8.	Český literární exil 1948–1989	https://doi.org/10.57680/asep.0576767
9.	Česká literatura v překladu	https://doi.org/10.57680/asep.0576539
	Verze 2	
10.	Česká literární bibliografie	https://doi.org/10.57680/asep.0586041
11.	Retrospektivní bibliografie české literatury 1770–1945	https://doi.org/10.57680/asep.0586015
12.	ALKARO – Almanachy, kalendáře, ročenky v českých zemích 1801–1945	https://doi.org/10.57680/asep.0586033
13.	Současná bibliografie (od roku 1945)	https://doi.org/10.57680/asep.0586027
14.	Bibliografie českého literárního internetu	https://doi.org/10.57680/asep.0586029
15.	Bibliografie českého literárního samizdatu	https://doi.org/10.57680/asep.0586031
16.	Bibliografie českého literárního exilu (1948–1989)	https://doi.org/10.57680/asep.0586125
17.	Český literární exil 1948–1989	https://doi.org/10.57680/asep.0586038
	Verze 3	
18.	Česká literární bibliografie	https://doi.org/10.57680/asep.0600376
19.	Retrospektivní bibliografie české literatury 1770–1945	https://doi.org/10.57680/asep.0600377
20.	ALKARO – Almanachy, kalendáře, ročenky v českých zemích 1801–1945	https://doi.org/10.57680/asep.0600378
21.	Současná bibliografie (od roku 1945)	https://doi.org/10.57680/asep.0600381
22.	Bibliografie českého literárního internetu	https://doi.org/10.57680/asep.0600382
23.	Bibliografie českého literárního samizdatu	https://doi.org/10.57680/asep.0600383
24.	Bibliografie českého literárního exilu (1948–1989)	https://doi.org/10.57680/asep.0600384
25.	Český literární exil 1948–1989	https://doi.org/10.57680/asep.0600385
26.	Národní knihovna Španělska. Národní knihovna Španělska – září 2024.	https://doi.org/10.57680/asep.0598584
27.	Národní knihovna Finska. Národní knihovna	https://doi.org/10.57680/asep.0598591

	Finska – září 2024.	
	Verze 4	
28.	Česká literární bibliografie	https://doi.org/10.57680/asep.0619242
29.	Retrospektivní bibliografie české literatury 1770–1945	https://doi.org/10.57680/asep.0619241
30.	ALKARO – Almanachy, kalendáře, ročenky v českých zemích 1801–1945	https://doi.org/10.57680/asep.0619233
31.	Současná bibliografie (od roku 1945)	https://doi.org/10.57680/asep.0619234
32.	Bibliografie českého literárního internetu	https://doi.org/10.57680/asep.0619235
33.	Bibliografie českého literárního samizdatu	https://doi.org/10.57680/asep.0619236
34.	Bibliografie českého literárního exilu (1948–1989)	https://doi.org/10.57680/asep.0619237
35.	Český literární exil 1948–1989	https://doi.org/10.57680/asep.0619238
36.	Česká literatura v překladu	https://doi.org/10.57680/asep.0619239
37.	Bohemistické konsorcium	https://doi.org/10.57680/asep.0619240
38.	Česká literární bibliografie v systému ISIS – interní dobové příručky a pomocné materiály.	https://doi.org/10.57680/asep.0584645

Příloha 1: Seznam datových sad ČLB uložených v repozitáři ASEP