



Archivační řád Českého sociálněvědního datového archivu (ČSDA)

Verze 2.1. 2023

Český sociálněvědní datový archiv
Sociologický ústav AV ČR
Jilská 1, 110 00 Praha 1,
e-mail: archiv@soc.cas.cz
URL: <http://archiv.soc.cas.cz>



1. ARCHIVAČNÍ ŘÁD	4
1.1. Přehled pracovních pozic v řízení informačního archivačního systému ČSDA	4
1.2. Společné a jednorázové činnosti	4
2. SPRÁVCE AKVIZICE A PŘÍJMU	6
2.1. Rozsah odpovědnosti	6
2.2. Plánování vyhledávání dat k archivaci	6
2.3. Kontaktování producentů a navazování spolupráce	7
2.4. Uzavření dohody o depozici dat	7
2.5. Odpověď na žádost o depozici dat	8
2.6. Přijetí dat	8
2.7. Přijímané datové formáty	8
2.8. Primární kontrola kvality	9
2.9. Vytvoření AIP	9
2.10. Spolupráce na auditu SIP a AIP a vypořádání výsledků auditu	10
2.11. Vytvoření popisné informace a předání AIP do archivního skladu	10
2.12. Koordinace aktualizací	10
3. ADMINISTRÁTOR SYSTÉMU ČSDA	11
3.1. Databázový systém archivu	11
3.1.1. Archiv konfigurace systému	11
3.1.2. Archiv elektronických materiálů k datovým souborům	11
3.1.3. Archiv analogových materiálů k datovým souborům	11
3.1.4. Archiv dokumentů	12
3.1.5. Databáze souborů určených k depozici - Databáze SIP	12
3.1.6. Databáze archivovaných datových souborů – Databáze AIP	12
3.1.7. Databáze souborů určených k distribuci - DIP	12
3.1.8. Databáze verzí souborů	12
3.1.9. Databáze uživatelských žádostí	13
3.1.10. Seznam pracovníků archivu	13
3.1.11. Seznam klientů archivu	13
3.2. Správa požadavků na systém	13
3.3. Monitoring výkonů archivu	13
3.4. Zpracování dotazů	13
3.5. Správa verzí	13
3.6. Správa technické báze archivu	14
3.7. Správa paměťových úložišť	14
3.8. Záložní paměťová média	14

3.9. Kontrola chyb	14
4. SPRÁVCE UŽIVATELŮ ČESKÉHO SOCIÁLNĚVĚDNÍHO DATOVÉHO ARCHIVU	16
4.1. Konkrétní činnosti vázané na funkci správce	16
4.2. Koordinace přístupových aktivit	16
4.3. Aktivace žádostí a doručení odpovědi	16
4.4. Vyhledávání dat a dokumentů	17
4.5. Poskytování přístupu k datům uloženým v ČSDA	17
4.6. Odpověď na žádost mimo systém Dataverse	17
4.7. Vytvoření „výstupního balíčku“ (DIP)	17

1. Archivační řád

Tento archivační řád popisuje činnosti vykonávané v rámci informačního archivačního systému Českého sociálněvědního datového archivu (dále jen ČSDA) a je strukturován podle postupu archivace datového souboru od fáze před-příjmové k fázi poskytování datového souboru uživatelům archivu. Jednotlivé fáze archivačního procesu naplňují činnosti digitálního archivu v souladu s principy Otevřeného archivačního informačního systému OAIS (Open Archival Information System)¹, přičemž popisované pracovní činnosti lze přiřadit k příslušným funkcím v OAIS².

1.1. Přehled pracovních pozic v řízení informačního archivačního systému ČSDA

Uvedené pracovní pozice jsou spojeny s odpovědností za příslušné oblasti, realizace činností probíhá ve spolupráci s ostatními pracovníky ČSDA v rámci aktuálního rozdělení práce včetně zapojení pracovníků mimo řídicí pozice. Přehled nezahrnuje řídicí pozice v rámci činností ČSDA mimo informační archivační systém. Podrobné popisy funkcí ve vztahu k funkcím OAIS jsou popsány v dokumentu „Zavedení principů OAIS v činnosti Českého sociálněvědního datového archivu“, který je k dispozici všem pracovníkům oddělení na společném disku na adrese X:\datarchiv\OAIS.

Pracovní proces v ČSDA je realizován v rámci následujících pozic:

1.1.1 Vedoucí archivu

Vedoucí archivu je odpovědný za řízení archivačního procesu jako celku.

1.1.2 Správce akvizice a příjmu

Správce akvizice a příjmu má na starosti před-příjmové fáze, Příjem, Audit dodávek a vyjednávání dohod o předání dat z Administrace. Monitoruje požadavky depozitorů.

1.1.3 Správce (administrátor) systému

Správce (administrátor) systému má na starosti všechny funkce Správy dat a Archivního skladu, funkci Monitorování technologie z Plánování ochrany a Správu konfigurace systému z Administrativy.

1.1.4 Správce uživatelů archivu

Správce uživatelů archivu má na starosti Přístup a funkce týkající se uživatelů v Administraci: aktivace žádostí a zákaznický servis. Monitoruje požadavky uživatelů.

1.2. Společné a jednorázové činnosti

Některé funkce jsou naplňovány jednorázově, zpravidla vypracováním příslušných dokumentů a jejich pravidelnou (nikoliv ale neustálou) aktualizací. Zejména se jedná o:

- Administrace - Aktualizace archivních informací
- Administrace - Vypracování standardů a pravidel

¹ <http://public.ccsds.org/publications/archive/650xom2.pdf>

² Blíže viz interní dokument „Zavedení principů OAIS v činnosti Českého sociálněvědního datového archivu“.

- Administrace - Kontrola fyzického přístupu do archivu
- Plánování ochrany - Vytváření designu informačních souborů SIP/AIP/DIP a Vývoj migračních plánů
- Plánování ochrany - Vývoj strategií a standardů uchování dat

2. Správce akvizice a příjmu

2.1. Rozsah odpovědnosti

Správce akvizice a příjmu má na starosti:

a) Před-příjmové (pre-ingest) fáze (nejsou součástí OAIS) od vyhledávání potenciálních depozitorů po vyjednávání dohod o předání dat do archivu. Ve spolupráci s vedoucím ČSDA plánuje činnosti spojené s akvizicí dat, předkládá plán ke schválení vedoucím, rozděluje úkoly v oblasti akvizice dat, kontroluje průběh činností a plnění plánu, vyhodnocuje výsledky a navrhuje aktivity na podporu akvizice dat v rámci projektové činnosti ČSDA.

b) Příjem dat a metadat (SIP) zahrnuje převzetí „datového balíčku“ (SIP), kontrolu součástí tohoto balíčku (primární audit dat), opravy a doplnění ve spolupráci s depozitorem. Správce má na starosti transformaci přijatých dat do podoby AIP, předání SIP a AIP k auditu a zpracování případných připomínek z tohoto auditu, vytvoření popisné informace a předání AIP správci systému k zařazení do databáze.

c) Zpracovávání případných aktualizací dat a metadat.

Konkrétně to zahrnuje následující činnosti:

- plánování vyhledávání dat
- monitoring výzkumných projektů
- kontaktování potenciálních depozitorů a vyjednávání dohod o předávání dat
- uzavření dohody o depozici dat
- odpověď na žádost o depozici dat
- přijetí dat
- primární kontrola kvality
- vytvoření AIP
- spolupráce na auditu SIP a AIP a vypořádání výsledků auditu
- vytvoření popisné informace a předání AIP do Archivního skladu
- koordinace aktualizací

2.2. Plánování vyhledávání dat k archivaci

Správce akvizice dat vytváří plán kontaktování producentů dat a shromažďování dat se záměrem (1) žádat o archivaci dat konkrétních studií a/nebo (2) získat informace o dostupných datových souborech za účelem archivace a zpřístupnění a/nebo (3) navázat širší spolupráci při akvizici dat. Plán zahrnuje zejména:

- vymezení oblastí a/nebo výčet institucí, na něž se zaměří vyhledávání dat;
- rozdělení odpovědnosti za jednotlivé oblasti a/nebo instituce mezi pracovníky ČSDA;
- harmonogram;
- specifikaci postupu.

Pro účely vytvoření plánu může správce zadat požadavek na monitoring výzkumných projektů, vedle toho využívá výsledky přehledových studií, šetření a analýz zdrojů dat realizovaných v rámci výzkumné činnosti ČSDA. Draft plánu předloží k širší diskusi v oddělení za účelem upřesnění volných kapacit a sladění s ostatními úlohami řešenými v ČSDA. Vedoucí ČSDA schvaluje plán a pověří pracovníky jeho řešením. Správce plán předá administrátorovi systému k zařazení do archivu konfigurace systému.

Správce akvizice dat průběžně provádí kontrolu plnění plánu a v případě potřeby, eventuálně na přímou žádost vedoucího ČSDA, plán aktualizuje. Aktualizované verze předává administrátorovi systému k zařazení do archivu konfigurace systému. Případné zásadní revize plánu předem schvaluje vedoucí ČSDA a pověřuje pracovníky jejich implementací.

2.3. Kontaktování producentů a navazování spolupráce

Pracovníci pověřeni plněním plánu kontaktují v rámci stanovených úkolů určené producenty dat. Tuto činnost, pokud možno, doprovází propagace sdílení dat a datových služeb včetně nabídky specifických služeb ČSDA (např. organizace prezentací a seminářů a konzultace).

V případě identifikace souborů vhodných ke zpřístupnění v archivu pracovníci pověřeni plněním plánu zajistí základní informace o materiálech k archivaci v souladu s aktuální šablonou SIP, v které jsou popsány akceptovatelné typy a formáty dat i nezbytné náležitosti týkající se informací o výzkumu. Z nutných informací se zejména jedná o:

- název projektu a/nebo databáze
- ustanovení kontaktní osoby na straně depozitora
- institucionální prostředí (hlavní výzkumník, instituce, pravděpodobný majitel práv)
- základní informace o projektu (téma, zaměření, metoda)
- základní informace o databázi (typ dat, anonymita, rozsah - počet souborů, formáty, média)
- základní informace o dokumentaci (dostupnost, stav, formáty)

V případě neúspěšného uzavření jednání s producentem podá pracovník komunikující s potenciálním depozitorem zprávu správci akvizice dat. V případě identifikace vhodných souborů a zjištění základních informací založí pro jednotlivé soubory záznamy v databázi souborů určených k depozici, získané informace zaznamená do této databáze a podá zprávu správci akvizice dat.

2.4. Uzavření dohody o depozici dat

Správce akvizice dat posoudí informace o materiálech k archivaci získané na základě žádosti o depozici (viz níže) nebo při kontaktování producentů dat a vydá pokyn k přípravě dohody o depozici nebo požádá o doplnění informací anebo podá vedoucímu ČSDA návrh na odložení či odmítnutí archivace. Posouzení informací může založit na konzultacích s dalšími pracovníky ČSDA nebo externími experty, přičemž případné náklady předem schvaluje vedoucí ČSDA.

Vyjednání dohody provádí zpravidla pracovník pověřený odpovědí na žádost o depozici. Správce akvizice dat může např. z kapacitních nebo organizačních důvodů pověřit jiného člena týmu ČSDA.

Tento pověřený pracovník:

- 1) projedná s depozitorem zařazení deponovaných materiálů do kategorie přístupnosti a případné specifické požadavky ohledně zpřístupnění dat a formulaci dohody,
- 2) na základě zjištěných informací a s využitím vzorové dohody připraví draft dohody o depozici,
- 3) draft předloží ke schválení správci akvizice dat a vedoucímu ČSDA,
- 4) za asistence vedoucího ČSDA předloží draft dohody depozitorovi a vedení SOÚ k posouzení a zajistí výměnu a zapracování připomínek,

5) za asistence vedoucího ČSDA zajistí podpis dohody ve třech kopiích ředitelem SOÚ a následně oprávněným zástupcem depozitora.

Minimálně jedna kopie zůstává u depozitora, jednu pracovník předá k uložení na sekretariát SOÚ, třetí předá administrátorovi systému uložení do archivu dokumentů.

Zprávu o uzavření dohody předá pracovníkovi pověřenému řízením příjmu a zaznamená do databáze souborů určených k depozici.

2.5. Odpověď na žádost o depozici dat

Správce přebírá žádosti o depozici ze správy uživatelských žádostí, průběžně kontroluje databázi uživatelských žádostí, zda neobsahuje žádosti o depozici a přiděluje úkoly vyjednání depozice kvalifikovaným pracovníkům archivu včetně sebe.

Žádost o depozici je přidělena k vyřízení obratem, nejpozději do 3 pracovních dnů po obdržení žádosti archivem. Pověřený pracovník založí záznam v databázi souborů určených k depozici.

Pověřený pracovník kontaktuje žadatele o depozici nejpozději do 5 pracovních dnů po obdržení žádosti archivem, sdělí mu stručnou základní informaci o postupu depozice, a pokud je žádost o depozici neobsahuje, požádá o doplnění nebo upřesnění informací nutných k základní identifikaci materiálů k archivaci – výčet těchto základních informací je uveden výše. Celý proces vyjednávání probíhá stejným, výše popsáním, způsobem jako v situaci kdy vyjednávání iniciuje archiv.

2.6. Přijetí dat

Funkce zajišťuje převzetí souborů určených k archivaci od depozitorů. Správce akvizice dat nebo jím pověřený pracovník iniciuje funkci příjmu dat. Pokud nejde o stejného pracovníka, který předání dat vyjednával, předá mu správce současně všechny potřebné údaje včetně unikátních kódů sloužících k identifikaci dat a dohody o předání dat mezi Archivem (resp. SOU) a depozitorem. V dohodě jsou specifikovány podmínky předání a rozsah předaných dat. Data v originální podobě (SIP) jsou přijata, jejich příjem je potvrzen depozitorovi (součástí potvrzení je i přehled přijatých materiálů) a současně zaznamenán v rámci databáze souborů určených k depozici. Data samotná jsou pak uložena pod příslušným kódem v databázi SIP. Ne-elektronická data jsou uložena v Archivu analogových materiálů k datovým souborům, součástí samotného SIP musí být informace o existenci těchto analogových informací.

2.7. Přijímané datové formáty

Archiv data zpracovává pro účely dlouhodobé archivace, zabezpečuje je proti poškození a na základě dojednaných podmínek je zpřístupňuje sekundárním uživatelům. Pro naplnění těchto funkcí musí být data uložena způsobem umožňujícím jejich využití mimo původní výzkumný tým. Tímto se řídí i vymezení formátů jednotlivých typů dat, které jsou do archivu přijímány. V principu platí, že do archivu lze přijmout data v jakémkoli formátu, pokud je možné je převést na standardní formáty, v nichž jsou data uchovávána v rámci archivních databází AIP (Archival information package). Pracovníci archivu pověřeni vyjednáváním o příjmu datového balíčku od depozitora jsou odpovědní i za to, aby byla data přijata v podobě, která zajistí jejich integritu, umožní jejich dlouhodobé uchování v původním formátu (SIP – Submission information package) a převod do formátů, v nichž jsou data uchovávána ve vnitřním archivním systému (jako AIP). Ideální případ představuje, kdy jsou data již od počátku zpracovávána ve formátu, v kterém jsou následně uložena v datovém archivu, a to jak ve formě SIP, tak i AIP.

Pokud je data nutno transformovat, pak by pokud možno měl tuto transformaci provádět primární výzkumník, který je s obsahem a formou dat nejlépe seznámen. Pokud tato

transformace přímo tvůrci dat není možná nebo je odmítnuta, provedou převod dat pověření zaměstnanci archivu a to takovým způsobem, aby byla minimalizována ztráta informace (například kvůli vymazání definic „missing values“ při převodu mezi formáty různých programů pro zpracování kvantitativních dat).

Český sociálněvědní datový archiv preferuje data ve formátech, které uvádí následující tabulka. Jsou v ní uvedeny i vybrané akceptovatelné formáty.

Tabulka: Přehled preferovaných a přijatelných datových formátů

Typ dat	Preferované formáty dat	Přijatelné formáty dat
Kvantitativní data ve formě datových matic	Formáty programu SPSS (.sav; .por; .sys) Formáty v sociálních vědách nejpoužívanějších statistických programů jako je Stata (.dta) a SAS (.sas)	Formáty nejrozšířenějších databázových programů jako jsou MS Access (.mdb/.accdb) a dBase (.dbf) Formát programu MS Excel (.xls/.xlsx),
Grafická data	JPEG (.jpeg, .jpg), TIFF (.tif, .tiff) Adobe Portable DocumentFormat (PDF/A, PDF) (.pdf)	RAW image format (.raw)
Audio data	MPEG-1 Audio Layer 3 (.mp3)	Free Lossless Audio Codec (FLAC) (.flac) Audio InterchangeFileFormat (AIFF) (.aif) Wave form Audio Format (WAV) (.wav)
Video data	MPEG-4 (.mp4)	motion JPEG 2000 (.mj2)
Texty a dokumenty	MS Word (.doc/.docx) PDF/A nebo PDF (.pdf) Rich Text Format (.rtf)	HTML (.htm) Open Document Text (.odt) Prostý text (.txt)

2.8. Primární kontrola kvality

Data následně (do 5 pracovních dnů) projdou primární kontrolou kvality. Dojde především k ověření „čitelnosti“ všech digitálních souborů a jejich úplnosti. Ověří se zde také předběžně, zda obsah dat není v rozporu s legislativou (především zákonem na ochranu osobních údajů). Platí, že ČSDA nemůže odpovídat za kvalitu dat, ovšem měl by se pokusit alespoň o základní kontrolu úplnosti (například na základě srovnání datového souboru a dotazníku) a logické nerozpornosti dat, a především se zasadit o takovou kvalitu a úplnost metadat, aby mohl kvalitu dat posoudit uživatel. Výsledky této kontroly kvality jsou zaznamenány do databáze souborů určených k depozici. Pokud dojde k nalezení nějakých nejasností, vad, či se zjistí, že nějaká část dat chybí, je o tom vytvořena zpráva o kontrole, která se stává součástí AIP a současně je informován správce akvizice. Pracovníci pověřeni správcem a odpovědní za přijetí konkrétních dat mají na starosti i případnou další komunikaci s depozitorem – vyžádání aktualizací SIP, dotazy, které mají odstranit nejasnosti, atd. Po dokončení fáze kontroly a případných aktualizací SIP se provede digitalizace analogových součástí SIP a vše je uloženo na příslušné místo v informačním systému. SIP slouží především k udržení autenticity a integrity materiálů předaných depozitorem, touto zásadou se musejí řídit všechny operace s daty v této fázi.

2.9. Vytvoření AIP

Správce nebo jím pověřený pracovník vytváří strukturovaný archivní informační balíček, který je srozumitelný a využitelný nezávisle na depozitorech i ČSDA. Pro zařazení do AIP jsou soubory konvertovány na jednotné formáty na základě předpisu standardních datových formátů ČSDA (definovaném v šabloně AIP). Metadata, pokud je to možné, jsou převedena do XML schémat, především s využitím formátu DDI a to přinejmenším v rozsahu doporučení CESSDA. Vytváření AIP se nemusí odehrávat pouze na základě aktuálně obdrženého SIP, ale také s využitím již v databázích archivu dostupných informací, kdy využití těchto informací zprostředkovává funkce Správy systému. Kromě samotných dat a metadat popisujících poskytnutá data se součástí AIP stávají i další údaje popisující samotný proces archivace. Zejména se jedná o:

- Záznam o časovém průběhu významných kroků během archivace
- Zprávu o primární kontrole úplnosti a kvality dat
- Po vnitřním auditu dat řízeném Administrací se jeho výsledky stanou součástí AIP
- Vztahy k dalším datům a kolekcím dat v archivu už uloženým.

Součástí AIP je soubor „obsah“, který představuje „mapu“ popisující obsah AIP, tedy to, co obsahují jeho jednotlivé součásti.

2.10. Spolupráce na auditu SIP a AIP a vypořádání výsledků auditu

Po vytvoření AIP jsou odeslány jak AIP, tak i původní SIP k auditu. Provedením auditu je pověřen správcem pracovník, který neměl na starosti předchozí fáze příjmu. Výsledek auditu se zaznamená do databáze souborů určených k depozici. Pokud je výsledek auditu neuspokojivý, AIP je zaslán zpět k přepracování, doplnění apod. Po skončení auditu je vypracována konečná zpráva o příjmu dat, která je poskytnuta producentům/depozitorům dat a stává se také součástí AIP.

2.11. Vytvoření popisné informace a předání AIP do archivního skladu

Po auditu AIP a případných opravách a doplněních je vytvořen dokument „popisné informace“. Dokument je svým kódem propojen s příslušným AIP a záznamy o něm v ostatních databázích a je uložen v databázi souborů určených k depozici. Tento dokument obsahuje údaje, získané především z AIP, které umožňují identifikaci a vyhledání příslušného AIP v databázích archivu. Dokument by měl také umožnit vyhledat AIP na základě klíčových slov popisujících jeho obsah – klíčová slova musí odpovídat thesauru ELSST. Následně správce nebo jím pověřený pracovník předává AIP k uložení do archivního skladu, který uložení nového/aktualizované AIP potvrdí. Pověřený pracovník příjmu (obvykle ten, který data od jejich přijetí do ČSDA zpracovával) zajišťuje předání AIP do archivního skladu a také předání deskriptivní informace administrátorovi systému. Údaj o uložení AIP do archivního skladu se přidá k popisné informaci a ta je předána administrátorovi systému spolu s žádostí o aktualizaci databáze.

2.12. Koordinace aktualizací

Pod koordinací správce pracovníci příjmu provádějí aktualizace a opravy SIP a AIP. Tyto aktualizace a/nebo opravy mohou být iniciovány jak depozitory původních souborů, tak uživateli nebo samotnými pracovníky archivu. O oprávněnosti opravy/aktualizace rozhodne na základě zdokumentované žádosti od správce příjmu vedoucí ČSDA. Revize AIP a nové verze součástí AIP je potřeba dokumentovat (například pomocí syntaxí ke statickému programu, kde byla revize provedena). Záznamy o aktualizacích jsou prováděny do databáze verzí souborů administrátorem systému. Záznamy o aktualizacích a revizích by měly být i součástí AIP. Staré verze souborů zůstávají součástí AIP s tím, že musí být zřejmé, že jde o starší verze.

3. Administrátor systému ČSDA

Pracovní náplň administrace systému archivu zahrnuje shromažďování, ukládání a zprostředkování dat a kontrolu informací o všech činnostech, které v archivu probíhají. Administrátor systému spravuje databáze informací o archivu a v něm uložených datových souborů. Administrátor je zodpovědný za integritu databáze archivu. Administrátor systému dále spravuje v součinnosti spolu s IT oddělením SOÚ hardwarovou bázi archivu, tedy příslušné servery a diskové oddíly a je zodpovědný za zálohování archivních databází a samotných datových souborů.

3.1. Databázový systém archivu

Databáze archivu obsahuje následující položky:

- Archiv konfigurace systému
- Archiv elektronických materiálů k datovým souborům
- Archiv analogových materiálů k datovým souborům
- Archiv dokumentů
- Databáze SIP
- Databáze AIP
- Databáze DIP
- Databáze verzí
- Databáze uživatelských žádostí
- Seznam pracovníků archivu
- Seznam klientů archivu

3.1.1. Archiv konfigurace systému

Archiv konfigurace systému je struktura adresářů určená k ukládání digitálních verzí dokumentů popisujících konfiguraci systémů ČSDA, manuálů, vzorových dokumentů, zpráv o výkonech systému, strategických dokumentů, politik a plánů. Odpovědní pracovníci předávají dokumenty administrátorovi systému, který je ukládá. Do archivu konfigurace systému má plný přístup včetně přepisu administrátor systému a vedoucí ČSDA, ostatní pracovníci ČSDA mají přístup omezený na čtení a stahování. Administrátor systému kontroluje aktuálnost a provádí správu verzí uložených dokumentů. Verze, které pozbyly platnosti, jsou archivovány v samostatné sekci archivu. Aktuální verze vybraných dokumentů jsou zároveň k dispozici v papírové verzi na vhodném místě na pracovišti ČSDA.

3.1.2. Archiv elektronických materiálů k datovým souborům

Archiv elektronických materiálů k datovým souborům je struktura adresářů určená k ukládání digitálních dokumentů a dalších materiálů vázících se k depositovaným souborům. Adresáře s materiály k souborům jsou označeny unikátními kódy, které je propojují s příslušnými databázemi. Archiv má sekci s otevřeným přístupem a sekci s omezeným přístupem. K sekci s otevřeným přístupem mají přístup všichni pracovníci pro vkládání i přepis souborů. K sekci s omezeným přístupem mají přístup všichni pracovníci pouze pro čtení a stahování, pro vkládání a přepis má přístup pouze administrátor systému a vedoucí ČSDA. Administrátor systému provádí pravidelné kontroly archivu a aktualizace.

3.1.3. Archiv analogových materiálů k datovým souborům

Archiv analogových materiálů k datovým souborům je struktura fyzických složek s materiály vázícími se k archivovaným datům, které existují v jiné, než digitální formě. Materiály se

ukládají do složek na pracovišti archivu. Pracovníci pověřeni příslušnými úkoly předávají materiály administrátorovi systému, který je ukládá v archivu. Složky s materiály jsou označeny unikátními kódy, které je propojují s příslušnými databázemi. Součástí složky je vytištěný seznam obsažených materiálů, který je zároveň v elektronické podobě uložen v archivu elektronických materiálů. Pokud je materiál příliš objemný je uložen na vhodném místě a v archivu je uložen záznam o jeho uložení a dostupnosti. O zapůjčení materiálů je vedena kniha výpůjček.

3.1.4. Archiv dokumentů

Archiv dokumentů je archiv oficiálních a jiných formálních dokumentů vážících se k činnosti archivu (dohody o depozici dat, dohody o poskytnutí dat, smlouvy o spolupráci atd.) či jejich kopií. Pracovníci pověřeni příslušnými úkoly předávají dokumenty administrátorovi systému, který je ukládá v archivu. Dokumenty nebo složky s dokumenty jsou označeny unikátními kódy, které odpovídají příslušné činnosti a zařazení. Dokumenty obsahující citlivé osobní údaje nebo důvěrné údaje jsou umístěny v archivním trezoru.

3.1.5. Databáze souborů určených k depozici - Databáze SIP

Databáze SIP obsahuje záznamy o průběhu akvizice dat. Nový záznam vzniká po převzetí žádosti depozitora nebo po identifikaci vhodného souboru při kontaktování producenta dat. Zápis do databáze provádějí pracovníci přebírající žádosti o depozici nebo pověření příslušnými úkoly při akvizici dat.

Každý záznam má přidělen unikátní a neměnný kód. Kód propojuje záznam v databázi s materiály uloženými v archivu elektronických materiálů, v archivu analogových materiálů, v archivu dokumentů a záznamy v databázích AIP a DIP. Odpovědný pracovník dále zaznamená následující informace: datum převzetí žádosti o depozici / zaznamenání záměru depozice pracovníkem archivu, název projektu a/nebo databáze, instituci depozitora, jméno hlavního výzkumníka / řešitele projektu, z něhož data pocházejí (pokud je relevantní), jméno a adresy kontaktní osoby na straně depozitora. Zároveň v elektronickém archivu dokumentů založí adresář se stejným kódem. Pokud vznikne potřeba, založí složku v archivu analogových materiálů. Do adresáře a složky ukládá všechny materiály vážící se k datovému souboru, přičemž do databáze provede zároveň záznam o jejich uložení. Soubory jsou identifikovány unikátním kódem propojujícím záznam s ostatními databázemi a uloženými soubory. Zahrnuje přehled o obsahu souboru a záznamy o vydání nových verzí, edicí nebo stažení souboru atp. Záznamy vytvářejí pověření pracovníci příjmu.

3.1.6. Databáze archivovaných datových souborů – Databáze AIP

Tato databáze obsahuje všechny datové soubory archivované v ČSDA. Soubory v databázi AIP jsou identifikovány unikátním kódem propojujícím záznam s ostatními databázemi a uloženými soubory. Databáze AIP zahrnuje přehled o obsahu souboru a záznamy o vydání nových verzí, edicí nebo stažení souboru atp. Záznamy vytvářejí pověření pracovníci příjmu a správy databáze.

3.1.7. Databáze souborů určených k distribuci - DIP

Tato databáze je databází všech souborů DIP určených pro uživatele. Soubory jsou identifikovány unikátním kódem propojujícím záznam s ostatními databázemi a uloženými soubory. Databáze DIP zahrnuje přehled o obsahu souboru a záznamy o vydání nových verzí, edicí nebo stažení souboru atp. Záznamy vytvářejí odpovědní pracovníci zodpovědní za přístup k datovým souborům..

3.1.8. Databáze verzí souborů

Databáze verzí souborů je samostatná databáze verzí a edicí souborů SIP, AIP a DIP, která obsahuje podrobné informace o provedených změnách a odkazy na uložené materiály dokumentující tyto změny. Záznamy provádí správce verzí a edicí.

3.1.9. Databáze uživatelských žádostí

Tato databáze je databází uživatelských žádostí zaslaných mimo systém Dataverse. Záznamy provádějí pracovníci vyřizující konkrétní žádost o informace nebo datové soubory.

3.1.10. Seznam pracovníků archivu

Seznam pracovníků slouží k udržování přehledu o přidělených pracovních činnostech v rámci archivu, je propojen se všemi databázemi, tak aby mohla být ke každému datovému souboru a činnosti v rámci archivu přidělena odpovědná osoba.

3.1.11. Seznam klientů archivu

Seznam obsahuje údaje o klientech archivu, jejich žádostech a smlouvách s nimi uzavřených.

3.2. Správa požadavků na systém

Administrátor systému shromažďuje žádosti od pracovníků i uživatelů na změny v systému archivu, ukládá je do adresáře na pracovním disku a předkládá je k projednání vedoucímu ČSDA. Na základě zadání vedení archivu je administrátor zodpovědný za provedení příslušných úprav systému archivu.

3.3. Monitoring výkonů archivu

Administrátor monitoruje veškeré procesy v archivu. Monitoring je zpracováván s ohledem na vykazované zprávy ČSDA a možnosti systému. Administrátor stanovuje monitorované ukazatele a způsob spolu s periodicitou jejich sledování. Z monitoringu jsou vypracovány zprávy, které jsou ve stanovené periodicitě předkládány vedení ČSDA a ukládány do databáze archivu.

3.4. Zpracování dotazů

Administrátor vyřizuje všechny dotazy pracovníků archivu ohledně databáze archivu a poskytuje informace, které lze z této databáze získat.

3.5. Správa verzí

Administrátor spravuje verze a edice informačních souborů (SIP, AIP, DIP).

- 1) přiděluje unikátní kód souborům při depozici do archivu;
- 2) organizuje audit informačních souborů, tj:
 - na základě zpráv z databází SIP, AIP a DIP o vyhotovení informačních balíčků přiděluje jednotlivé balíčky k auditu pracovníkům a kontroluje jejich vyhotovení a
 - zaznamenává výsledek auditu do databází SIP, AIP a DIP a předává pokyny k odstranění nedostatků příslušnému pracovníkovi;
- 3) provádí správu verzí a edicí informačních balíčků, tj:

- shromažďuje a vyhodnocuje požadavky na aktualizace informačních balíčků, informace o chybách a problémech v SIP, AIP a DIP;
- podává návrhy na vytvoření nových verzí nebo edicí informačních balíčků;
- v případě identifikace vážných problémů podává návrhy na stažení SIP, AIP nebo DIP z datové knihovny;
- předává pokyny k aktualizaci příjmu, správě dat a na přístup;
- přebírá zprávy o realizovaných aktualizacích z příjmu, správy dat a přístupu;
- vede záznamy o verzích a edicích informačních balíčků v databázi verzí a archivuje podklady k aktualizacím.

Audit informačního balíčku je přidělen kvalifikovanému pracovníkovi ČSDA, který se nepodílel na jeho vytvoření. Audit zahrnuje kontrolu souladu s dohodou o depozici a logickou kontrolu úplnosti balíčku a realizace kroků přípravy balíčku. Pracovník o provedení kontroly provede záznam do databází SIP, AIP nebo DIP, případně zpracuje zprávu o nedostacích, kterou zařadí k databázi.

Nová verze vzniká na základě provedení jakékoliv změny v informačním souboru, nová edice vzniká na základě významného rozšíření (např. přidání dat za další zemi) nebo rozsáhlejší modifikace (např. transformace struktury datového souboru) informačního souboru. Návrhy na vytvoření nových edicí nebo definitivní stažení informačních souborů schvaluje vedoucí ČSDA.

3.6. Správa technické báze archivu

Administrátor je odpovědný za ukládání datových souborů archivu na k tomu určená datová úložiště. Administrátor pravidelně kontroluje uložená data, aby byla čitelná a bez chyb. Správa dále zahrnuje pravidelné zálohování datových souborů, tak aby je v případě výpadku hlavního datového úložiště bylo možné obnovit.

3.7. Správa paměťových úložišť

Administrátor spolu s IT oddělením Sociologického ústavu AV ČR spravuje paměťová úložiště archivu. Hlavním paměťovým úložištěm ČSDA je struktura adresářů vyčleněná na serveru SOÚ. Dalšími paměťovými úložišti je server Dataverse, kde jsou uloženy datové soubory (DIP), v určených datových formátech, které umožňují okamžitý přístup k datovému souboru ze strany klientů archivu. Dataverse server neslouží k archivaci dat, ale pouze k jejich distribuci klientům archivu.

3.8. Záložní paměťová média

Administrátor ukládá datové soubory na záložní média, která budou určena dokumentem „Plán na ochranu dat v případě katastrofy“. Administrátor v určeném časovém intervalu duplikuje datové soubory archivu a zajišťuje jejich uložení na odlišném a bezpečném místě. Tato činnost je uskutečňována kopírováním datových souborů na specifikovaná externí paměťová média, nebo pomocí síťového uložení dat mimo ČSDA a SOÚ.

Datové soubory archivu jsou uloženy na serveru, který je pravidelně zálohován. Dále je celý archivní obsah ukládán na externí datové úložiště, umístěné přímo na pracovišti archivu, tato záloha je prováděna pravidelně. Při aktualizaci je prováděna kontrola čitelnosti uložených dat, která je zaznamenána do interní databáze archivu.

3.9. Kontrola chyb

Administrátor v určeném časovém intervalu kontroluje datové soubory na všech paměťových úložištích archivu, tak aby při jakékoliv operaci s těmito soubory nedocházelo k žádným

chybám. Kontroly jsou prováděny pomocí zavedených metod kontroly integrity dat (CRC součty, PDI fixity information). Archiv používá MD5 checksums. „Fixity information“ je ve formě textového souboru uložena přímo u každého archivovaného datového souboru a také je uložena v interní databázi archivu a je kontrolována pokud dojde k jakékoliv manipulaci s daty.

4. Správce uživatelů Českého sociálněvědního datového archivu

Správce uživatelů má na starosti komplexní zákaznický servis. Zde ve spolupráci s vedoucím ČSDA plánuje činnosti spojené s uspokojením potřeb uživatelů. V rámci správy zákaznického servisu je správce uživatelů zodpovědný za včasné a správné předání „výstupního balíčku“ (DIP) v požadovaném formátu. Zároveň správce provádí průběžnou kontrolu úplnosti a správnosti předávaného balíčku. Do činnosti správce spadá také komunikace s uživateli a monitoring jejich požadavků.

4.1. Konkrétní činnosti vázané na funkci správce

- koordinace přístupových aktivit
- monitoring spokojenosti klientů
- kontrola funkčnosti systému Dataverse
- vyřízení požadavků mimo systém Dataverse
- správa a kontrola databáze registračních údajů uživatelů ČSDA
- kontrola plnění registračních podmínek
- vytvoření výstupního balíčku DIP

4.2. Koordinace přístupových aktivit

Správce uživatelů se ve spolupráci s vedoucím ČSDA podílí na zpracování koncepce zákaznického servisu, předkládá vedoucímu plán ke schválení, rozděluje úkoly v oblasti poskytování dat, kontroluje průběh činností a plnění plánu, vyhodnocuje výsledky a navrhuje aktivity na podporu přístupu k datům v rámci projektové činnosti ČSDA. Správce uživatelů zajišťuje správné fungování příjmů žádostí od uživatelů a je také zodpovědný za jejich včasné vyřízení.

Většina požadavků je v ČSDA zpracována automaticky bez aktivního zásahu správce, prostřednictvím systému Dataverse. Správce je zodpovědný za kontrolu správnosti funkce tohoto systému vůči uživatelům a v případě potřeby spolupracuje s administrátorem systému na odstranění nedostatků nebo znovu zprovoznění (restartování) systému. U tzv. ad-hoc požadavků na data (či metadata) vyřizovaných mimo systém Dataverse je správce uživatelů zodpovědný za vyřízení žádosti. Postupy při zpracování a lhůty pro vyřízení jsou konzultovány a upraveny ve spolupráci s vedoucím ČSDA.

Spokojenost klientů s datovými službami ČSDA je monitorována pomocí pravidelného průzkumu spokojenosti. Přístupové aktivity jsou na základě podnětů od uživatelů upravována tak, aby lépe odpovídala požadavkům uživatelů.

4.3. Aktivace žádostí a doručení odpovědi

Tato funkce zaznamenává seznam žádostí o data, porovnává aktuální obsah archivu ke zjištění zda jsou příslušná data k dispozici a zajišťuje dodání dat nebo informací uživatelům ČSDA. U požadavků zadaných prostřednictvím systému Dataverse je doručení odpovědi automatické. V případě požadavků zpracovávaných mimo katalog Dataverse se postupuje individuálně.

4.4. Vyhledávání dat a dokumentů

Vyhledávání dat a dokumentů je v systému Dataverse automatizováno. Pokročilé vyhledávání umožňuje hledání datových souborů či tabulek a také jednotlivých proměnných. Kritéria podrobného vyhledávání v popisu výzkumů a v popisu proměnných jsou možné podle metadatových položek popisu datového souboru. Vyhledávání umožňuje kombinaci více kritérií a je umožněno i pro neregistrované klienty archivu. Úkolem správce uživatelů je mimo jiné zjišťovat, zda nastavení systému vyhledávání odpovídá požadavkům klientů.

4.5. Poskytování přístupu k datům uloženým v ČSDA

Kromě funkce vyhledávání mají neregistrovaní uživatelé ČSDA přístup k metadatům výzkumu. Poskytování primárních datových souborů a doprovodných materiálů, včetně starších výzkumů, kde jsou k dispozici tabulky agregovaných dat, je podmíněno registrací uživatele.

V procesu registrace je uživatel vyzván k souhlasu s podmínkami pro užívání datových souborů z ČSDA, včetně podmínky použití dat pro nekomerční účely nebo výuku, respektování autorských práv a správné citování zdroje dat, nešíření datových souborů mezi další uživatele, důvěrné zacházení s daty atd. Zároveň uživatel souhlasí se zpracováním osobních údajů uvedených při registraci pro potřeby ČSDA a zavazuje se k poskytnutí bibliografických údajů všech publikací vzniklých na základě získaných dat.

Po registraci uživatel získává přístup k primárním datům on-line prostřednictvím systému Dataverse. V této fázi má registrovaný uživatel možnost prohlížet již nedefinované tabulky ze starších šetření, kde nejsou dostupná mikro-data, nebo stahovat datové soubory za účelem statistické analýzy.

Po registraci je správce uživatelů povinen zprostředkovat plný přístup uživatele ke všem datům a materiálům uloženým v ČSDA s výjimkou omezeného množství datových souborů, kde je požadován písemný souhlas depozitora. Přidělení přístupových práv by mělo proběhnout co nejrychleji, nejpozději však do tří pracovních dnů ode dne registrace uživatele.

4.6. Odpověď na žádost mimo systém Dataverse

V případě požadavků na vyhledávání dat a dokumentů uživatelé ČSDA používají vyhledávací systém Dataverse. Zpracovávání požadavků mimo systém je možné ve výjimečných případech a je řešeno individuálně pro každou žádost. Postupy při zpracování těchto požadavků a lhůty pro jejich vyřízení jsou konzultovány a upraveny ve spolupráci s vedoucím ČSDA.

4.7. Vytvoření „výstupního balíčku“ (DIP)

Výstupní balíček DIP, tedy vlastní datový soubor s metadatovým popisem a všemi souvisejícími materiály, opatřený unikátním PID je vytvářen procesem záznamu v prostředí Dataverse. On-line je dostupná metadatová dokumentace a informace o obsažených souborech. Prohlížení a stažení datových souborů je podmíněno registrací uživatele.

V případech vyřizování mimořádných požadavků mimo systém Dataverse je obsah DIP stanoven individuálně dle požadavků klienta a možností ČSDA.