

# Příloha č. 2

Datová struktura pro účely provedení zkušebního přesunu archiválií v digitální podobě do Národního archivu

## 1. Cíl

Cílem dokumentu je naplnit předpoklad uvedený v § 60a odst. 2 písm h) zákona č. 499/2004 Sb., který stanovuje, že žádost o udělení oprávnění k ukládání archiválií v digitální podobě musí kromě obecných náležitostí podání podle správního řádu obsahovat potvrzení Národního archivu o úspěšném provedení zkušebního přesunu archiválií v digitální podobě do Národního archivu. Předpokladem pro získání takového potvrzení je provedení přesunu archiválií v datové struktuře, která umožní zpracování dat Národním archivem. Záměrem je tedy stanovit datovou strukturu, která poslouží jak žadatelům o udělení oprávnění k ukládání archiválií v digitální podobě, tak i Národnímu archivem.

## 2. Východiska

Žadatel uchovává archiválie v digitální podobě ve formě archivního informačního balíčku (AIP) podle modelu OAIS. Protože žádná obecná norma nespécifikuje datovou strukturu AIP, lze předpokládat velkou různorodost podob AIP u jednotlivých žadatelů s ohledem na jejich potřeby (ty jsou dány zejm. druhy uchovávaných archiválií, vyhodnocením znalostní základny určené skupiny podle modelu OAIS, použitím konkrétního LTP systému). Podobně je tomu na straně Národního archivu. V takovém případě by bylo užitečné návrh datové struktury vytvořit s vědomím zásad snadné interoperability pro obě strany přesunu, softwarové nezávislosti jednotlivých LTP systémů a dlouhodobého uchovávání. Pro příslušný návrh je jeví jako nejvhodnější využít specifikací vytvořených v rámci projektu European Archival Records and Knowledge Preservation (E-ARK) v letech 2014–2017 a spravovaných Digital Information LifeCycle Interoperability Standards Board (DILCIS Board) s podporou Evropské komise a jejího programu CEF eArchiving Building Block.

Využita bude specifikace Common Specification for Information Packages (CSIP), která popisuje společné vlastnosti jakéhokoli informačního balíčku (SIP, AIP, DIP) v rámci OAIS, základní principy popisující tvorbu informačních balíčků a technická implementace těchto principů. Dále pak specifikace Archival Information Package (AIP), která doplňuje CSIP o konkrétní aspekty AIP. Obě specifikace ve verzi 2.0 ze dne 31. 5. 2019.

<https://github.com/DILCISBoard/E-ARK-CSIP/blob/gh-pages/pdf/eark-csip.pdf>

<https://github.com/DILCISBoard/E-ARK-AIP/blob/gh-pages/pdf/aip-specification.pdf>

## 3. Rámcový popis

Výstupní balíček je soubor ve formátu ZIP, který po rozbalení obsahuje kořenovou složku. Pro každý výstupní balíček SIP se generuje UUID, kterým se pojmenovává jak soubor ve formátu ZIP, tak i kořenová složka.

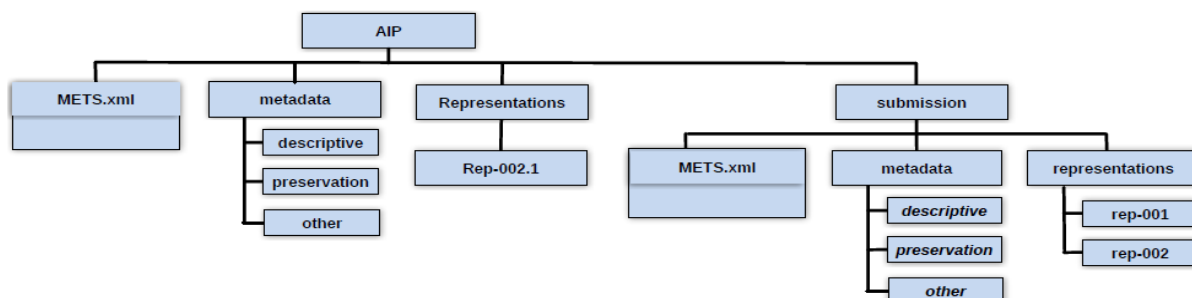
V kořenové složce se vytvářejí podsložky: „metadata“, „representations“, „submission“ a soubor mets.xml pro zaznamenání strukturálních metadat balíčku.

Složka „metadata“ obsahuje podsložku „descriptive“ se souborem XML, který zaznamenává popisná metadata intelektuální entity podle schématu EAD. Dále podsložku „preservation“ se souborem XML, který zaznamenává uchovávací metadata reprezentace a jednotlivých komponent.

Složka „representations“ obsahuje podsložku s verzí reprezentace intelektuální entity, ve které je podsložka „data“ s jednotlivými komponentami reprezentace.

Složka „submission“ stínuje strukturu balíčku AIP, přičemž navíc obsahuje složku „schemas“, která obsahuje schémata, podle kterých jsou používány soubory XML (např. METS, XLink, EAD, PREMIS).

Obecný příklad:



## 4. Popisná metadata

Popisná metadata jsou uložena v souboru ead3.xml, který leží v podsložce „descriptive“ složky „metadata“. Soubor musí být validní podle schématu formátu EAD3 (<https://www.loc.gov/ead/>). Zohlednit je nutné níže uvedené náležitosti, které upravují zaznamenání metadat podle § 12a vyhlášky č. 645/2004 Sb.

ead/control/recordid	UUID balíčku AIP
ead/maintenancestatus/maintenanceagency	informace o pečujícím archivu
ead/maintenancestatus/maintenanceagency/agencycode	identifikátor instituce z číselníku PEvA
ead/maintenancestatus/maintenanceagency/agencyname	název instituce
ead/archdesc	informace o archivním souboru
ead/archdesc/@level	„fonds“ pro archivní fond nebo „collection“ pro archivní sbírku
ead/archdesc/did	
ead/archdesc/did/unitid	číslo evidenčního listu NAD
ead/archdesc/did/unitid/@localtype	„NAD“
ead/archdesc/did/unittitle	název archivního souboru

ead/archdesc/did/origination	informace o původci
ead/archdesc/did/origination/corpname   persname	určení, zda jde o korporaci nebo fyzickou osobu
ead/archdesc/did/origination/corpname   persname/@identifier	identifikátor původce
ead/archdesc/did/origination/corpname   persname/part	název nebo jméno původce
ead/archdesc/dsc/.../c	informace o archiválii v digitální podobě
ead/archdesc/dsc/.../c/@level	„file“ pro úroveň složka nebo „item“ pro úroveň jednotlivost
ead/archdesc/dsc/.../c/did	
ead/archdesc/dsc/.../c/did/unitid	identifikátor archiválie přidělený digitálním archivem
ead/archdesc/dsc/.../c/did/unittitle	stručný obsah archiválie
ead/archdesc/dsc/.../c/did/unitdate	doba vzniku archiválie
ead/archdesc/dsc/.../c/did/unitdate/physdescstructured	informace o evidenčních jednotkách
ead/archdesc/dsc/.../c/did/unitdate/physdescstructured/quantity	počet evidenčních jednotek
ead/archdesc/dsc/.../c/did/unitdate/physdescstructured/unittype	druh evidenčních jednotek (uvádí se zkratka podle ZP z r. 2015)
ead/archdesc/dsc/.../c/did/langmaterial	informace o jazyku, ve kterém je archiválie pořízena
ead/archdesc/dsc/.../c/did/langmaterial/language	název jazyku
ead/archdesc/dsc/.../c/did/langmaterial/language/@langcode	zkratka jazyku
ead/archdesc/dsc/.../c/did/dao	odkaz na digitální komponentu archiválie a informace o datovém formátu archiválie, jeho verzi a velikosti v bytech
ead/archdesc/dsc/.../c/accessrestrict	informace o omezení nahlížení