

# Transformace prostorových dat pro účely trvalého uložení

Mgr. Martin Rehtorik, NAČR

# Schéma prezentace

1. Digitální archivace (dle AZ „trvalé uložení“)
2. Platná legislativa a povinnosti z ní plynoucí
3. Informační balíček SIP pro prostorová data
4. Metadata a jejich vytěžování

# Co to je trvalé uložení/archivace?

- Dle Archivního zákona č.499/2004 sb. se ukládají dokumenty
- Zpravidla se nearchivují aplikace nebo nosiče, služby (WMS)
- Archivují se data, logika, struktura, relace, vazby, objekty
- Archivace není datová záloha, ta má krátkodobý účel
- Archivace je založena na tzv. výběru, kdy se z množiny informací vyberou data (informace/dokumenty aj.) s hodnotou
- Digitální archivace je založena na zdokumentovaných formátech, doporučených postupech atd. aby se zajistila datová autenticita, integrita, interoperabilita a dlouhodobá čitelnost

# Proč, co a kdy digitálně archivovat?

- Protože je to dáno zákonem
- Protože chceme uchovávat, vytěžovat, zpracovávat, porovnávat a poskytovat cenné informace
- Informace, a to po dlouhou dobu, data jsou ropou 21.století a samotné archivy stojí možná před největší výzvou své dosavadní existence
- Archivaci bude nutné provádět nejen na konci životnosti systémů, ale v určitých případech i průběžně, např. změna legislativy, v pravidelných intervalech do tzv. snapshotů

# Legislativní povinnosti v AZ

AZ vymezuje v § 3 (1) Povinnost uchovávat *dokumenty a umožnit výběr archiválií* pro:

- **a)** organizační složky státu, **b)** ozbrojené síly, **c)** bezpečnostní sbory, **d)** státní příspěvkové organizace, **e)** státní podniky, **f)** územní samosprávné celky, **g)** organizační složky územních samosprávných celků, vytvářejí-li dokumenty uvedené v přílohách č. 1 nebo 2 k tomuto zákonu, **h)** právnické osoby zřízené nebo založené územními samosprávnými celky, vytvářejí-li dokumenty uvedené v přílohách č. 1 nebo 2 k tomuto zákonu, **i)** vysoké školy, **j)** školy a školská zařízení s výjimkou mateřských škol, výchovných a ubytovacích zařízení a zařízení školního stravování (dále jen „školy“), **k)** zdravotní pojišťovny, **l)** veřejné výzkumné instituce, **m)** právnické osoby zřízené zákonem,

# Co je to „dokument“?

Legislativa chápe jako dokument:

- dle AZ e) dokumentem každá písemná, obrazová, zvuková nebo jiná zaznamenaná informace, ať již v podobě analogové či digitální, která byla vytvořena původcem nebo byla původci doručena
- Dle eIDAS 35) „elektronickým dokumentem“ jakýkoli obsah uchovávaný v elektronické podobě, zejména jako text nebo zvuková, vizuální nebo audiovizuální nahrávka
- **Prostorová data = dokumentem je celý dataset (struktura)**

# Co to pro původce/tvůrce znamená?

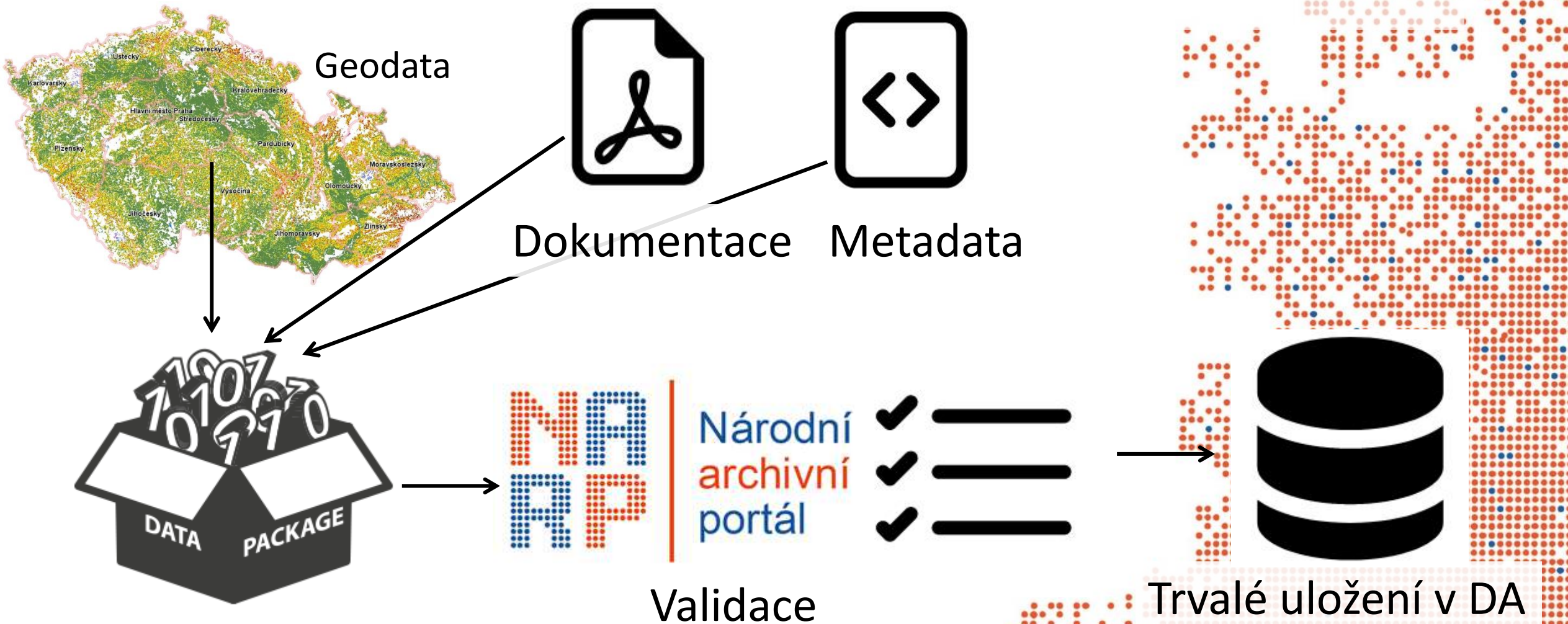
1. Maximální možnou pomoc ze strany archivu
2. Provádět datový export do informačního balíčku SIP
3. Volbu vhodných formátů (dle formátových doporučení DA)
4. Export, resp. tvorbu popisných metadat
5. Shromáždění, případně vytvoření potřebné dokumentace k datasetu (metadata pro datovou obnovu, SQL, Selecty atd.)
6. Nahrání informačního balíčku SIP do Archivního portálu

# Jak vlastně data k archivaci vypadají?

- Informační balíček SIP má (a bude mít) podobu předepsané povinně-volitelné adresářové struktury
- Navržená struktura v SIP se snaží maximálně zachovat logiku datové struktury (vychází se z evropského modelu)
- Povinné části jsou nezbytné kvůli validaci, neboť se chystá společný validátor na úrovni EU
- Volitelné části jsou na nižších úrovních, aby se struktura co nejvíce přizpůsobila potřebám původce/tvůrce



# Schéma balíčku s geodaty/GIS



# Struktura balíčku s geodaty/GIS

- Kořenový adresář */komponenty* + [mets.xml](#)

- */Geo\_SIP\_nazev*

[EARK – CITS Geospatial](#)

*/Documentation*

*/Representations*

*/Metadata*

*/Schemas*



# Adresář /Documentation a co tam patří?

- obsahuje doporučenou kontextovou dokumentaci k prostorovým datům v adresářové struktuře:

*/Structure* – např. katalog prvků pro vektory, legenda pro mapy

*/Rendering* – jednoduchý náhled podoby prostorových dat, vzorky datových sestav generovaných z GIS

*/Behaviour* – dotazy Select, kterými se získávají informace ze systému

*/Other* – další kontextová dokumentace, rozhovory, videa, ocenění atp.

**Ideálně výstupní formáty dle AZ!!**

# Adresář /Representation a co tam patří?

- Obsahuje vlastní prostorová data/data GIS

*/Nazev\_reprezentace* – podadresář je opět strukturovaný a obsahuje pro každý dataset podadresáře /Data se strukturou

*/GML nebo /GPKG aj., které odpovídají datovým formátům*

*/Metadata* – popisná metadata prostorových dat (INSPIRE)

*/Documentation* – opět možné data logicky strukturovat do adresářů (soubory SQL atd.)

# Adresáře /Metadata a /Schemas

- /Metadata je primárně určen pro popisná metadata používána archivy/paměťovými institucemi
- Povinně obsahuje podadresář /Descriptive a v něm bude vyžadován soubor *dc.xml*, který obsahuje základní popisná metadata datasetu dle schématu Dublin CORE a který bude vytěžován na Národním archivním portálu
- /Schemas obsahuje veškerá schémata pro různá metadatová a popisná XML, která jsou v balíčku obsažena, kvůli datové a metadatové validaci

# Mapování metadat

- Metadata dle ISO 19115, INSPIRE vycházejí stejně jako archivní standard EAD z tzv. Dublin CORE a je tam částečná shoda
  - = název, abstrakt, typ, klíčová slova, referenční datum, časový rozsah, licence, omezení přístupu,
  - ≥ Jazyk, nadřízený, podřízený (tzv. rodokmen)
  - ≥ tematické kategorie, norma metadat, souřadnicový systém, typ prostorové reprezentace (jako klíčová slova)
  - + Měřítko, prostorový rozsah, URL odkazy
  - \* Identifikátor, původ, zodpovědná organizace, norma metadat
  - Vázaný zdroj, údržba, formát, a informace o metadatech (id., datum)

# Teoretické cíle

- Dlouhodobá čitelnost archivovaných dat, provádění formátových migrací dle nejnovějších poznatků, poskytování dat majitelům, tvůrcům, původcům, snaha o zpřístupnění badatelům atd.
- De facto se jedná o pokus interoperability dle motta EU „Data driven economy“
- Abyste měli co odpovědět a ukázat až se Vás vnoučata zeptají: „Co jsi dělal/a?“

# Praktické Cíle

- Minimalizovat náklady finanční, časové atd. na archivaci a jít cestou maximální automatizace procesů
- Připravit a vyzkoušet vhodné postupy, upravit modul Národního archivního portálu tzv. ePřejímku na pokud možno automatický příjem datasetů s prostorovými daty
- Vytěžování prostorových metadat do archivních .xml, se kterými pracuje archivní katalogový systém
- Vybudování rozšířeného vyhledávání nad datasety s prostorovými daty



# Praktické Cíle

- Trvale archivované datasey budou zpřístupňovány zásadně v podobě balíčku DIP
- Chtěli bychom disponovat/mít přístup do geoportálu, který by mohl zobrazit prostorová data badatelům v elementární podobě

!! Archivní katalogový systém asi nikdy nebude mít plné funkce jako katalogy geoportálů

!! Archivní portál nesměřuje na pozici geoportal.gov.cz, jeho účel je jiný

# Transformace prostorových dat pro účely trvalého uložení

Děkuji za pozornost

Mgr. Martin Rehtorik, NAČR



Výzkumný ústav meliorací  
a ochrany půdy, v.v.i.