



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

Martin Černý

METADATA

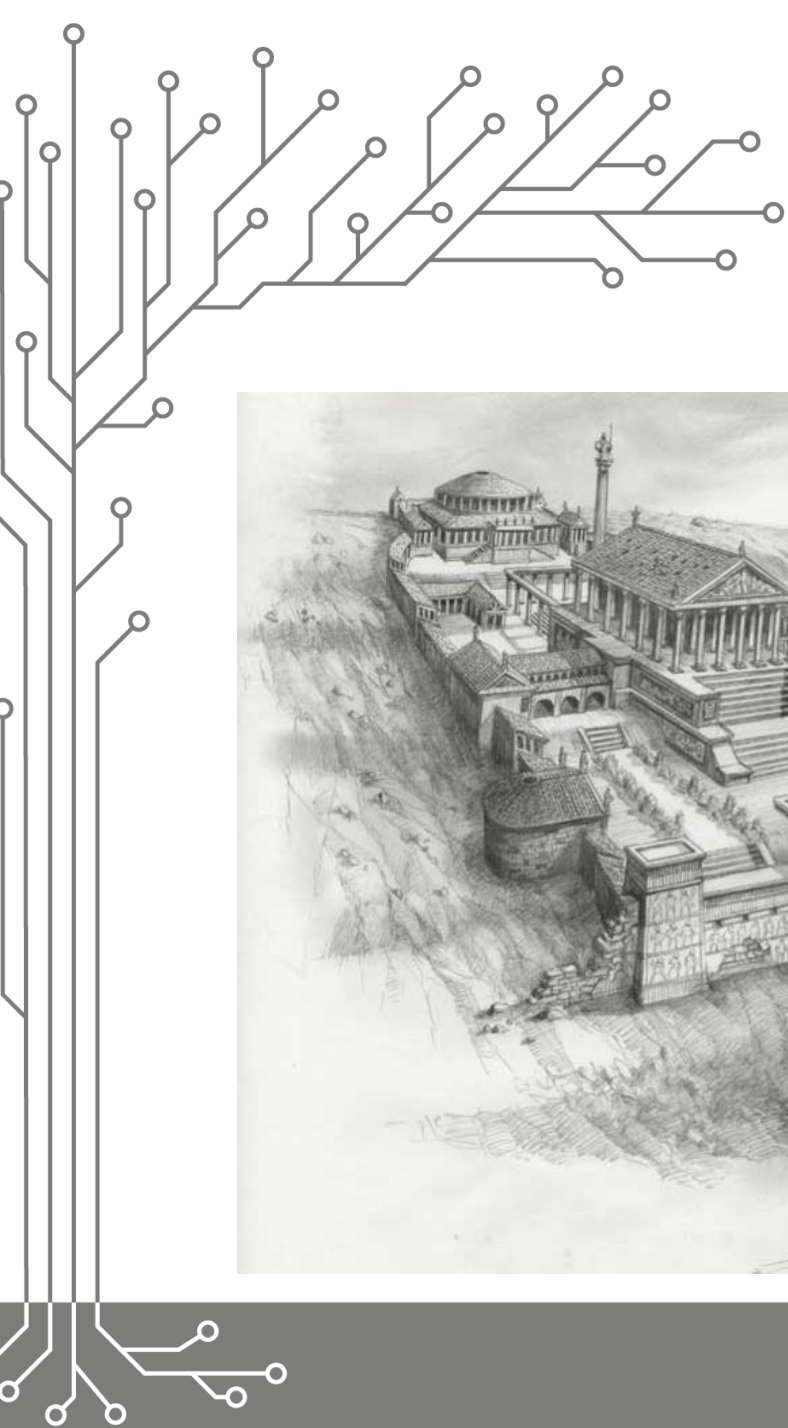
13.1.2022



Co to jsou METADATA

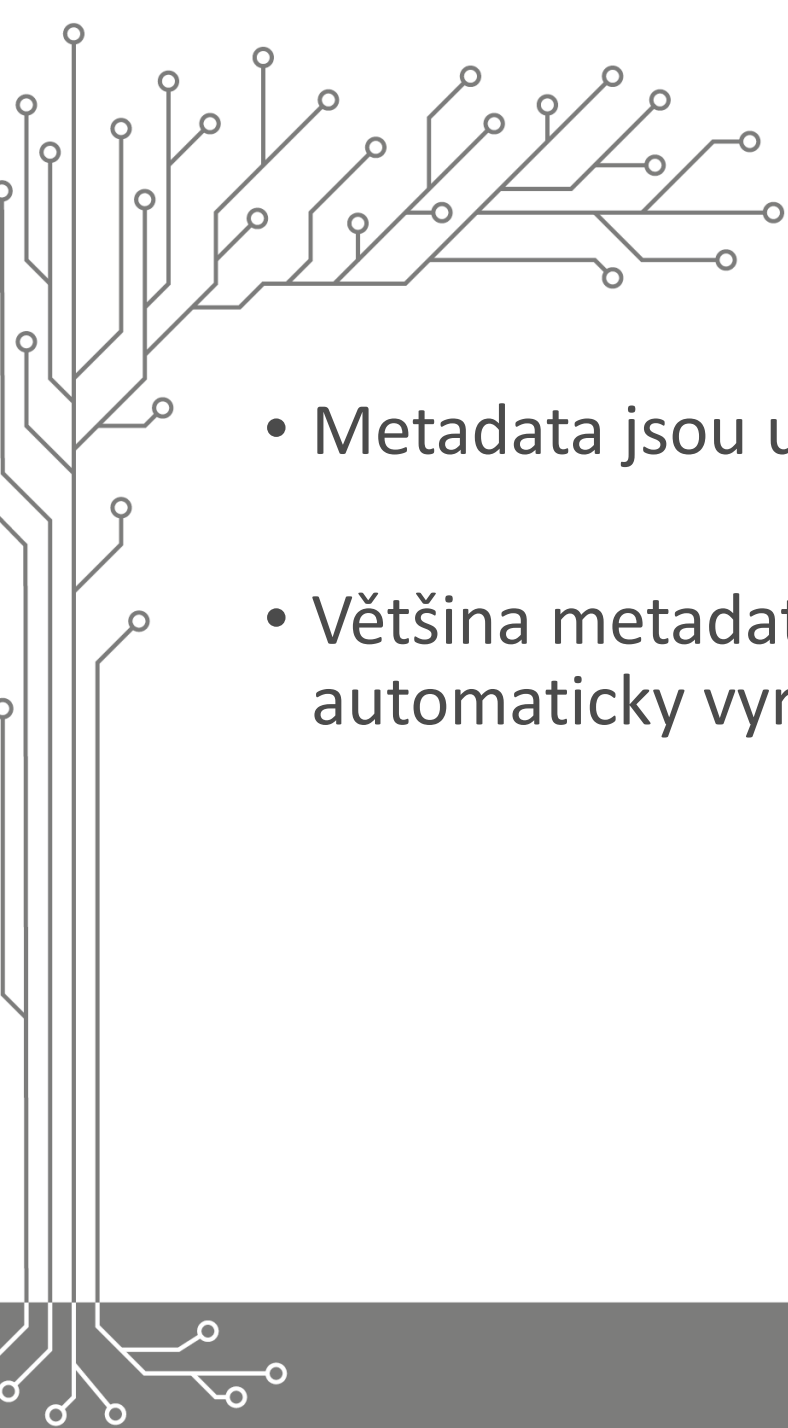
- data o datech
- strukturovaná data popisující jiná data či datové zdroje
- jsou stroji srozumitelné informace o webových zdrojích nebo dalších věcech -Tim Berners-Lee
- data obsahující informace o jiných datech
- data popisující souvislosti, obsah a strukturu zaznamenaných informací a jejich správu v průběhu času.





Co to jsou METADATA





Co to jsou METADATA

- Metadata jsou umělá, konstruktivní, užitečná
- Většina metadat je generována automaticky – software je automaticky vyrábí



Co to jsou METADATA

- Existují webové aplikace, kde jednoduše zjistíte, jaká metadata soubor obsahuje

<https://www.metadata2go.com/>

- Nebo existují aplikace, kde můžete metadata k již vytvořenému souboru změnit, nebo doplnit

<https://www.sejda.com/edit-pdf-metadata>

- A nebo můžete metadata z dokumentu zcela odstranit

<https://www.adarsus.com/en/remove-metadata-online-document-image-video/>

- První myšlenka metadat - knihovní katalog je 245 př. n. l. Alexandrijská knihovna
- Moderní katalogizační praxe knihoven - 30. léta 19. století
- Pojem metadata starý jen pár desetiletí (v angličtině se tento termín poprvé objevil v roce 1968)
- 90. léta 20. století znamenají postupný nárůst dokumentů věnovaných problematice metadat.
- Po roku 2000 s rozvojem a implementací nových informačních technologií
- 2003 – první ISO norma 19115



Vlastnosti METADAT

Za klíčové vlastnosti metadat jsou různými autory považovány:

- zjednodušení oproti realitě
- referenční povaha (říkají něco o něčem jiném)
- strojová čitelnost
- objektivita
- strukturovanost
- užitečnost pro uživatele
- interoperabilita
- elektronická forma



Rozdělení METADAT

Podle způsobu vytvoření

- metadata vytvořená ručně (nebo metadata vytvořená lidmi),
- metadata vytvořená automatizovaným způsobem (nebo metadata vytvořená stroji).



Rozdělení METADAT

Podle způsobu uložení

- interní metadata
- externí metadata



Rozdělení METADAT

Podle úrovně struktury (toto členění můžeme aplikovat jak na úrovni metadatového schématu, tak na úrovni metadatového záznamu)

- nestrukturovaná metadata,
- částečně strukturovaná metadata,
- strukturovaná metadata.



Rozdělení METADAT

Podle účelu (na úrovni metadatového schématu):

- metadata pro obecné účely (např. Dublin Core),
- metadata pro specifické účely, např. metadata pro geoprostorové informace (Dublin Core™ Metadata Element Set, Version 1.1, 2012)

Podle úrovně standardizace (na úrovni metadatového schématu):

- metadatové standardy de iure, které prošly procesem standardizace a jsou většinou uznávány mezinárodně (např. standardy ISO),
- metadatové standardy de facto, které mají dominantní pozici v rámci své oborové nebo geografické oblasti, ale nejsou standardy de iure.



Rozdělení METADAT

Podle využití (na úrovni metadatového schématu, záznamu nebo elementu):

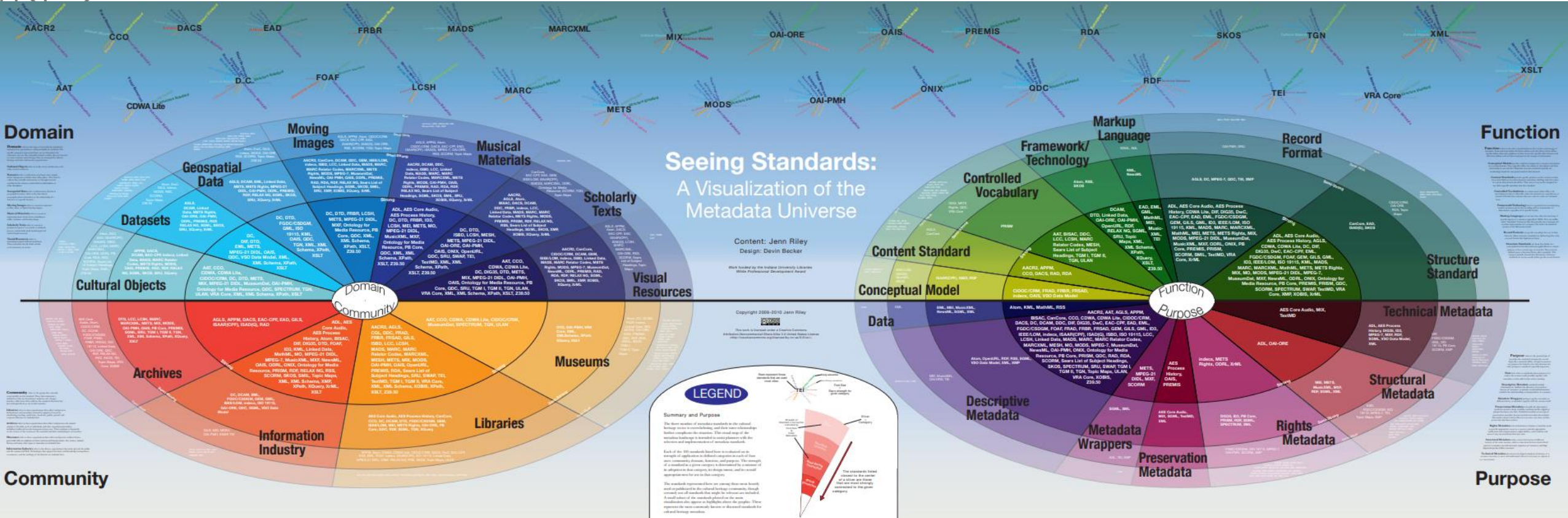
- Popisná metadata
- Administrativní metadata
- Strukturální metadata a technická metadata



METADATA jako riziko

- pro pachatele
- pro poškozené - bezpečnostní riziko

1000 metadatových standardů





METADATA a otevřená data

- Pro zápis do NKOD je nutné vyplnit potřebná metadata
- Povinná Metadata
- Nepovinná Metadata
- Interoperabilní metadata napříč otevřenými datovými portály umožňují, aby datové sady byly zjistitelné, znovu použitelné a prohledávatelné napříč portály, spíše než v nich „skladovat“



Interoperabilita

Právní interoperabilita

- překonání rozdílností mezi právními předpisy v jednotlivých členských státech,

Sémantická interoperabilita

- výměna informací mezi stranami bude správně chápán a zachován smysl vyměňovaných informací (smysl datových prvků a vztahů mezi nimi).

Syntaktická interoperabilita

- popis formátu vyměňovaných informací (gramatika, formát, schémata)

Technická interoperabilita

- technické stránky zajištění propojení informačních systémů a sdílených informací (specifikace rozhraní, spojové služby atd..)

Registrace do NKOD

1 Datová sada 2 Distribuce 3 Shrnutí

! Nedělejte stejné chyby jako řada poskytovatelů před vámi. Před registrací datové sady se ujistěte, že neděláte nějaké z [nejčastějších chyb v otevřených datech!](#)

Název datové sady česky	Název datové sady anglicky (volitelné)
Popis datové sady česky	Popis datové sady anglicky (volitelné)
Klíčová slova česky	Klíčová slova anglicky
Téma datové sady	Téma datové sady (volitelné)
Specifikace (volitelné)	
Periodičta aktualizace měsíční	Související území
Dotčené časové období od (volitelné) Vyberte začátek období. Pro změnu roku klikněte na rok v kalendáři.	Dotčené časové období do (volitelné) Vyberte konec období. Pro změnu roku klikněte na rok v kalendáři.
Časové rozlišení (volitelné)	Prostorové rozlišení v metrech (volitelné)
Jméno kurátora dat (volitelné)	E-mail kurátora dat (volitelné)
Odkaz na dokumentaci datové sady (volitelné)	
EuroVoc témata (volitelné)	

NAČÍST ZÁZNAM ZE SOUBORU NAČÍST ZÁZNAM Z URL



Povinná metadata

- **Název datové sady** - Měl by být krátký, věcný a výstižný. Jde o jedno z hlavních vyhledávacích kritérií, podle kterých budou uživatelé vaše data hledat. Zvažte například, zda slova jako "seznam" nejsou v názvu nadbytečná vzhledem k tomu, že většina datových sad bývají seznamy.
- **Popis datové sady** - co datová sada obsahuje, jak se dá použít a také na co si případně při jejím používání dávat pozor.
- **Klíčová slova** - charakterizující publikovanou datovou sadu. Slouží k vyhledávání datových sad napříč poskytovateli. Nemělo by se tedy jednat o interní identifikátory poskytovatele, nýbrž o slova, která by k popisu datové sady použilo co nejvíce uživatelů. Pro inspiraci doporučujeme se podívat na podobné datové sady, které jsou již v NKOD zaregistrovány.
- **Téma datové sady** - z evropského číselníku datových témat. Slouží k zařazení datové sady do jednoho ze základních témat, kterými lze klasifikovat datové sady napříč Evropou.
- **Související území** - Identifikátory území, kterého se publikovaná datová sada týká. Prvky se vybírají z existujících číselníků, popř. lze zadat jiné URL. Novou položku vložíte kliknutím na Přidat nové a vyplněním informací v dialogu.
- **Periodicita aktualizace datové sady** - jejíž uvedení a následné dodržování je předpokladem pro opakované využití otevřených dat uživateli. Pětina poskytovatelů

Registrace do NKOD

1 Datová sada **2** Distribuce **3** Shrnutí

< **2** + >

Podmínky užití distribuce ?

Autorské dílo

Neobsahuje autorská díla Jména autorů díla Odkaz na vlastní licenci k autorskému dílu

Originální databáze

Není autorskoprávně chráněnou databází Jména autorů originální databáze Odkaz na vlastní licenci k originální databázi

Zvláštní právo pořizovatele databáze

Není chráněna zvláštním právem pořizovatele databáze Odkaz na vlastní licenci týkající se zvláštního práva pořizovatele databáze

Osobní údaje

Neobsahuje osobní údaje

Přístup k distribuci

Typ distribuce

Soubor ke stažení

Odkaz na soubor ke stažení

Formát souboru ke stažení Media type souboru ke stažení

Media type komprese (volitelné) Media type balíčku (volitelné)

Odkaz na schéma distribuce (volitelné)

Název distribuce česky (volitelné) Název distribuce anglicky (volitelné)

[← DATOVÁ SADA](#) [SHRNUTÍ →](#)



Časté chyby při plnění metadat





1. Dokumentace či podmínky užití ve formátu PDF či Word
2. Nevhodně zvolená klíčová slova
3. Špatné dělení dat do distribucí datové sady
4. Chybně zasláná registrační zpráva do NKOD
5. Schémata v angličtině

Jak vypadají naše metadata

Datová sada [Skupiny](#) [Události](#)

Podíl paliv a technologií na výrobě elektřiny brutto (2005 - 2018)

Data a zdroje

-  [eru-bilance-elektriny-2018-202105141011.xlsx](#) [Prozkoumat -](#)
-  [eru-bilance-elektriny-2018-202105141011.csv](#) [Prozkoumat -](#)
-  https://github.com/cupido-minimus/knime_star/tr... [Prozkoumat -](#)
-  [eru-bilance-elektriny-2018-202105141011.json](#) [Prozkoumat -](#)

elektrina [spotřeba elektřiny](#)

Doplňující informace

Pole	Hodnota
Naposledy aktualizováno	6. leden 2022, 13:15 (UTC+01:00)
Vytvořeno	14. květen 2021, 10:12 (UTC+02:00)
Do databáze	Ano
Do portálu	Ne
Environmentální téma	Energetika
Jazyk datové sady	CZ
Kontaktní bod – kurátor CENIA – e-mail	miluse.rollerova@cenia.cz
Kontaktní bod – kurátor CENIA – jméno	Milúše Rollerová
Kontaktní bod – poskytovatel dat – e-mail	adriana.vesela@eru.cz
Kontaktní bod – poskytovatel dat – jméno	Ing. Adriana Veselá
Kontaktní bod – správce CENIA – e-mail	miroslav.elko@cenia.cz
Kontaktní bod – správce CENIA – jméno	Miroslav Elko
Otevřená data	Ano
Periodicita aktualizace	Ročně
Poskytovatel	Energetický regulační úřad
Publikovatelnost do	
Publikovatelnost od	
Související geografické území	ČR
Stav redakce záznamu	
Téma datové sady	Životní prostředí
Zdroj dat	
Časové pokrytí do	2018
Časové pokrytí od	2005



Budoucnost metadat

- Počet subjektů zainteresovaných v oblasti metadat bude růst
- Při rostoucím počtu metadatových schémat budou vznikat jejich registry tak, aby byly dostupné v platných verzích
- Dojde ke zúžení několika málo standardů
- Metadatové standardy budou vícejazyčné
- Vzhledem k rostoucímu významu metadat, se bude vyhodnocovat jejich efektivnost – způsoby hodnocení a metriky



Děkuji za pozornost

Martin Černý

martin.cerny@cenia.cz



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

