

Zuzana Nevěřilová
Hana Žižková
2024/25

CORE147

DIGITÁLNÍ DATA
V HUMANITNÍCH
A SOCIÁLNÍCH



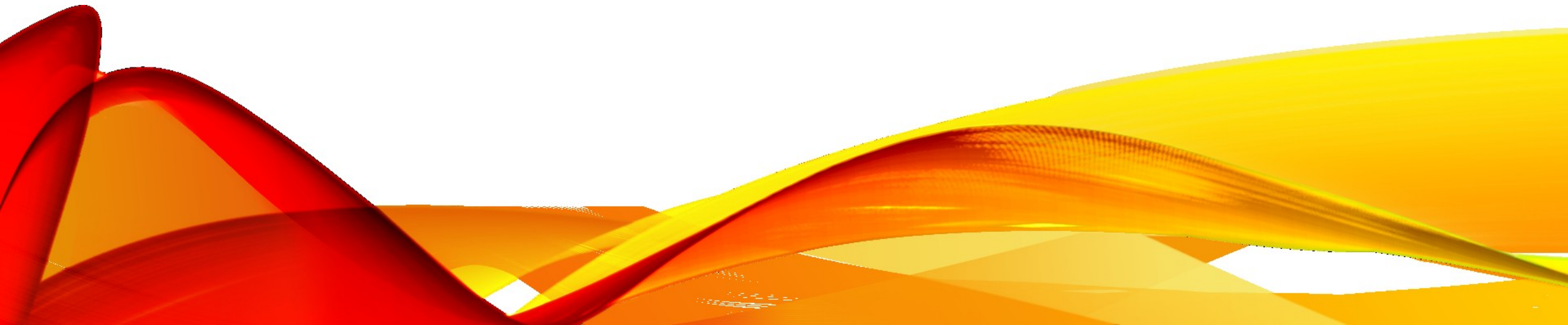


OBSAH

- Projektový management
- Datový management
- Metadata
- Publikování výsledků
- Na co nezapomenout

PROJEKTOVÝ MANAGEMENT

Co to obnáší a proč nás to má zajímat?



DIGITÁLNÍ HUMANITNÍ VĚDY PROJEKTOVÝ POHLED

Výzkum v humanitních oborech se

- demokratizoval
- zkomplikoval (nakládání s daty)
- zjednodušil (díky znovupoužitelnosti)
- stal záležitostí týmů (potřebuje definovat workflow)

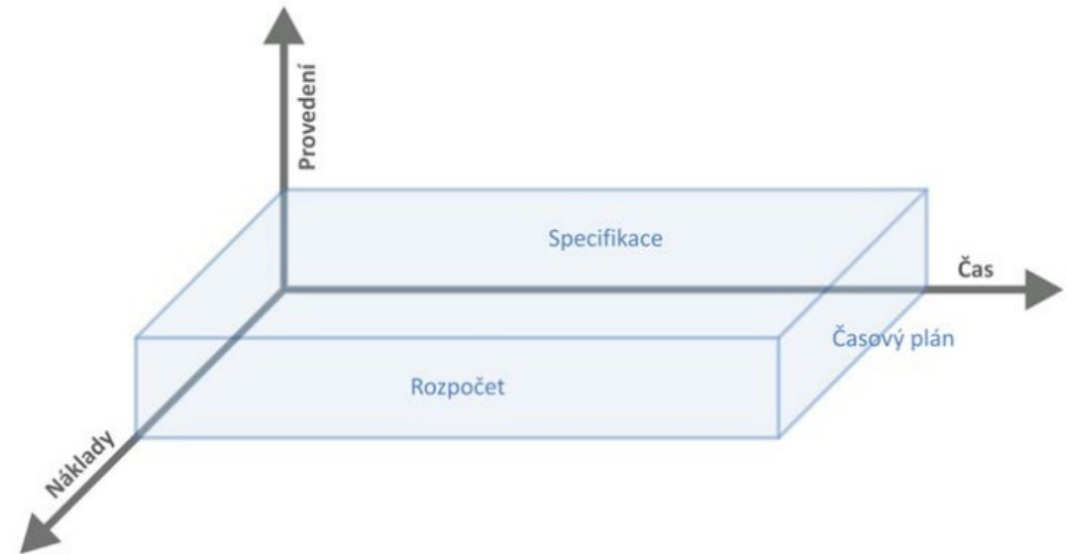
PROJEKTOVÝ MANAGEMENT DATA MANAGEMENT

Projekt je dočasný úkol s přesně stanoveným **cílem**, jehož splnění vyžaduje **organizované** využití odpovídajících **zdrojů**.

Úspěch projektu je charakterizován oceněním **výsledku** těmi, kteří jsou na projektu zainteresováni (**stakeholders**).

Úspěch:

- Projekt je funkční
- Stakeholders jsou spokojení
- Projekt je dodán včas
- Rozpočet není překročen
- Negativní vlivy jsou v normě



Rosenau, 2007 in Adámek, 2017

PROJEKTOVÉ AKTIVITY

- Plán projektu
- Zdroje (Assets)
- Komunikace
- Rizika
- (Rozpočet)

Plán

- Zahájení, ukončení
- Průběžné kontroly

Komunikace

- Uvnitř týmu
- Navenek (stakeholders)

Rizika

- Co se stane, když ...

DIGITÁLNÍ HUMANITNÍ VĚDY PROJEKTOVÝ POHLED

Zdroje, nástroje, metody a data

- Data
- Software
- Hardware
- Vzdálené služby
- Členové týmu
- Dovednosti
- Čas

Softwarová podpora workflow = **pipeline**

- Sestava softwarových nástrojů
- Nezávislé komponenty
 - Např. OCR, překlad

Ideální pipeline

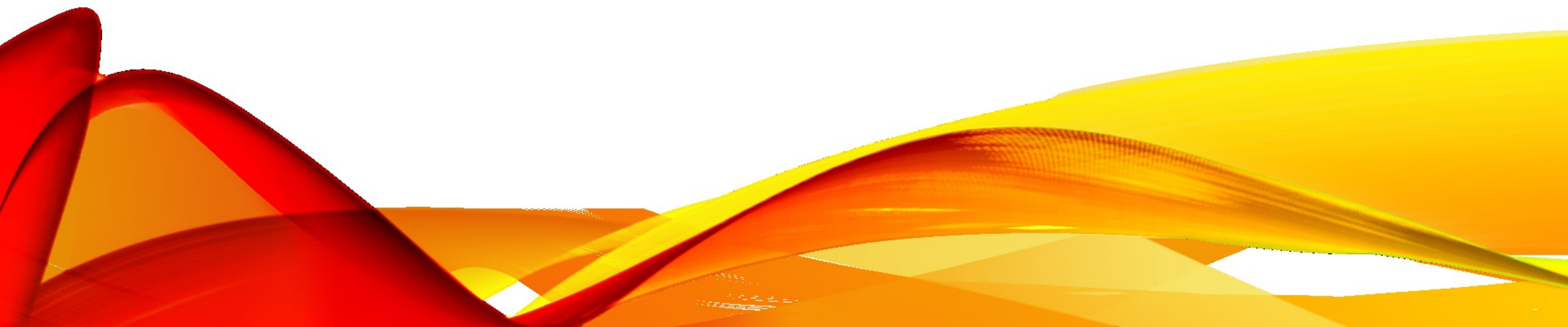
- Modulární (vyměnitelné komponenty)
- Interoperabilní (komponenty si „rozumí“)
- Transparentní (víme, co dělá co)
- Dokumentovaná (budou to vědět i ti po nás)

DIGITÁLNÍ HUMANITNÍ VĚDY EXISTUJÍCÍ NÁSTROJE

- Project Management for the Digital Humanities <https://scholarblogs.emory.edu/pm4dh/>
 1. Proposal – formulate guiding question, design the project
 2. Initiation – project charter, project team
 3. Planning – project scope, budget
 4. Execution – work plan, manage changes
 5. Closing – evaluation
- DevDH.org: Development for the Digital Humanities <https://devdh.org/>

DATA MANAGEMENT

Data jako v peřince.



DATA MANAGEMENT

Digitální kurátorství původně označovalo proces **vytváření a rozvoje důvěryhodných digitálních úložišť** pro výzkumná data, která mohou využívat současní i budoucí vědci, historici a badatelé.

Digitální kurátorství však využívají také obchodní podniky, aby uchovávaly, sdílely a přidávaly hodnotu svým digitálním datům a objektům s cílem zlepšit své provozní a strategické procesy.

Úspěšná digitální kurátorská činnost zajišťuje, že digitální objekty zůstanou **srozumitelné, přístupné, použitelné a bezpečné** v průběhu času i za hranicí technického zastarávání.

DATA MANAGEMENT

Proč management dat

- Potřebují to stakeholders
- Integrita dat (co je stejné, vypadá stejně)
- Zabráníme nechtěným duplikátům

OTÁZKY O DATECH

- Jaká data bude projekt produkovat?
- Kolik?
- Jak rychle budou narůstat?
- Jak často se budou měnit?
- Pro koho data jsou?
- Komu data patří?
- Kdo za data odpovídá?
- Jak dlouho musíme data uchovávat?

Volba úložiště

- Důvěryhodná úložiště
- Tam, kde jsou lidé zvyklí hledat
- Úložiště, která zabezpečují dlouhodobé uchování
- Úložiště, která umí sdílet standardizovaná metadata

DATA MANAGEMENT

Technické otázky

- Datové formáty
- Lokální úložiště a lokální politiky
 - Kde, odkud co kam kopírovat
- Zálohování
- Bezpečnost
 - Včetně fyzického zabezpečení počítačů
- Odpovědné osoby
 - Včetně kontaktů

Etické a právní otázky

- Duševní vlastnictví (intellectual property, IPR = IP regulation)
- Omezení pro použití dat (např. národní bezpečnost)
- Osobní údaje (GDPR – general data protection regulation)
- Související předpisy

DATA MANAGEMENT

FAIR

- Findable
- Accessible
- Interoperable
- Reusable

- F1. (Meta)data are assigned a **globally unique and persistent identifier**
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata **clearly and explicitly include the identifier** of the data they describe
- F4. (Meta)data are registered or **indexed** in a searchable resource
- A1.1 The protocol is **open, free**, and universally **implementable**
- A1.2 The protocol allows for an **authentication and authorization** procedure, where necessary
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable **language for knowledge representation**.
- I2. (Meta)data use **vocabularies** that follow FAIR principles
- I3. (Meta)data include qualified **references** to other (meta)data
- R1.1. (Meta)data are released with a clear and accessible data usage **license**
- R1.2. (Meta)data are associated with detailed **provenance**
- R1.3. (Meta)data meet domain-relevant **community standards**

DATA MANAGEMENT

CARE

- Collective Benefit
- Authority to Control
- Responsibility
- Ethics

- Data ecosystems shall be designed and function in ways that enable **Indigenous Peoples to derive benefit** from the data.
- Indigenous Peoples' rights and interests in Indigenous data must be recognized and **their authority** to control such data be empowered.
- Those working with Indigenous data have a responsibility to share how those **data are used to support Indigenous Peoples'** self-determination and collective benefit.
- Indigenous Peoples' **rights and wellbeing should be the primary concern** at all stages of the data life cycle.

(Carroll SR et al., 2022)

EXISTUJÍCÍ NÁSTROJE

Data Management Plan Online <https://dmponline.dcc.ac.uk>

Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

* What research project are you planning?

mock project for testing, practice, or educational purposes

* Select the primary research organisation

Organisation

- or -

No research organisation associated with this plan or my research organisation is not listed

* Select the primary funding organisation

Funder

- or -

No funder associated with this plan or my funder is not listed

Create plan

Cancel

EXISTUJÍCÍ NÁSTROJE

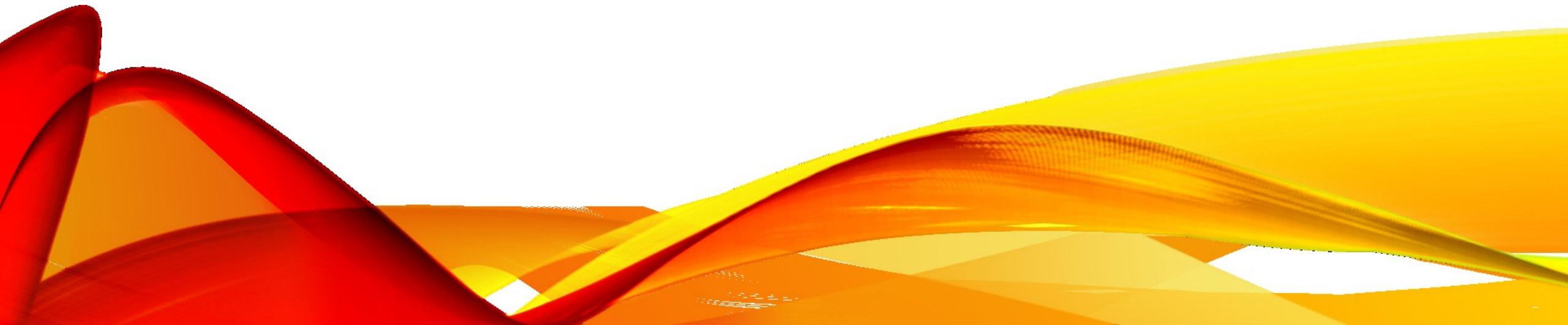
How to GO FAIR <https://www.go-fair.org/how-to-go-fair/>

FAIR principle	Question	FAIR enabling resource types
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type
F2	Which metadata schemas do you use for findability?	Metadata schema
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism
F4	In which search engines are your metadata records indexed?	Search engines
F4	In which search engines are your datasets indexed?	Search engines
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique
A2	Which metadata longevity plan do you use?	Metadata longevity
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema
I3	Which models, schema(s) do you use for your datasets?	Data schema
R1.1	Which usage license do you use for your metadata records?	Data usage license
R1.1	Which usage license do you use for your datasets?	Data usage license
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model



METADATA

Data o datech. Ale k čemu?



METADATA – K ČEMU?

Zdroje

Web pages, books, journal articles, images, songs, products, processes, people (and their activities), research data, concepts, and services

Popis zdroje

- Stručná sumarizace
 - Pomocí standardizovaných termínů jako „název“, „autor“, „datum vytvoření“, ...
 - Slovníky (vocabularies)
- Jednoznačná identifikace
 - Např. Dublin Core používá URI

METADATA – K ČEMU?

Nalezitelnost (findability)

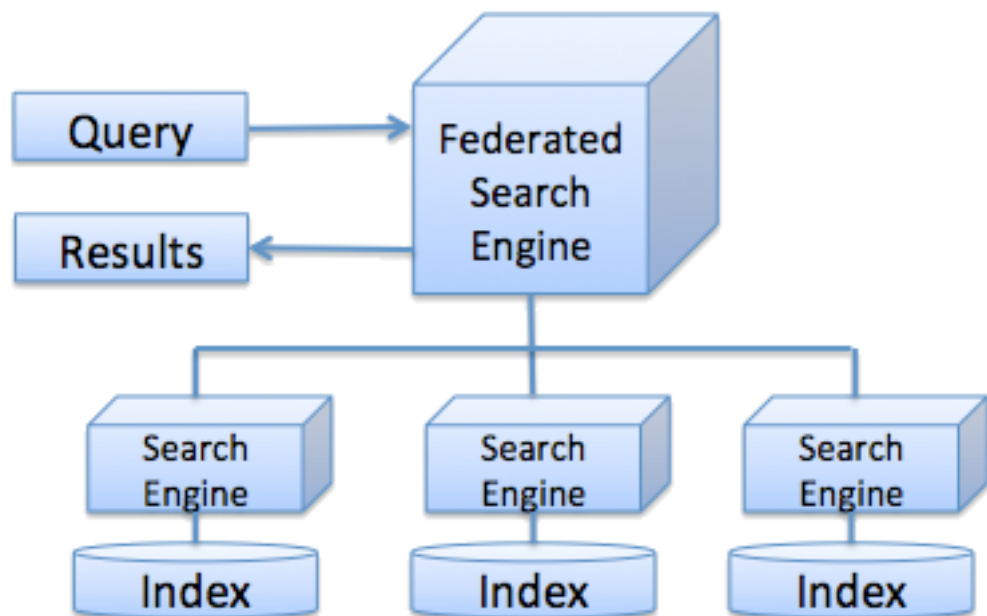
- Rysy (features)
 - Hledáme podle různých kritérií
- Klastrování (co patří k sobě, ať je u sebe)
 - Např. Rozdílné transliterace jmen
 - Explorativní vyhledávání – co by vás také mohlo zajímat



METADATA – K ČEMU?

Sdílení metadat

- Agregované hledání
 - Uživatel hledá na jednom místě, ale agregátor se zeptá zpřátelených vyhledávačů.
- Interoperabilita
 - Jedna služba postoupí metadata jiným službám.



METADATA JAKO SUMARIZACE ZDROJE

- Proč někdo zdroj vytvořil?
- Jak by zdroj vytvořen?
- Kdy?
- Kdo to byl?
- Kde?
- Velikost? Rozsah? Obsah?
Technická kvalita?
- Vztah k jiným zdrojům?

Základní prvky slovníku (vocabulary) Dublin Core

Creator, Contributor, Publisher, Title, Date, Language, Format, Subject, Description, Identifier, Relation, Source, Type, Coverage, and Rights

DOBRÁ METADATA

1. Musí být vhodná pro sbírku a materiál v ní a také pro **uživatele** a použití.
2. Musí podporovat **interoperabilitu**.
3. Musí podporovat **dlouhodobou ochranu objektů**, které popisují.
4. Musí používat standardní kontrolované **slovníky** (co, kde, kdy a kdo).
5. Musí být sama **digitálním objektem** a musí mít vlastnosti pro archivaci, dlouhodobé uchovávání, jednoznačnou identifikaci apod.
6. Musí být **ověřitelná** a validovatelná.

ZÁPIS METADAT POMOCÍ KEYS AND VALUES

Různá fakta o zdroji lze zapsat pomocí **klíčů (keys)** a **hodnot (values)**

John Doe is one of the nine authors of today's article How to use metadata for searching.

type=article

author=[John Doe, ...]

title=How to use metadata for searching

topics=[metadata, searching]

date=2020/12/14

TYPY METADAT

- **Deskriptivní** metadata popisují zdroj kvůli tomu, aby je bylo možné najít a identifikovat. Například: název, abstrakt, autor, klíčová slova.
- **Strukturní** metadata popisují, z čeho se komplexní zdroje skládají. Například: seřazení stránek v kapitolách.
- **Administrativní** metadata obsahují informace důležité pro správu zdroje. Například: datový formát, verze, přístupnost.
Administrativní metadata se dělí na:
 - **Management práv** metadata, která souvisejí s duševním vlastnictvím
 - **Metadata o uchování**, která souvisejí s tím, jak zdroj uchovat a archivovat.

STANDARDY METADAT

- **Knihovnictví:** Dewey Decimal Classification, Mezinárodní desetinné třídění (MDT), MARC21, Dublin Core (DC), Metadata Object Description Schema (MODS), Metadata Encoding & Transmission Standard (METS)
- **Zpracování textu:** TEI – Text Encoding Initiative: <https://tei-c.org/>
- **Zpracování znalostí:** Knowledge interchange format (KIF), Knowledge organization systems/services/structures (KOS), networked KOS (NKOS)
- **Semantický web:** RDF, OWL, Linked (open) data, Friend of a friend (FOAF), Schema.org
- **Digitální objekty:** MPEG-21 Digital Item Declaration Language (DIDL)

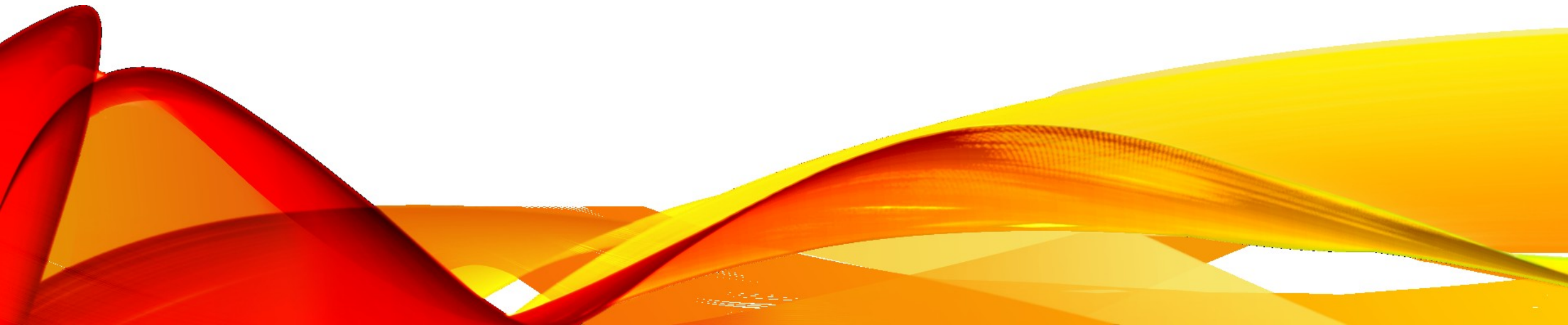
METADATA PODLE TYPU ZDROJE

- Zvuk: ID3
- Obraz: Exif
- Speciální metadata:
 - Geoprostorová data
 - Citační záznamy
 - Biochemická data
 - A mnoho dalších

METADATA VE SVĚTĚ

- A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.
<https://fairsharing.org/>
- WorldCat: <https://www.oclc.org/en/worldcat.html>
 - WorldCat is the world's most comprehensive database of information about library collections.

PUBLIKOVÁNÍ VÝSLEDKŮ



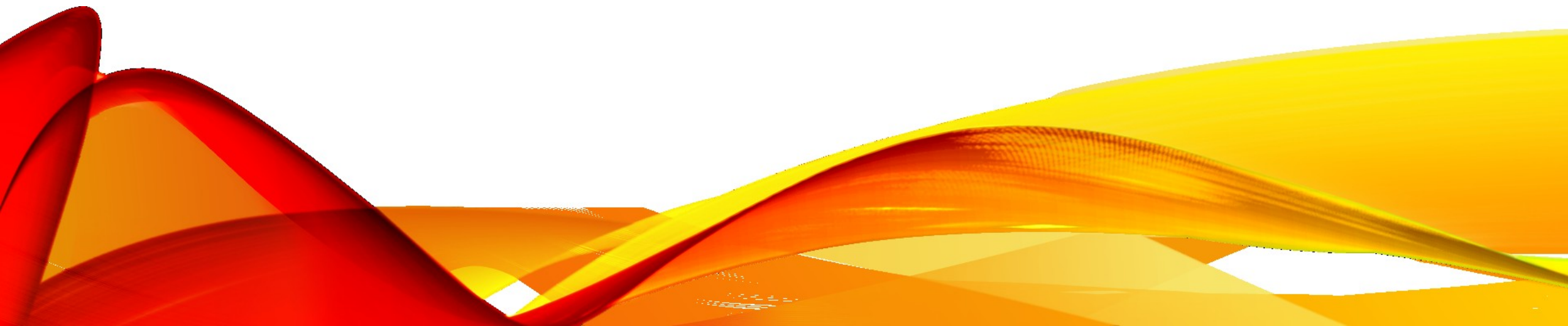
KOMUNIKACE A KOMUNITY

- K publikování zdrojů, datasetů, nástrojů a výsledků lze využít infrastruktury k tomu určené:
 - Bezpečnost
 - Perzistence
 - Nalezitelnost
- Budování komunit
 - Vědecká (interdisciplinární) komunita
 - Zapojení veřejnosti

SMĚREM K VEŘEJNOSTI

- Storytelling
 - Web
 - Interaktivní aplikace (i webová)
 - Sociální média
 - Plakáty
 - Podcasty

NA CO NEZAPOMENOUT



ZEPTĚJ SE PRÁVNÍKA ODPOVĚĎ BUDE „NE“

- Licence
 - https://wiki.creativecommons.org/wiki/Recommended_practices_for_attribution
- Práva všech dotčených
 - Autorské právo
 - Ochrana osobních údajů

DALŠÍ ČTENÍ

- Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., and Schnapp, J. (2012). Digital Humanities. The MIT Press.
<<https://mitpress.mit.edu/books/digitalhumanities>>
- de Smedt, K., Gardiner, H., Ore, E., Orlandi, T., Short, H., Souillot, J., and Vaughan, W. (1999). Computing in humanities education: A European perspective. University of Bergen.
<<http://korpus.uib.no/humfak/AcoHum/book/>>
- Terras, M., Nyhan, J., and Vanhoutte, E. (2013). Defining Digital Humanities: A Reader. Ashgate Publishing Company, Brookfield, VT, USA.
<<http://ccftp.scu.edu.cn/Download/505594d1-330f-4831-94a1-ec87d218345a.pdf>>
- Chambliss, J., Moll, E. (2022). Making Sense of Digital Humanities. Michigan State University. <<https://openbooks.lib.msu.edu/makingsensedh/>>
- Carroll SR, Garba I, Plevel R, Small-Rodriguez D, Hiratsuka VY, Hudson M and Garrison NA (2022) Using Indigenous Standards to Implement the CARE Principles: Setting Expectations through Tribal Research Codes. Front. Genet. 13:823309. doi: 10.3389/fgene.2022.823309

DALŠÍ ČTENÍ

- Davidson, J. and Ashley, K. (2005-2012). **Curation reference manual**.
- Lin, X., Khoo, M., wook Ahn, J., Tudhope, D., Binding, C., Massam, D., and Jones, H. J. (2016). **Mapping metadata to DDC classification structures for searching and browsing**. International Journal on Digital Libraries, 18:25–39.
- Park, J.-R. and Brenza, A. (2015). **Evaluation of semi-automatic metadata generation tools: A survey of the current state of the art**. Information Technology and Libraries, 34(3):22–42.
- Randtke, W. (2013). **Automated metadata creation: Possibilities and pitfalls**. The Serials Librarian, 64(1-4):267–284.
- Zhang, A. B. and Gourley, D. (2009). **4 - metadata strategy**. In Zhang, A. B. and Gourley, D., editors, Creating Digital Collections, Chandos Information Professional Series, pages 31– 53. Chandos Publishing.
- Adámek, P. (2017). **Projektový management**. Distanční studijní opora. Slezská univerzita. Obchodně podnikatelská fakulta v Karviné.
- Appleford, Simon, and Jennifer Guiliano. “Writing a Data Management Plan.” DevDH.org, 2013. <http://devdh.org/lectures/thinkingdata/datamanagement/>.