



Vybrané otázky (meta)datové heuristiky

PREZENTACE PRO DOKTORANDY ÚHV FF MU

JOSEF SCHWARZ, CIT FF MU, LINDAT/CLARIAH-CZ

15. 3. 2023

„Data isn't just numbers. It's stories, it's ideas, it's emotions, it's what makes our world and ourselves unique“.

– [HANS ROSLING](#)

Několik základních pojmů

JAK SE LIŠÍ DATA A METADATA? CO JSOU TO PARADATA?
PROČ NEEEXISTUJÍ „TVRDÁ“ DATA? A PROČ V KNIHÁCH
NENAJDEME ZNALOSTI...

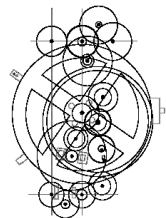
Data, informace, znalosti I

Data

- Vymezení: Zaznamenané symbolické reprezentace reálného světa nebo světů fikčních
- Médium (nosič): Analogové (hlína, kámen, papír atd.) nebo digitální
- Struktura: Data strukturovaná (seznamy, tabulky, databáze atd.) nebo nestrukturovaná (text)
- Formát: Číselný, textový, obrazový, zvukový, audiovizuální, kombinovaný

Příklady:

- [460.8](#)
- [NA1995](#)
- [Příštpo](#)



Informace

- Vymezení: data s přiděleným významem
- Médium (nosič): Analogové (hlína, kámen, papír atd.) nebo digitální
- Struktura: Informace faktografické, plnotextové, bibliografické atd.
- Formát: Číselný, textový, obrazový, zvukový, audiovizuální, kombinovaný

Příklady:

- 460,8 m.n.m.
- NA1995 Architecture as a profession
- Příštpo u Jaroměřic nad Rokytnou

Znalost

- Vymezení: informace s přiděleným kontextem
- Médium (nosič): Mozek
- Struktura: Poznatkový tezaurus
- Formát: Elektrochemický signál

Příklady:

- 460,8 m.n.m. (Říp)
- NA1995 Architecture as a profession (LCC)
- Příštpo u Jaroměřic nad Rokytnou

Data, informace, znalosti II

POZOR!!!:

- Neexistují **tzv. tvrdá data** ve smyslu nezpochybnitelných faktů, protože data jsou výsledkem výzkumu, měření a interpretace za použití specifických metrik, nástrojů a metodologie.
- Nezaměňovat **fakta a data** – fakta jsou údaje nebo tvrzení, která lze ověřit a jsou považována za pravdivá; data naproti tomu jsou pouhým záznamem bez pravdivostní hodnoty.
- Jazyková poznámka:

Termín	Singulár	Plurál
Fakta	Fakt, faktum	Fakta
Data	[neexistuje] (nebo: údaj)	Data (údaje)
Informace	Informace	Informace
Znalost	Znalost („jednotková“ znalost: poznatek)	Znalosti

Exkurz:

Sémiotický trojúhelník a [teorie tří světů Karla Poppera](#) a [Johna Ecclese](#)

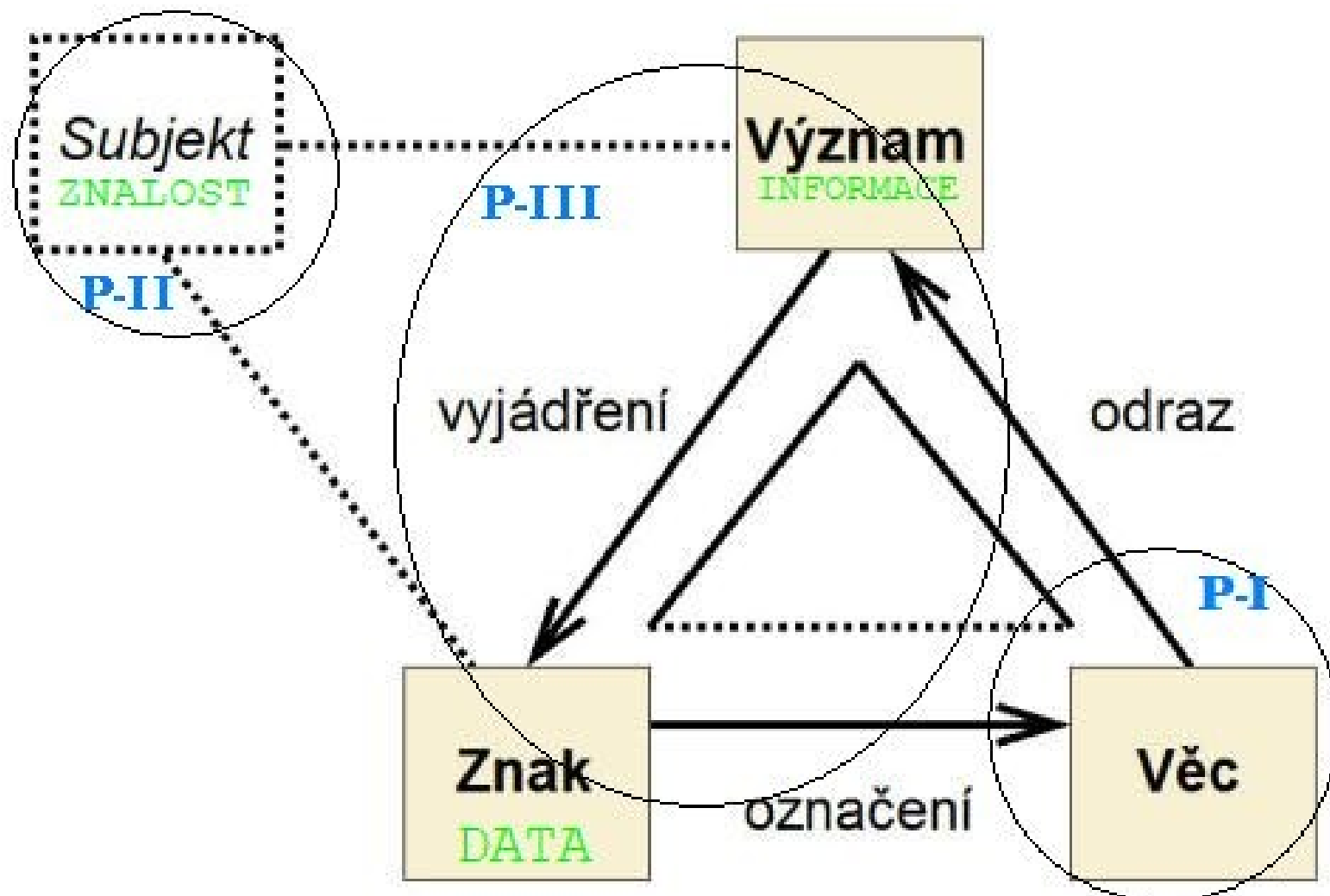
Schéma převzato z KUČEROVÁ, Helena.

Co analyzujeme při obsahové analýze dokumentů? K pojmu aboutness v organizaci znalostí. Knihovna [online]. 2014, roč. 25, č. 1, s. 36-54 [cit. 2023-03-15]. Dostupný z WWW:

<http://knihovna.nkp.cz/knihovna141/141036.htm>

upraveno.

a



Data, metadata, paradata

Data

Metadata

- Vymezení: Data o datech související s jejich obsahem, vlastnostmi, formátem a dalšími charakteristikami.
- Dělení: popisná, strukturální, technická, administrativní aj. (záleží na kritériu členění)
- Příklady:
 - 460,8 m.n.m. | Bpv
 - NA1995 Architecture as a profession | Last revision 2013/30/12
 - Příštpo u Jaroměřic nad Rokytnou | CZ0634 544752

Paradata

- Vymezení: Údaje o procesu vzniku/sběru, použití, interpretaci a kvalitě dat a jeho metodách; zajišťuje správné pochopení dat, jejich adekvátní využití a reprodukovatelnost výzkumu
 - 460,8 m.n.m. || Ověřeno vlastním měřením
 - NA1995 Architecture as a profession || Duplicitní k TU234B
 - Příštpo u Jaroměřic nad Rokytnou || Definitivní stav

Jak na (meta)data

Vědecký výzkum a data

1. Formulace výzkumného problému a cílů
2. Návrh výzkumného designu
3. **Sběr (získávání) dat**
4. **Zpracování dat (včetně tvorby metadat a paradat)**
5. **Analýza dat**
6. Prezentace výsledků výzkumu, závěry, doporučení

Kde vezmu data?

A. Kontext: [Otevřená věda \(Open Science\)](#)

B. Existující data

- Zčásti nebo zcela odpovídají mému výzkumu a jsou v něm bezprostředně využitelná
- S mým výzkumem souvisí jen zčásti, ale lze je použít např. jako metodologickou inspiraci
- Zdroje existujících dat velice různorodé, pro strukturovaná vědecká data možnost použití řady datových repozitářů, viz [sekundární i primární zdroje](#).
- [SSH Open Marketplace](#)
- Nechte si poradit [umělou inteligencí](#)

C. Vlastní průzkum a sběr v primárních zdrojích

Jaký nástroj pro zpracování dat použiji?

A. Specializovaný software

- Předpoklad: SW naplňuje všechny mé představy o tom, jak chci manipulovat se svými daty; popř. použiji sadu více SW pro různé úkony.
- Důležité podmínky: SW používá běžné standardy a umožňuje bezproblémový přenos dat (import a zejména export)
- Typické příklady: [Zotero](#) pro bibliografické informace
- Další nástroje lze nalézt na specializovaných portálech, např. [SSH Open Marketplace](#), [LINDAT/CLARIAH-CZ](#), [EOSC](#) aj.

B. Tabulkový procesor (TabP)

- Předpoklad: Žádný z existujících nástrojů nenaplňuje moje potřeby (nebo se obávám, že by se v průběhu práce objevily požadavky, které by nebyl specializovaný SW schopen splnit)

Pracujeme s TabP: Základní kroky

1. Vytvoření datového modelu
2. Stanovení datových objektů
3. Určení struktury dat a metadat včetně případných standardních metadatových formátů
4. Pravidla pro zápis jednotlivých údajů
5. Generování sestav a provádění analýz pomocí funkcí TabP

Pracujeme s TabP: Datový model

1. Jaká data potřebuji pro svůj výzkum?
2. Jak tato data budu popisovat metadata a parady tak, abych mohl s daty efektivně manipulovat?
3. Budu pracovat s neurčitostí dat a informací a jak se to projeví ve struktuře dat?
4. Použiji pro některé údaje standardizované taxonomie, řízené slovníky nebo authority?
5. Jaké očekávám výstupy ze zpracování dat, jak má vypadat prezentační vrstva dat?

Pracujeme s TabP: Datové objekty

Datový objekt = entita, ke které shromažďují data, metadata a paradata a která pro mě představuje samostatnou logickou jednotku. Datové objekty by měly být disjunktní (měly by představovat různé kategorie entit).

Realizace v TabP: jeden datový objekt = jeden list

Příklady datových objektů:

- Paměťové instituce
- Archivní fondy
- Osoby
- Počítačové hry
- Archivační softwary

Pracujeme s TabP: Struktura dat

1. Určení jednotlivých údajů, které budu na úrovni dat, metadat a paradat sledovat (včetně různých atributů, např. v jakém stavu se nachází jednotlivý záznam: rozpracovaný, hotový, definitivní)
2. Stanovení **granularity** údajů, tj. míry specifičnosti jednotlivých dat
3. Použiji nějaký standardizovaný metadatový formát, ať už přímo, nebo jako inspiraci?
 - Příklad: [VRA Core](#)

Realizace v TabP: jednotlivý údaj = jeden sloupec tabulky

Doporučení: označování jednotlivých záznamů barvami se sémantickou funkcí (popř. jiným grafickým zvýrazněním, např. kurzivou, velikostí písma atd.) je vhodné jen jako vizualizační pomůcka, nikoli pro analytickou práci s daty

Pracujeme s TabP: Pravidla pro zápis dat

1. Jak budu tvořit záznam instance a vyplňovat jednotlivé údaje? Jakými pravidly se budu řídit? Jak budu zapisovat opakovatelné údaje?
2. Použiji pro zápis dat nějaký existující standard?
 - Příklady: [CDWA – Categories for the Description of Work of Arts](#), [Cataloging Cultural Objects](#)
3. Mohu vyplňování dat nějak poloautomatizovat?
4. Budu používat importy dat? Jaký si stanovím unikátní oddělovač údajů?

Realizace v TabP: jeden záznam instance = jeden řádek

Pracujeme s TabP: Funkce

1. Filtry
2. Řazení
3. Kontingenční tabulky
4. Grafy

Děkuji za pozornost
a přeji úspěšnou
práci s daty!
