

Digital humanities

Kulturní data

Jindřich Marek

Co je digitální obsah?

- kulturní data
 - přes 50 milionů digitalizovaných knih (2024)
 - Europeana: 60+ milionů kulturních objektů
 - Internet Archive: 40+ milionů textů
- texty, obrazy, audio, video, interaktivní média

Jaký je rozdíl mezi naskenovanou knihou na webu a digitálním obsahem z hlediska DH?

Data v DH

- strojově zpracovatelná data
- strukturovaná metadata
- možnost analýzy a vizualizace
- interaktivita a propojitelnost

Digitální obsah: dostupnost

- online platformy, databáze a projekty
- materiály z různých institucí a regionů: demokratizace přístupu
- bohatý a dynamický zdroj kulturních dat, který umožňuje výzkumníkům zkoumat, analyzovat a interpretovat lidské dějiny a společnost prostřednictvím inovativních výpočetních metod

Typy digitálního obsahu

- texty
- vizuální a multimediální obsah

Texty

- digitalizované knihy, rukopisy, dopisy a historické dokumenty
 - pochopení a interpretace různých aspektů lidské kultury a společnosti
 - analýzy s využitím výpočetních metod k odhalení vzorců, trendů a poznatků, které nemusí být okamžitě zřejmé při tradičním zkoumání
 - textové korpusy umožňují aplikovat techniky zpracování přirozeného jazyka

Vizuální a multimediální obsah

- obrazy, fotografie, umělecká díla, audiovizuální materiály
 - zkoumání kultury, ikonografie a vlivu obrazů na historické a kulturní narativy
 - analýza vizuálních dat: rozpoznávání obrazů, analýza výrazu tváře, studium vizuální estetiky

Klíčové koncepty

Kulturní dědictví

- minulé hmotné artefakty a nehmotné prvky nějaké skupiny nebo společnosti zděděné od předchozích generací

Dokumentární dědictví (definice UNESCO)

- dokument
 - něco, co je zapsáno s intelektuálním záměrem
 - rozlišení intelektuálního obsahu a nosiče, chráněno obojí
- dokumenty mohou být symboly a obsahovat rysy relevantní pro kolektivní paměť komunity, národa, regionu nebo společnosti

Kde najít digitální obsah?

Digitální knihovny (příklady)

World Digital Library

<https://www.loc.gov/collections/world-digital-library/about-this-collection/>

kurátorský výběr

multikulturní perspektiva

Original Format	
Photo, Print, Drawing	13,305
Book/Printed Material	4,038
Map	1,808
Manuscript/Mixed Material	175
Newspaper	112
Periodical	45
Film, Video	11
Notated Music	11
Audio Recording	9
Software, E-Resource	1
Web Page	1
Online Format	
Image	19,491
Compressed Data	5,070
PDF	4,323
Online Text	690
Video	11
Audio	9

COLLECTION

World Digital Library

About this Collection

[Collection Items](#)

Featured Content



Morning Glory Flowers:
National Diet Library



The Origin of Tenjin:
National Diet Library



Annals of Creation: National
Library of China



One Hundred Aspects of
the Moon: Moon and
Smoke: National Diet ...



Itinerary Book Kept During
the Journey to East India:
National ...



About this Collection

[Rights and Access](#)

Expert Resources

[Main Reading Room](#)

[General Collections](#)

[Ask a Librarian](#)

[African & Middle Eastern Reading Room](#)

[Ask a Librarian](#)

[Listen to this page](#)

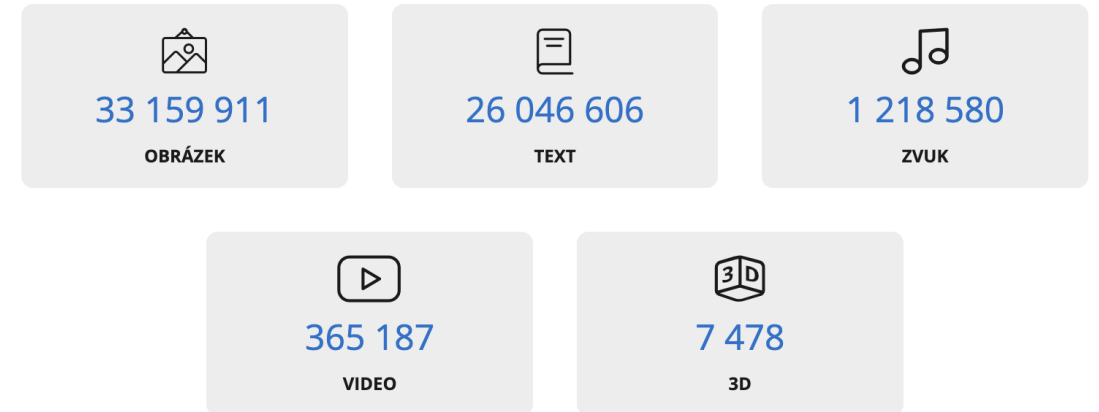
About this Collection

This collection contains cultural heritage materials gathered during the World Digital Library (WDL) project, including [thousands of items contributed by partner organizations](#) worldwide as well as content from Library of Congress collections. The original World Digital Library site (preserved in [LC's Web Archives here](#)) and all descriptive metadata were translated from English and made available in six additional languages: [Spanish](#), [Portuguese](#), [French](#), [Arabic](#), [Russian](#), and [Chinese](#). All item records include narrative descriptions submitted by the contributing partners and enhanced by WDL researchers to contextualize the item and its cultural and historical importance. Books, manuscripts, maps, and other primary materials in the WDL collection are presented in their original languages; more than 100 languages are represented, including many lesser known and endangered languages.

Europeana (Collections)

<https://www.europeana.eu/cs>

agregátor
standardizace metadat



Sbírky

Prozkoumejte naši sbírku podle tématu, tématu, století a organizace.

 SDÍLEJTE

Témata



Migrace

Lidé se vždy stěhovali z jednoho místa na druhé. Objevte příběhy slavných migrantů a cesty a životy obyčejných lidí do Evropy, z Evropy ...



Móda

Prozkoumejte historickou a současnou módu a její vliv na naše oblečení, obuv, doplňky a styl. Objevte náčrtky, návrhy a oděvy,...



Noviny

Prozkoumejte tištěné publikace od roku 1618 do 80. let 20. století. Objevte titulky, fulltextové články, reklamy a autory, kteří za nimi stojí.



Přírodopis

Prozkoumejte studium zvířat a rostlin prostřednictvím kreseb, fotografií a textů exemplářů a biologů a botaniků, kteří je studovali.



Jak pracovat s daty?

Získávání dat

- stažení archivu
- přístup přes API
- web scraping

Využití dat a metadat ke knihám

- prostá prezentace digitalizovaného obsahu
 - i úzce specializované projekty
- další využití dat a metadat (strojové zpracování) a interpretace
 - různé významy: strojové zpracování dat a metadat
 - důležitá je možnost uložit si vlastní dataset

Data ke knihám v digitálních knihovnách

- data
 - obrazy stran
 - JPG, TIFF, JPEG2000
- metadata
 - popisy
 - rukopisy: TEI, tisky: MARC21 - MODS
 - data související s digitalizací, data pro digitální ochranu, kontejnerové formáty
 - PREMIS, METS

Postup při analýze kulturních dat

1. Získání dat

- API
 - např. Europeana REST API
- web scraping
 - pozor na podmínky služeb a přetěžování serverů
- export z databází
 - nejčastější formáty: CSV, JSON, XML

2. Čištění dat

- cíl: přehledná datová struktura
 - „tidy data“
- prostředky
 - rozdělení/spojení buněk
 - řešení chybějících hodnot
 - sjednocení údajů
 - oprava chyb

Tidy data



A data set is **tidy** iff:

1. Each **variable** is in its own **column**
2. Each **case** is in its own **row**
3. Each **value** is in its own **cell**

3. Explorativní analýza

- základní statistika
 - počty, distribuce
- vizualizace
 - časová řada, geografická mapa
- příp. identifikace anomálií

4. Interpretace

- výzkumná otázka: co data říkají o historickém/kulturním jevu?
 - vždy vzhledem ke kontextu

Výzkumné nástroje (příklady)

- umožňují zpracování a zobrazení digitálních dat (včetně těch kontextuálních)
- interaktivita
- analytické, často mapové zobrazení

Inkunábule: místa tisku

- také univerzity, obchodní cesty, konflikty
- <http://atlas.lib.uiowa.edu/>
- DH metoda: geoprostorová analýza

THE ATLAS OF EARLY PRINTING

The Map

About the Atlas ▾

About Early Printing ▾

①



1450

1470



Date Span:

1450 to 1470

☒ Spread of Printing

Data collated from digital and print sources. [Improve this data.](#)

☐ Output by Location

☐ Typography

☐ Ecclesiastical Borders

☐ Political Borders

☐ Trade Routes

☐ Bishoprics

☒ Paper Mills

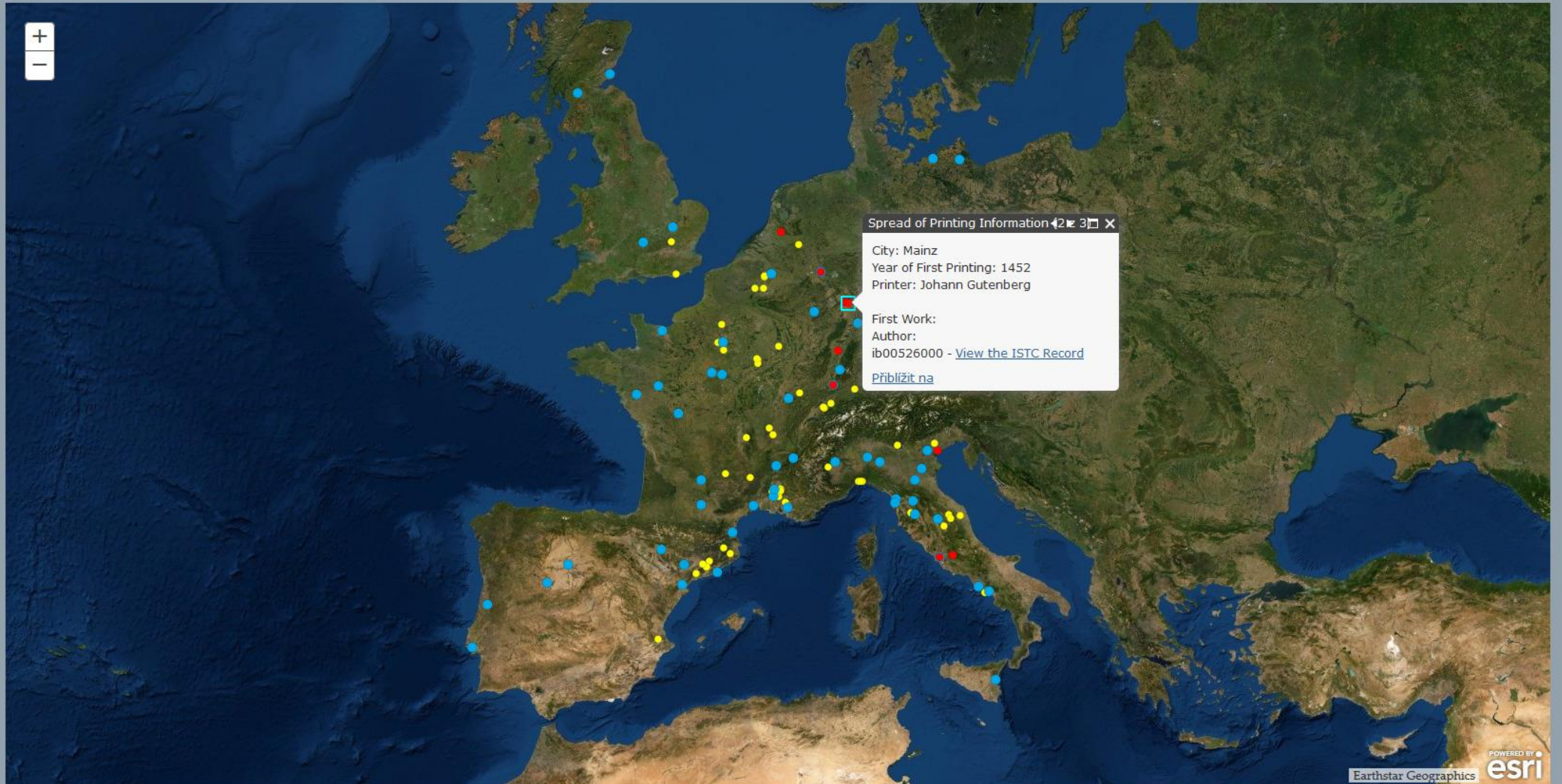
Data collated from digital and print sources. [Improve this data.](#)

☒ Universities

Data collated from digital and print sources. [Improve this data.](#)

☐ Fairs

☐ Conflicts



THE ATLAS OF EARLY PRINTING

The Map

About the Atlas ▾

About Early Printing ▾

①



1475

1500



Date Span:

1483 to 1500

☒ Spread of Printing

Data collated from digital and print sources. [Improve this data.](#)

☐ Output by Location

☐ Typography

☐ Ecclesiastical Borders

☐ Political Borders

☐ Trade Routes

☐ Bishoprics

☒ Paper Mills

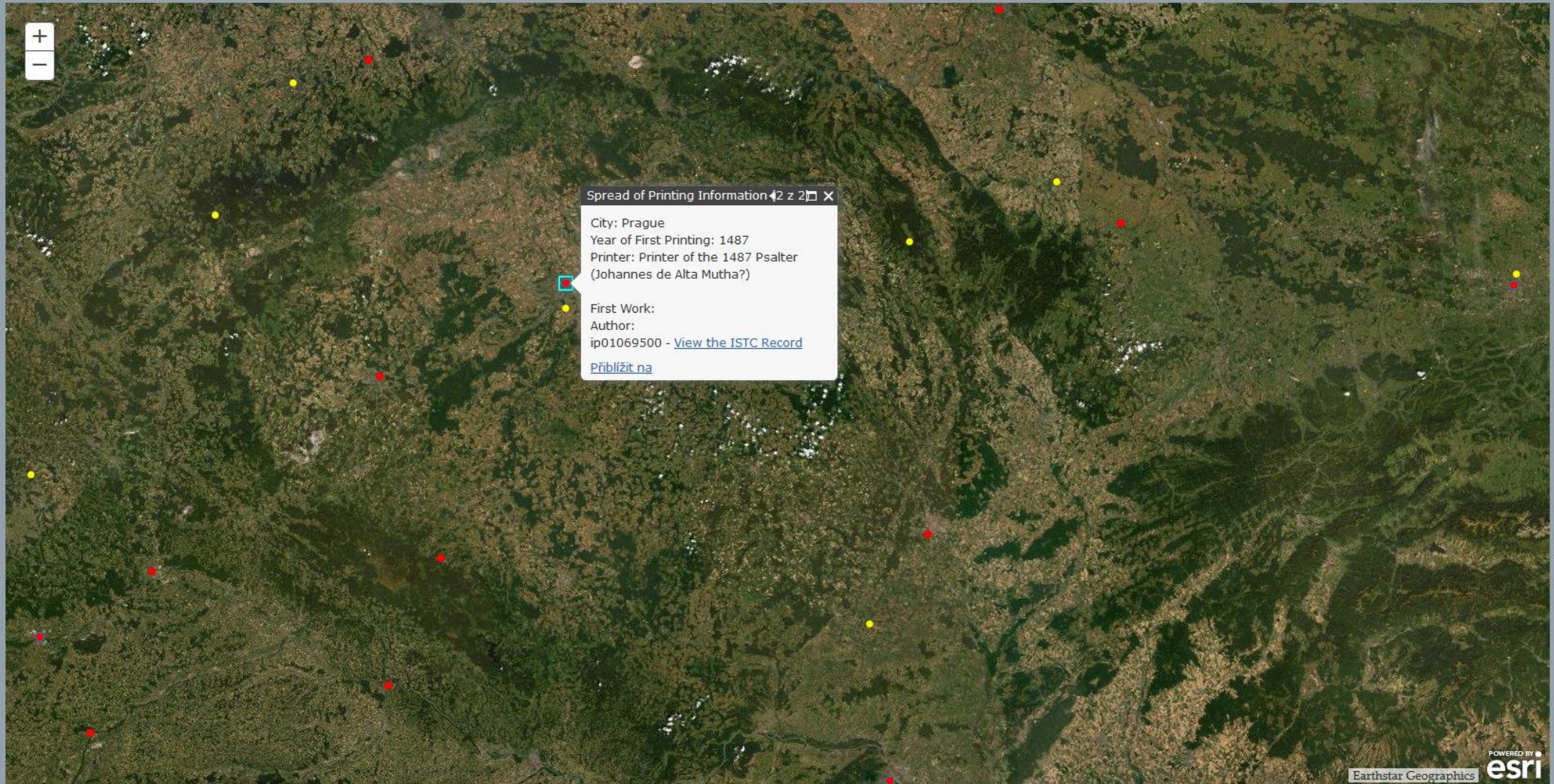
Data collated from digital and print sources. [Improve this data.](#)

☒ Universities

Data collated from digital and print sources. [Improve this data.](#)

☐ Fairs

☐ Conflicts



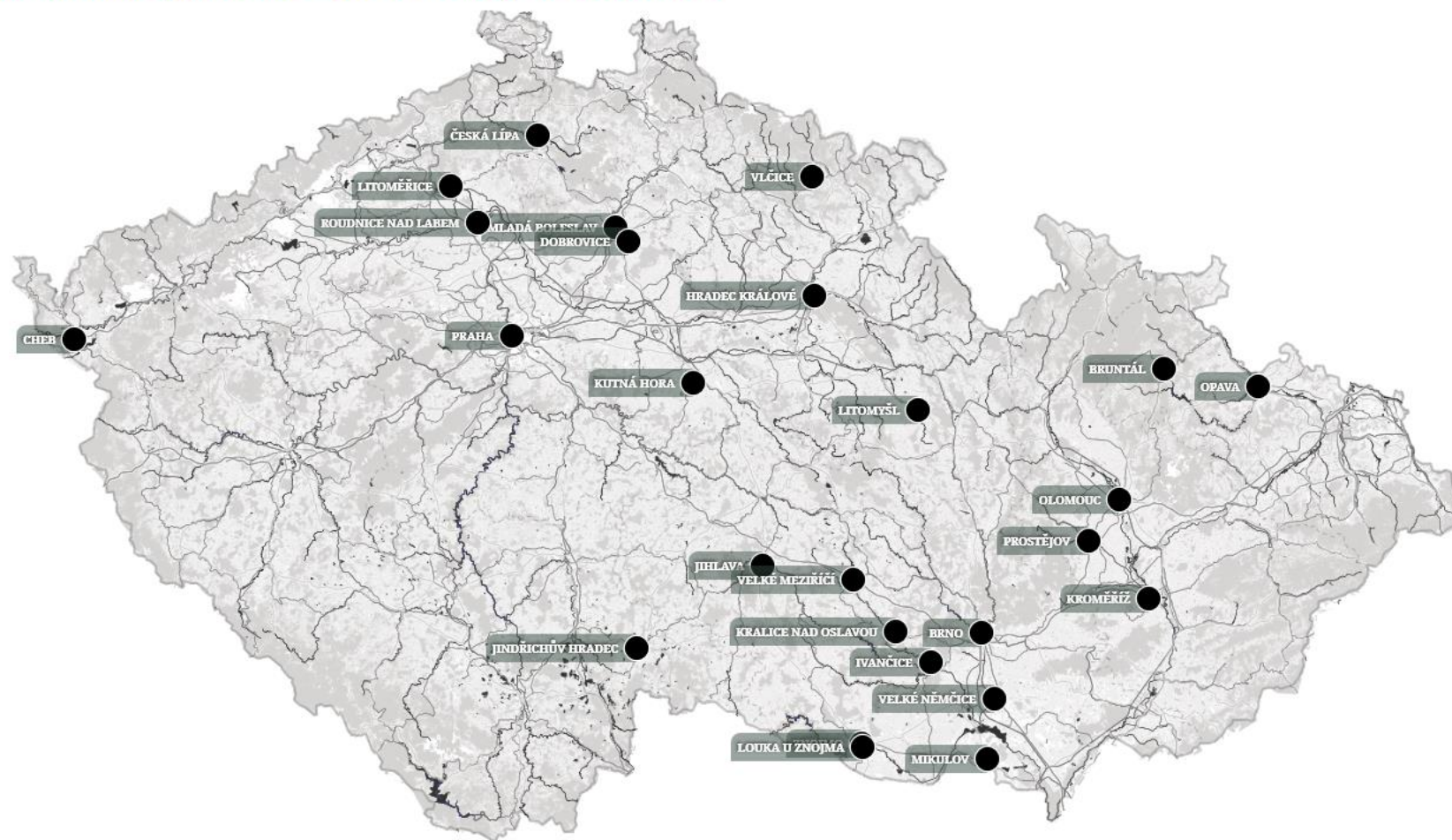
Earthstar Geographics

POWERED BY
esri

Knihověda.cz

- <https://www.knihoveda.cz/>
 - integrované databáze tiskařské produkce s odkazy na digitalizáty
 - encyklopedie knihy
 - nástroj: **mapa tiskařské produkce**

Mapa tiskařské produkce Čech a Moravy do roku 1800



12345

Celkem 10 296 vydání

Místo26

Vybrat

Tiskař222

Vybrat

Téma44

Vybrat

Žánr48

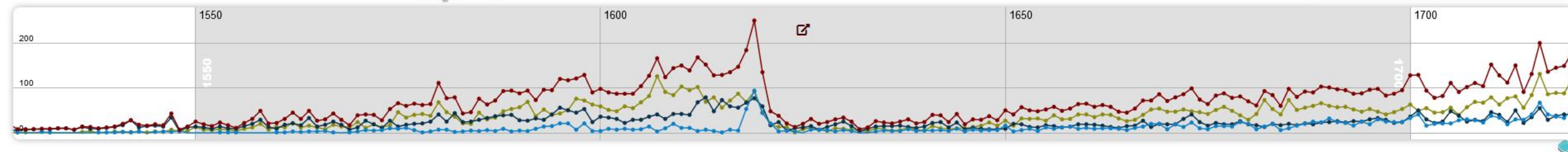
Vybrat

Jazyk20

Vybrat

Časový interval

Od1550do1700



Kritická reflexe a závěry

Jak pracovat s daty

- povaha dat
 - data bývají neúplná, nekonzistentní a obsahují chyby
 - tj. klíčová je příprava dat (čištění, transformace)
- kontext je vše
 - data sama o sobě nic neříkají
 - pro interpretaci a kladení otázek je třeba znát kontext
- doporučené nástroje
 - OpenRefine pro explorativní analýzu a čištění dat
 - Python pro zpracování a vizualizaci dat (knihovny: pandas, plotly apod.)

Závěry

- ne všechno, co na je k humanitním vědám na webu, je možné označit za přístup digital humanities
- často se jedná o tradiční výstupy prezentované v elektronickém prostředí
- přístup založený na digital humanities zpravidla prezentuje velké množství dat, které je možné různě filtrovat a vizualizovat
- přidanou hodnotou digital humanities je zejména interaktivita

Literatura

- Huhtamo, Erkki, ed. a Parikka, Jussi, ed. Media archaeology: approaches, applications, and implications. Berkeley: University of California Press, 2011.
- [Video] [The Story of Python and how it took over the world | Python: The Documentary](#)