

LINDAT/CLARIAH-CZ

Digital Research Infrastructure for Language Technologies, Arts and Humanities

Czech node of the CLARIN ERIC and DARIAH ERIC networks

Jan Hajič

<https://ufal.mff.cuni.cz/jan-hajic>

Charles University, Prague

Faculty of Mathematics and Physics

Computer Science School

Institute of Formal and Applied Linguistics

Malostranské nám. 25, 118 00 Prague 1

Czech Republic

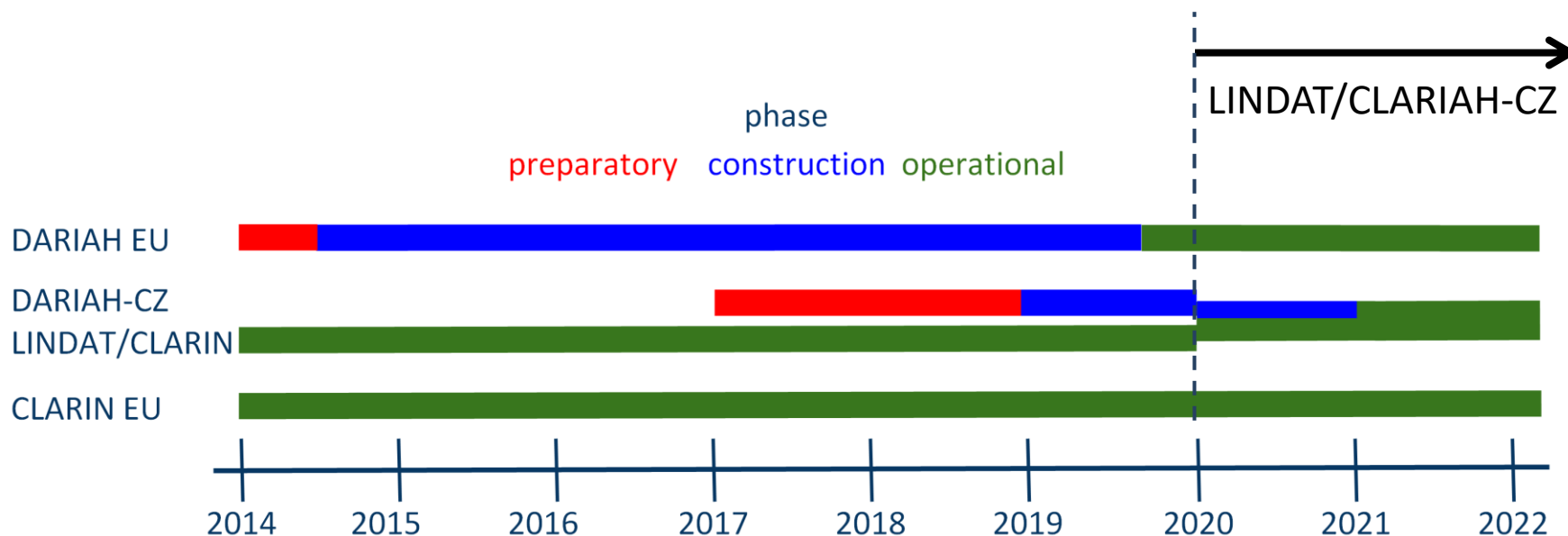
- Language Technology
 - Language Resources (open repository)
 - Natural Language Processing tools and services
- “Digital humanities”
 - Enable & support digitally-enabled research
- Users
 - Researchers in Language Technology Arts and Humanities
 - General public („Internetová jazyková příručka“ at ICL AS CR)
- Technology
 - Repository (resources), Services, (Software) Tools, Applications
- Current status
 - LINDAT/CLARIAH-CZ, joining CLARIN and DARIAH in CZ
 - Started Jan. 1, 2019 in parallel with LINDAT/CLARIN

Main Goals for 2019-2022 LINDAT



- Continue CLARIN ERIC participation, become stable part of DARIAH ERIC
 - Membership of Czech Republic in DARIAH ERIC – *approved Nov. 7, 2019*
 - Integrated with LINDAT/CLARIN by Jan. 1, 2020
 - Operational and fully connected by Jan. 1, 2021
- Construction phase (2019-2020), following preparations (2017-8)
 - Collect data, assess quality, completeness, licensing, determine main repository(ies)
 - Prioritize, convert data / metadata
- Operational phase (2021-)
 - Repositories running and connected to DARIAH and CLARIN central hub(s)
 - Services
 - integrated language services, new services: search in multimodal data, statistics, visualization
- Continued operation for the CLARIN part, integration to EOSC (SSHOC project)

Timeline of LINDAT/CLARIAH-CZ w/relation to LINDAT/CLARIN



- LINDAT/CLARIN est. 2010, operational since Jan. 1, 2014
- CLARIN ERIC est. 2012 (Utrecht, NL)
- DARIAH ERIC est. 2014 (Paris, FR)

Institutions



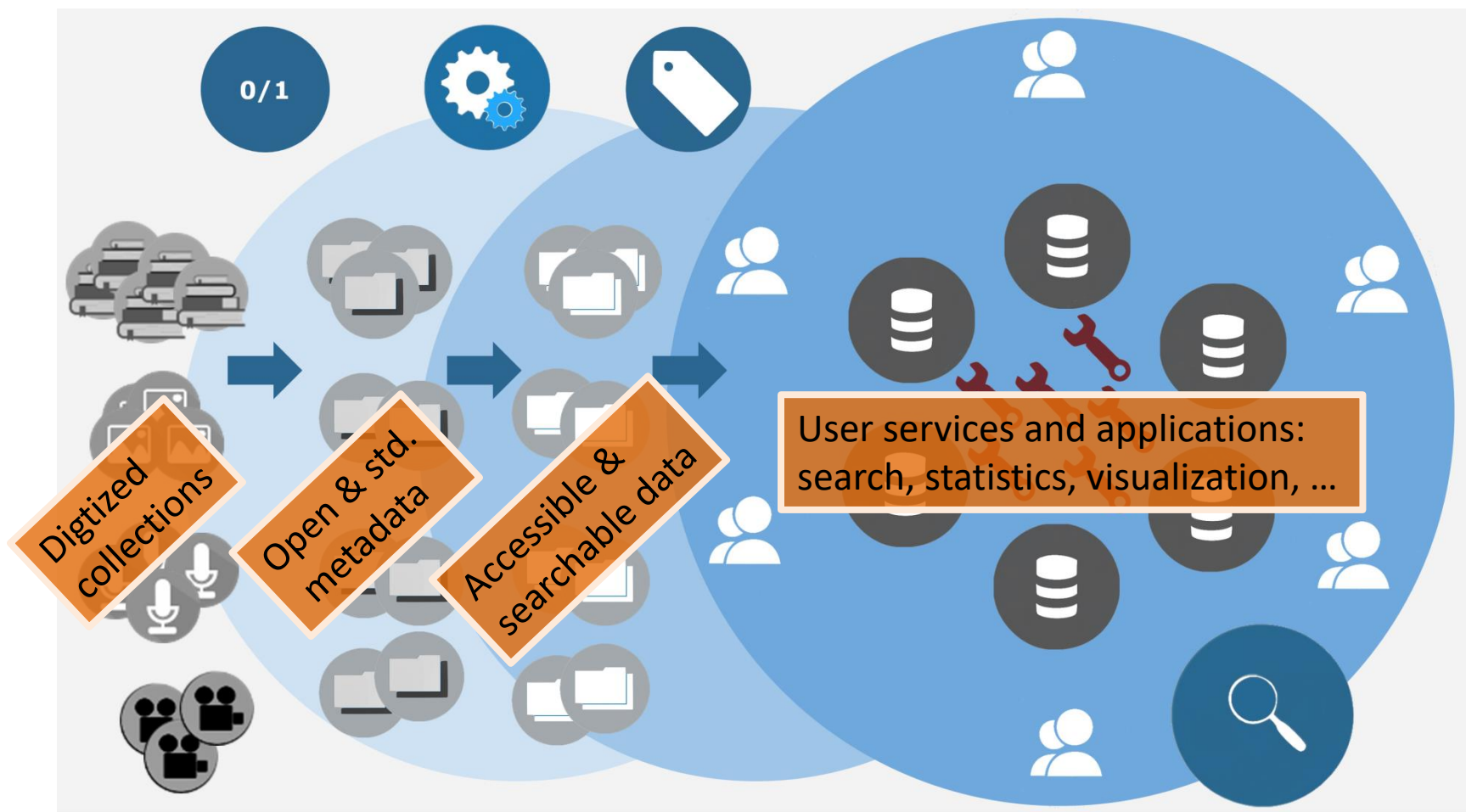
- Charles University, Prague
 - Faculty of Mathematics and Physics (CS School)
 - Faculty of Arts (multiple departments)
- Univ. of West Bohemia, Pilsen (Applied Sciences)
- Masaryk University, Brno
 - Faculty of Informatics, Faculty of Arts (multiple departments)
- Academy of Sciences of the Czech Republic
 - Library of the Academy of Sciences
 - Institute of the Czech Language
 - Institute of Philosophy
 - Institute of History
- National Library, Prague
- Moravian Library, Brno
- National Film Archive, Prague
- National Gallery, Prague

Research Fields



- Linguistics, computational linguistics
- History, historical bibliography
- Study of culture
- History of art
- Philosophy
- Film and film history including new audiovisual media
- Visual arts
- Musicology and history of music
- Ethnology
- Archeology/Egyptology
- ... and interdisciplinary areas of the above

Workplan scheme



Example Services: Machine Translation



LINDAT
CLARIAH

Repository

Corpus Search

TreeQuery

TreeX

More Apps

About

CLARIN

LINDAT Translation

Translate

Docs

You'll be shown the status (started/queued) of your translation task above this panel.

Source

Czech

☐ advanced

Input sentences

Translate

Choose file

Target

English

English

French

German

Hindi

Russian

Credits: The service runs systems trained by Martin Popel ([en-cs](#); [cs-en](#); en-fr; fr-en), Shantipriya Parida (en-hi) and Dušan Variš (en-ru, ru-en, en-de, de-en).

Example Services: Machine Translation



Repository

Corpus Search

TreeQuery

TreeX

More Apps

About

LINDAT Translation

Success

Translate

Docs

You'll be shown the status (started/queued) of your translation task above this panel.

Source

Czech

Target

English

☐ advanced

Input sentences

Meč datovaný do mladší doby bronzové je zdobený jednoduchou rytou linií obíhající kolem ostří, které je stále ostré jako břitva, konstatovala archeoložka rychnovského muzea Martina Beková.
„Bronzový meč s jazykovitou rukojetí je datovaný do období kolem roku 1200 před naším letopočtem, náleží kultuře lužické. Nálezy této kultury jsou ve východních Čechách četné, ale neplatí to o mečích,“ řekla Beková.
Podle ní v posledních desetiletích nálezů pravěkých mečů není v celé ČR více než pět.
Přesné místo objevu archeologové tají z důvodu ochrany lokality.

Translation

The sword, dated to a younger Bronze Age, is decorated with a simple engraved line orbiting a blade that is still razor-sharp, Rychnov Museum archaeologist Martina Bek noted.
"The bronze sword with the tongue-like handle is dated to around 1200 B.C., belonging to a culture of debt. Findings of this culture are numerous in Eastern Bohemia, but this is not true of swords," Bekova said.
According to her, in recent decades there are no more than five finds of ancient swords in the entire Czech Republic.
The exact location of the discovery is being kept secret by archaeologists to protect the site.

Translate

Choose file

Credits: The service runs systems trained by Martin Popel (en-cs; cs-en; en-fr; fr-en), Shantipriya Parida (en-hi) and Dušan Variš (en-ru, ru-en, en-de, de-en).

Example Services: Language Analysis

Input Text

Meč datovaný do mladší doby bronzové je zdobený jednoduchou rytou linií.

Process Input

Output Text

Show Table

Show Trees

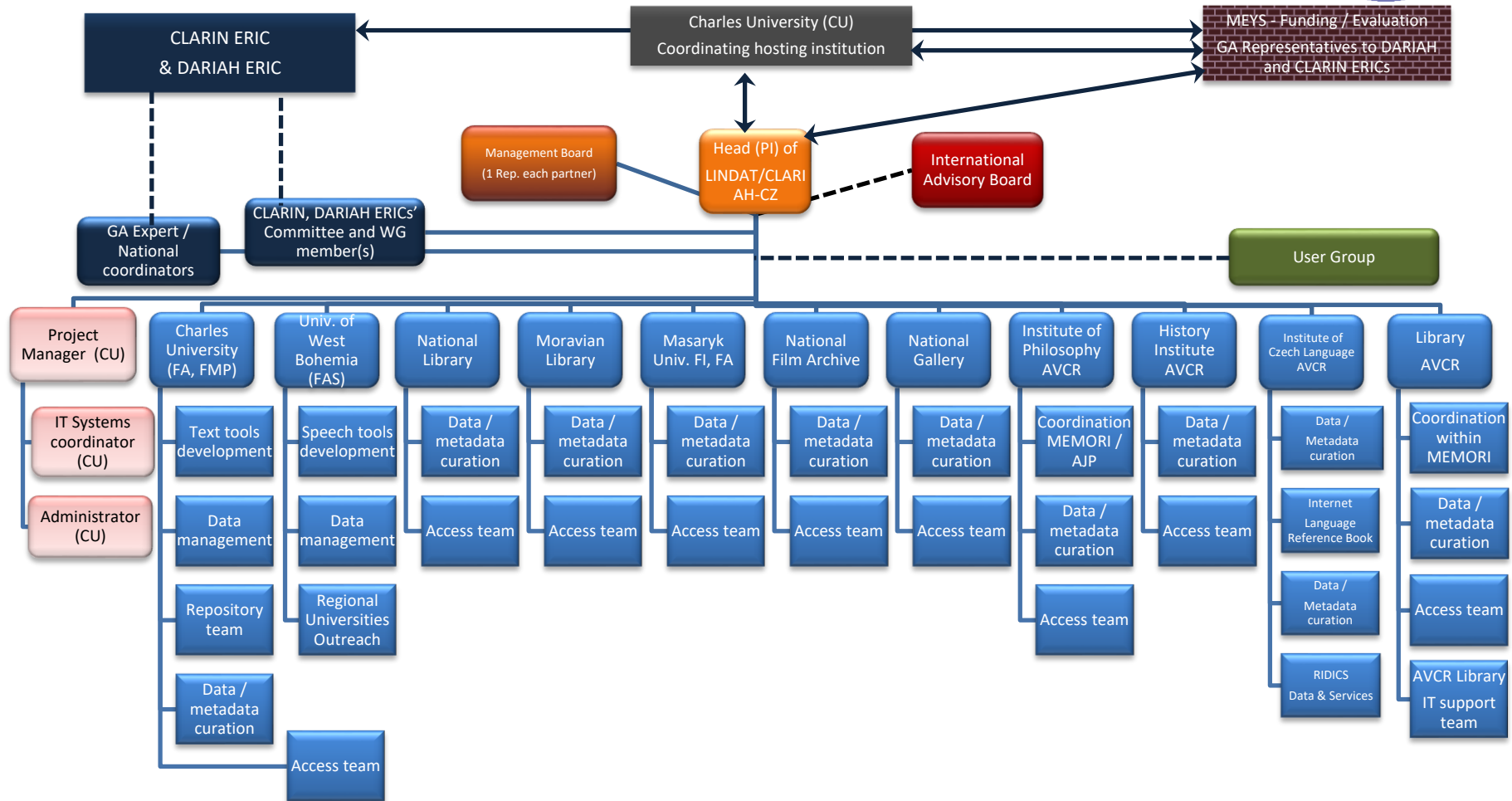
Save Tree as SVG

Meč datovaný do mladší doby bronzové je zdobený jednoduchou rytou linií .

```
graph TD
    root["<root>"] --> z["zdobený<br/>root<br/>ADJ"]
    root --> m["Meč<br/>nsubj<br/>NOUN"]
    z --> l1["linií<br/>obl<br/>NOUN"]
    z --> j["je<br/>cop<br/>AUX"]
    z --> r["rytou<br/>amod<br/>ADJ"]
    z --> p["punct<br/>PUNCT"]
    m --> d["datovaný<br/>amod<br/>ADJ"]
    d --> d1["doby<br/>obl<br/>NOUN"]
    d1 --> do["do<br/>case<br/>ADP"]
    d1 --> ml["mladší<br/>amod<br/>ADJ"]
    d1 --> br["bronzové<br/>amod<br/>ADJ"]
```

LINDAT/CLARIAH-CZ

Management Structure



Thank you!

<https://clariah.lindat.cz>