

# LINDAT-Clarín Centre in the Czech Republic

Jan Hajič

UFAL MFF UK

Charles University in Prague

*LINDAT-Clarín Project principal investigator*



ZÁPADOČESKÁ  
UNIVERZITA  
V PLZNI



LM2010013

Supported by



# LINDAT-Clarín



- Infrastructure for Language Research in...
  - Humanities (linguistics and related)
  - Computational Linguistics
- Provides:
  - Data repository, linked to European networks
    - Clarín ERIC, open to META-SHARE
  - National Center for language data; funding to 2015 (4.7ME)
    - Goal: Centre A+B type
    - Acquisition, specification, licensing
    - Linguistic and language-related annotation
    - Archiving, preservation, distribution
    - Masaryk Univ. Brno, Univ. of West Bohemia Pilsen, UJC Prague



ZÁPADOČESKÁ  
UNIVERZITA  
V PLZNI



Clarín ERIC 1st meeting Oct. 26, 2012

LM2010013  
Supported by



# LINDAT-Clarín Clarín Centre



- Data-related services
  - Repository (dSpace, now v. 1.8); access and authentication:
    - Single sign-on (SAML2/Shibboleth) within Clarín ERIC
      - » 2011/2: problematic in several respects; local registration added
    - EduGAIN: UK (LINDAT) IdP working, SP just opted in. Other (17) Czech IdPs encouraged to opt-in
    - Permissions, *licensing*
      - Persistency, archiving, metadata API (Clarín ERIC, also META-SHARE)
        - Persistent ID assignment (EPIC handles)
- Language services
  - PML-TQ Search (see the demo later)
  - Morphology, tagging, parsing, anaphora resolution, ...
  - Plans: Clarín ERIC compatible API (Weblicht), workflows,
  - Manatee/Bonito (NoSketchEngine) with FCS for large corpora



ZÁPADOČESKÁ  
UNIVERZITA  
V PLZNI



Clarín ERIC 1st meeting Oct. 26, 2012

LM2010013  
Supported by



# Some links...

- LINDAT-Clarín
  - <http://www.clarin.cz>
- Repository
  - <http://ufal-point.mff.cuni.cz>
- LINDAT lexical resources
  - VALLEX 2.6: <http://ufal.mff.cuni.cz/vallex/2.6>
  - PDT-Vallex 2.5+PDT: <http://ufal.mff.cuni.cz/lindat/PDT-Vallex.html>
  - Dialogs: <http://ujc.dialogy.cz>,  
<http://ufal.mff.cuni.cz/~toman/pdtsc>,  
<http://ufal.mff.cuni.cz/~toman/pdtse>
  - Treebanks info / browser (example)
    - PEDT <http://ufal.mff.cuni.cz/pedt2.0>
    - PCEDT <http://ufal.mff.cuni.cz/pcedt2.0>

# LINDAT-Clarín



<http://www.lindat.cz>

Thank you for your attention!



ZÁPADOČESKÁ  
UNIVERZITA  
V PLZNI



Clarín ERIC 1st meeting Oct. 26, 2012

LM2010013  
Supported by

