CLARIN-DSpace repozitář v LINDAT/CLARIN

Straňák, Pavel; Košarko, Ondřej; Mišutka, Jozef
2019

Konference o šedé literatuře a repozitářích, Praha, 17. 10. 2019

LINDAT/CLARIN

FAIR REPOSITORY FOR LANGUAGE DATA

PAVEL STRAŇÁK, ONDŘEJ KOŠARKO, JOZEF MIŠUTKA

LINDAT/CLARIN

LINDAT/CLARIN

LINDAT/CLARIN

LINguistic

LINDAT/CLARIN

LINguistic

DATa

LINDAT/CLARIN

LINguistic

DATa ... very broadly

LINDAT/CLARIN

LINguistic

DATa

/

CLARIN

LINDAT/CLARIN

LINguistic

DATa

/

Common

LAnguage

Research and technology

INfrastructure

LINDAT/CLARIN

- Czech national project; node of CLARIN ERIC
- Operational since 2014
- Users:
  - Researchers in SSH and Computational Linguistics
- Technology:
  - Repository (resources), Services, Applications
- Knowledge, Support and Training

LANGUAGE TECHNOLOGY

- Natural Language Processing - NLP
  - Analysis, synthesis of spoken and written language
  - Machine Translation, Information Extraction, ...
  - Search in texts, audio, video, images

- State-of-the-art technology in NLP
  - “Statistical” methods:
    - Machine learning incl. neural networks
    - Need for (large) Language Resources – Texts, multimodal
      - Repositories, identification, replication of experiments, standards

USERS

- Everyone
  - communicates in and works with natural language!
- ... immediate users of the infrastructure:
  - Language Technology researchers
    - Universities, Research organisations
      - Need lots of data, easy to get, clean open licensing
- “Content” users:
  - Linguists, historians, teachers, psychologists, sociologists, ...
    - Need identifiable data, preprocessed, searchable, easy-to-use services and applications

LINDAT CLARIN

Data Repository

PRESERVE AND FIND LANGUAGE DATA AND NLP TOOLS

Advanced Search

Author

- Hajič, Jan (47)
- Žabokrtský, Zdeněk (32)
- Straka, Milan (29)
- Zeman, Daniel (29)
- Bojar, Ondřej (28)
- ... View More

Subject

- Germanistik (47)
- machine translation (39)
- corpus (34)
- treebank (30)
- morphology (26)
- ... View More

Language (ISO)

- English (222)
- Czech (192)
- German (159)
- Dutch (92)
- Spanish (83)
- ... View More

What's New

ToolService — LINDAT / CLARIN

CorpusExplorer

Author(s):
Rüdiger, Jan Oliver

Description:

Software for corpus linguists and text/data mining enthusiasts. The CorpusExplorer combines over 45 interactive

What can you do?

Browse

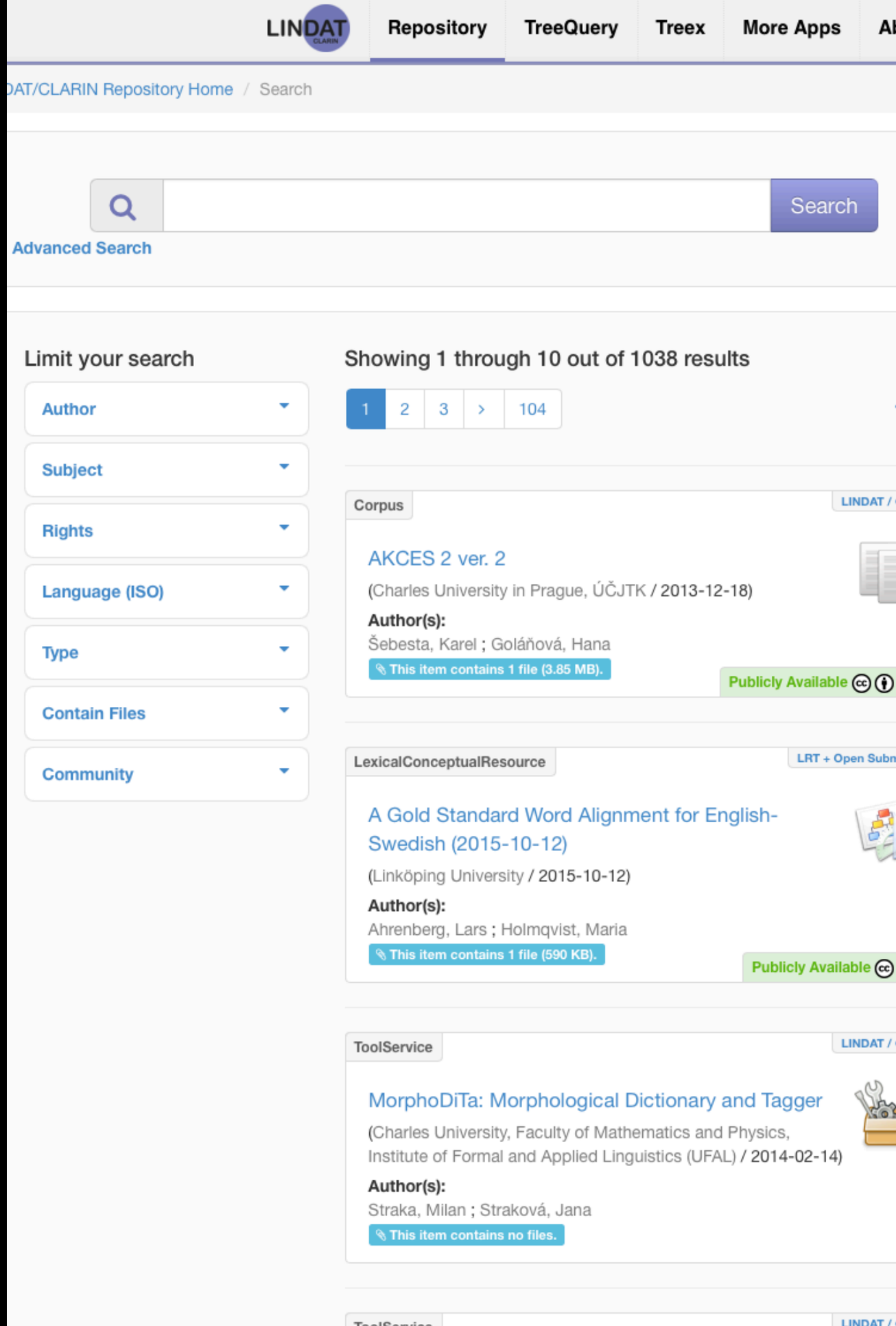
› All of the Repository

DATA REPOSITORY

- **OPEN ACCESS** (whatever can be – Public License Selector)
- > 500 registered users
  - submitters & users signing licenses (not everything can be OA)
- 200+ Data Records
  - > 1000 Metadata Records
  - 80 languages
- 100 TB+ Data in Repository (+ 1PB of UCS Shoah Foundation Archive)

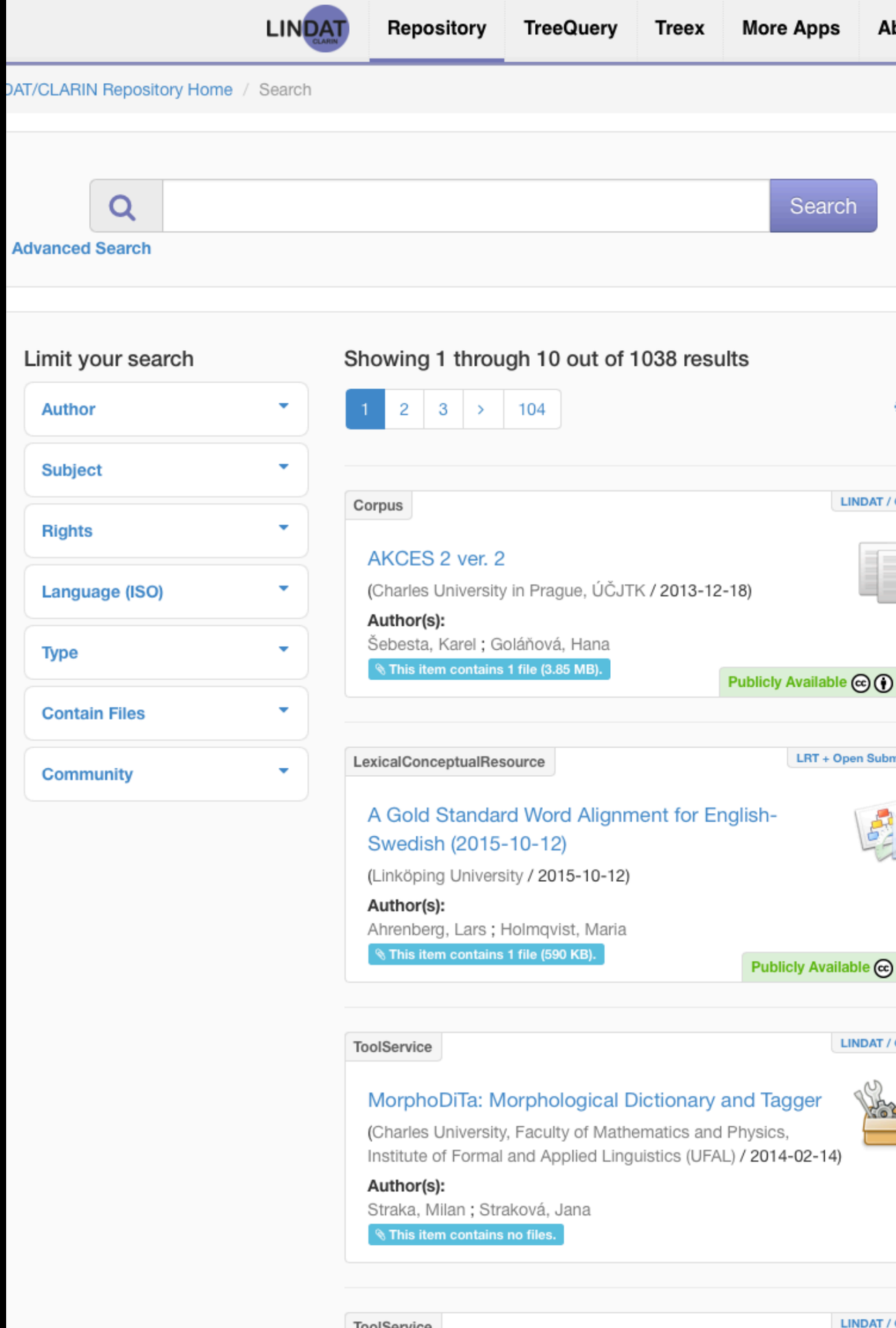
DATA REPOSITORY

- > 500 registered users
  - submitters & users signing licenses (not everything can be Open Access)
- 200+ Data Records
  - > 1000 Metadata Records
  - 80 languages
- 100 TB+ Data in Repository (+ 1PB of UCS Shoah Foundation Archive)



DATA REPOSITORY

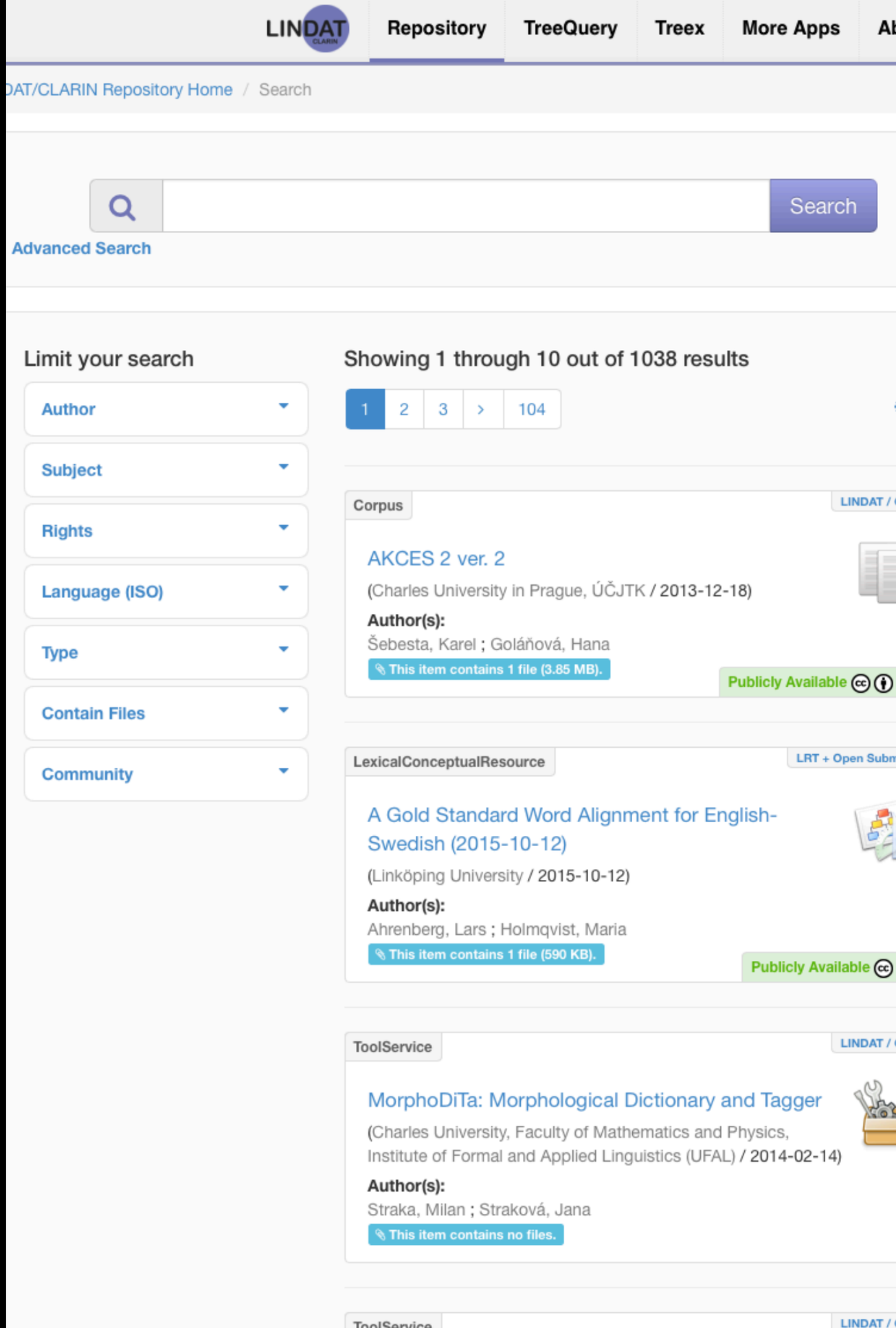
- Safe preservation
  (upload and don't worry)
- Discovery & Reuse
- Direct data citation
  (works in Google Scholar)
- Licensing
  (Open Access, but also more options)
- Versioning
- Language data and tools
- Worldwide (for everyone), easy to use



DATA REPOSITORY

- Safe preservation
  (upload and don't worry)
- Discovery & Reuse
- Direct data citation
  (works in Google Scholar)
- Licensing
  (Open Access, but also more options)
- Versioning
- Language data and tools
- Worldwide (for everyone), easy to use

OPEN ACCESS FORCE11 The Future of Research Communications and e-Shcolarship GO FAIR OpenAIRE



SAFE PRESERVATION

UPLOAD AND DON'T WORRY

How to Deposit

Only authenticated users can deposit items. If you cannot find your home organisation in the Login dialog list of organisations then register at clarin.eu and authenticate using "clarin.eu website account". In case you cannot use any authentication method above or if you encounter a problem, do not hesitate to contact our Help Desk and we can create a local account for you.

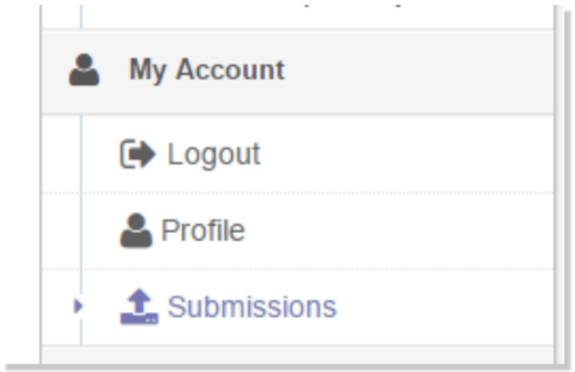
Step 1: Login

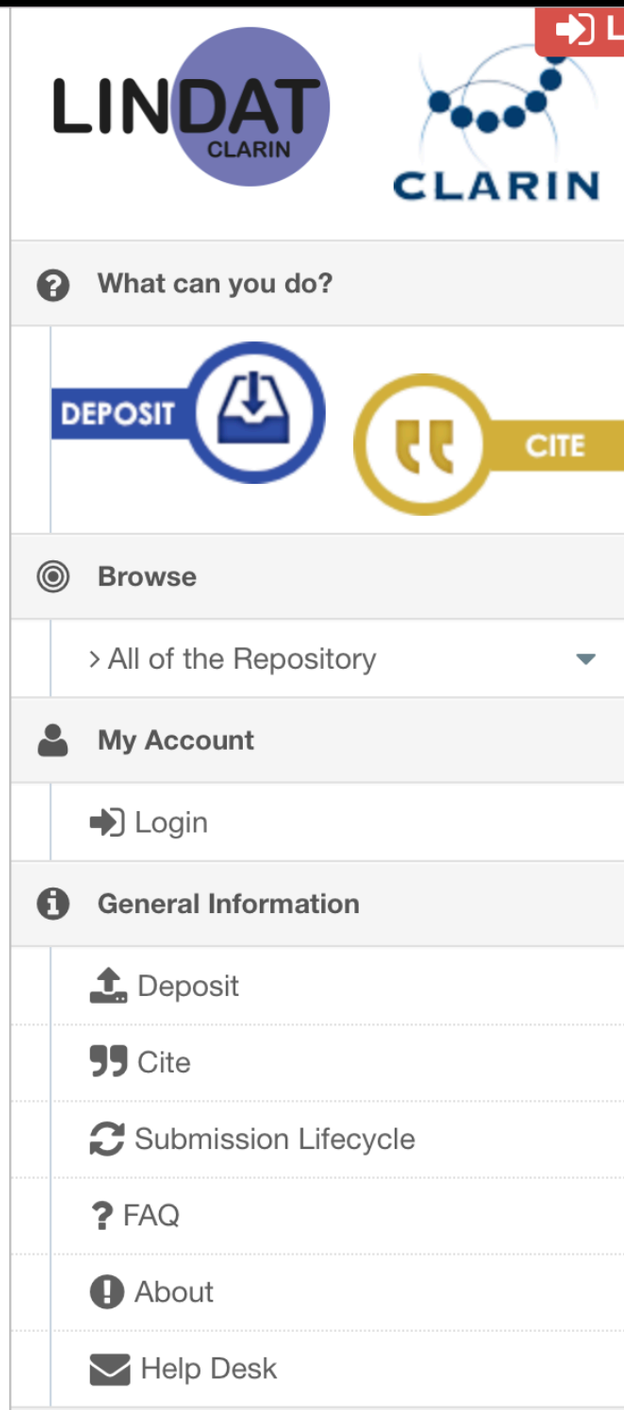
To start a new submission you have to login first. Click Login under My Account in the right menu panel.

Fig1. Menu Login

Step 2: Starting a new submission

Now you have a new menu item 'Submissions' under My Account. Click on Submissions to go to the Submissions screen.





SAFE PRESERVATION

UPLOAD AND DON'T WORRY

How to Deposit

Only authenticated users can deposit items. If you cannot find your home organisation in the Login dialog list of organisations then register at clarin.eu and authenticate using "clarin.eu website account". In case you cannot use any authentication method above or if you encounter a problem, do not hesitate to contact our Help Desk and we can create a local account for you.

Step 1: Login

To start a new submission you have to login first. Click Login under My Account in the right menu panel.

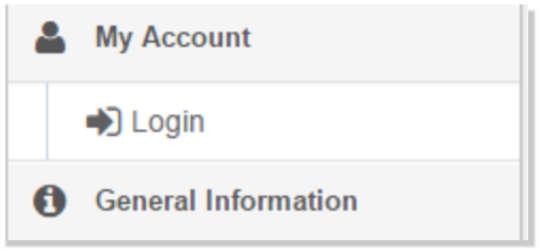
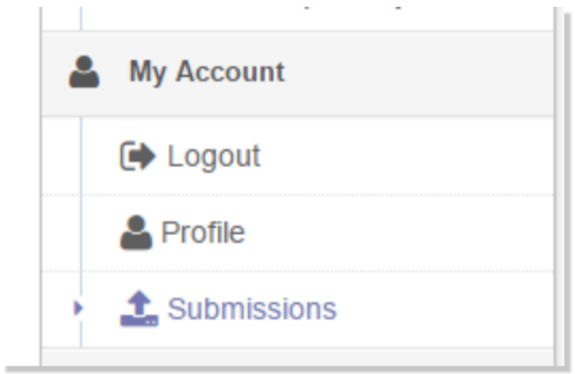
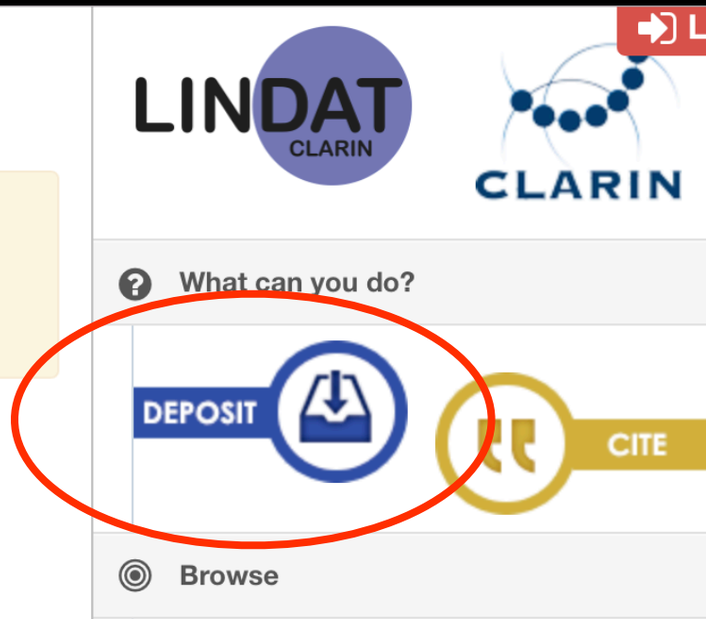


Fig1. Menu Login

Step 2: Starting a new submission

Now you have a new menu item 'Submissions' under My Account. Click on Submissions to go to the Submissions screen.





DEPOSITION GUIDE

- step-by-step description

1. login
2. fill-in metadata
3. drag&drop data
4. select a license
5. submit

Step 3: Select type of your submission

You have initiated a new workflow item. In the next few steps you will provide the details select the type of the resource you are about to submit.

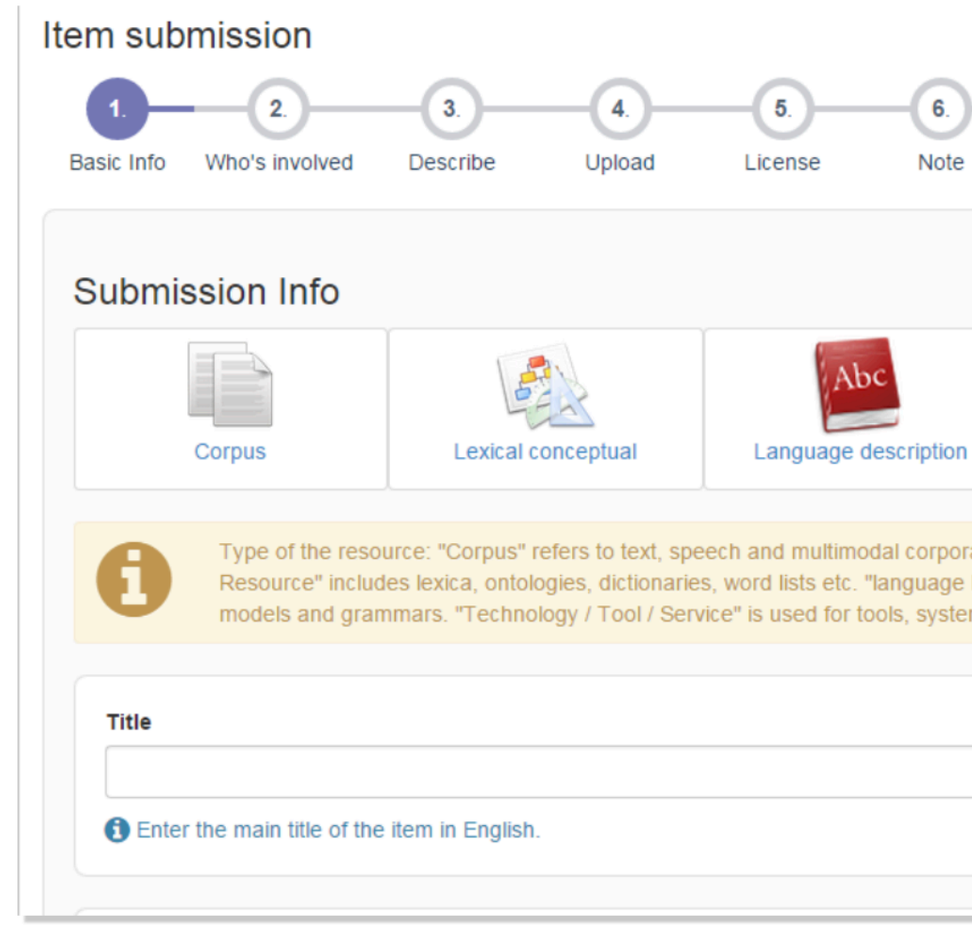


Fig4. Submission info

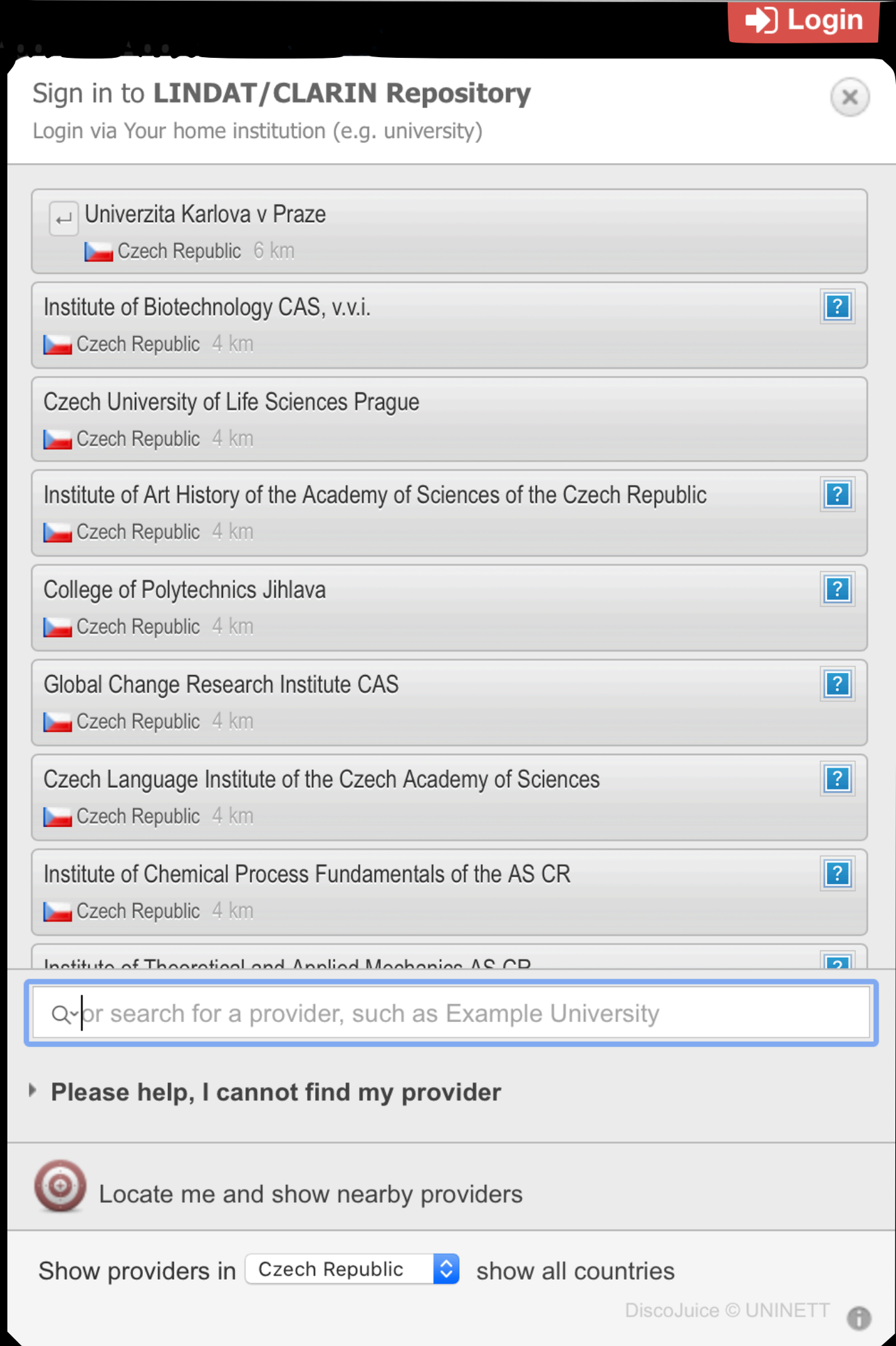
Click on one of the type buttons e.g. Corpus. Proceed with filling the basic information following step.

Step 4: Describe your item

In the following two steps you will provide more details for your item. First describe the the item.

LOGIN TO DEPOSIT

- institutional logins (EduID-cz, EduGAIN)
- CLARIN account for the “homeless researchers”
- minimal personal info



DISCOVERY

FACETED SEARCH

LINDAT CLARIN

Search

Advanced Search

Limit your search

Author

Subject

Rights

Language (ISO)

Type

Contain Files

Community

Showing 1 through 10 out of 1038 results

1 2 3 › 104

Corpus

LINDAT / CLARIN

AKCES 2 ver. 2

(Charles University in Prague, ÚČJTK / 2013-12-18)

Author(s):

Šebesta, Karel ; Goláňová, Hana

This item contains 1 file (3.85 MB).

Publicly Available

LexicalConceptualResource

LRT + Open Submissions

A Gold Standard Word Alignment for English-Swedish (2015-10-12)

(Linköping University / 2015-10-12)

Author(s):

Ahrenberg, Lars ; Holmqvist, Maria

This item contains 1 file (590 KB).

LINDAT CLARIN

CLARIN

What can you do?

DEPOSIT

CITE

Browse

› All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

Help Desk

DISCOVERY

FACETED SEARCH

LINDAT CLARIN

Search

Advanced Search

Limit your search

Author

Subject

Rights

Language (ISO)

Type

Contain Files

Community

Showing 1 through 10 out of 1038 results

1 2 3 › 104

Corpus

LINDAT / CLARIN

AKCES 2 ver. 2

(Charles University in Prague, ÚČJTK / 2013-12-18)

Author(s):

Šebesta, Karel ; Goláňová, Hana

This item contains 1 file (3.85 MB).

Publicly Available

LexicalConceptualResource

LRT + Open Submissions

A Gold Standard Word Alignment for English-Swedish (2015-10-12)

(Linköping University / 2015-10-12)

Author(s):

Ahrenberg, Lars ; Holmqvist, Maria

LINDAT CLARIN

CLARIN

What can you do?

DEPOSIT

CITE

Browse

› All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

Help Desk

DISCOVERY

FACETED SEARCH

LINDAT CLARIN

Search

Advanced Search

Limit your search

- Author
- Subject
- Rights
- Language (ISO)
- Type
- Contain Files
- Community

Showing 1 through 10 out of 1038 results

1 2 3 › 104

Corpus — LINDAT / CLARIN

AKCES 2 ver. 2

(Charles University in Prague, ÚČJTK / 2013-12-18)

Author(s):

Šebesta, Karel ; Goláňová, Hana

This item contains 1 file (3.85 MB).

Publicly Available

LexicalConceptualResource — LRT + Open Submissions

A Gold Standard Word Alignment for English-Swedish (2015-10-12)

(Linköping University / 2015-10-12)

Author(s):

Ahrenberg, Lars ; Holmqvist, Maria

This item contains 1 file (590 KB).

LINDAT CLARIN

CLARIN

What can you do?

DEPOSIT

CITE

Browse

› All of the Repository

My Account

Login

General Information

Deposit

Cite

Submission Lifecycle

FAQ

About

Help Desk

DISCOVERY

GOOGLE

Google

prague dependency treebank 3.0

Vše | Obrázky | Nákupy | Mapy | Videa | Více | Nastavení | Nástroje

Přibližný počet výsledků: 13 900 (0,44 s)

Vědecké články o **prague dependency treebank 3.0**

Prague dependency treebank 3.0 - Bejček - Počet citací tohoto článku: 47

Prague Dependency Treebank - Hajič - Počet citací tohoto článku: 385

The **Prague dependency treebank** - Böhmová - Počet citací tohoto článku: 423

Prague Dependency Treebank 3.0 | ÚFAL

https://ufal.mff.cuni.cz/pdt3.0 ▾ Přeložit tuto stránku

Introduction. The **Prague Dependency Treebank 3.0** (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDiT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed and improved in various aspects. Moreover ...

The Prague Dependency Treebank 2.0.

https://ufal.mff.cuni.cz/pdt2.0/ ▾ Přeložit tuto stránku

The **Prague Dependency Treebank** 2.0 (PDT 2.0) contains a large amount of Czech texts with complex and interlinked morphological (2 million words), syntactic (1.5 MW) and complex semantic annotation ... Please note that new versions of this corpus have been published: PDT **3.0** (2013), PDiT 1.0 (2012), PDT 2.5 (2012).

Prague Dependency Treebank 3.0 (PDT 3.0)

https://lindat.mff.cuni.cz/repository/xmlui/bitstream/.../PDT30_index_lindat.html?...

Prague Dependency Treebank 3.0 (PDT 3.0). Overview. The **Prague Dependency Treebank 3.0** (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDiT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed ...

DISCOVERY

GOOGLE

Google

prague dependency treebank 3.0

Vše | Obrázky | Nákupy | Mapy | Videa | Více | Nastavení | Nástroje

Přibližný počet výsledků: 13 900 (0,44 s)

Vědecké články o **prague dependency treebank 3.0**

Prague dependency treebank 3.0 - Bejček - Počet citací tohoto článku: 47

Prague Dependency Treebank - Hajič - Počet citací tohoto článku: 385

The **Prague dependency treebank** - Böhmová - Počet citací tohoto článku: 423

Prague Dependency Treebank 3.0 | ÚFAL

https://ufal.mff.cuni.cz/pdt3.0 ▾ Přeložit tuto stránku

Introduction. The **Prague Dependency Treebank 3.0** (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDiT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed and improved in various aspects. Moreover ...

The Prague Dependency Treebank 2.0.

https://ufal.mff.cuni.cz/pdt2.0/ ▾ Přeložit tuto stránku

The **Prague Dependency Treebank** 2.0 (PDT 2.0) contains a large amount of Czech texts with complex and interlinked morphological (2 million words), syntactic (1.5 MW) and complex semantic annotation ... Please note that new versions of this corpus have been published: PDT **3.0** (2013), PDiT 1.0 (2012), PDT 2.5 (2012).

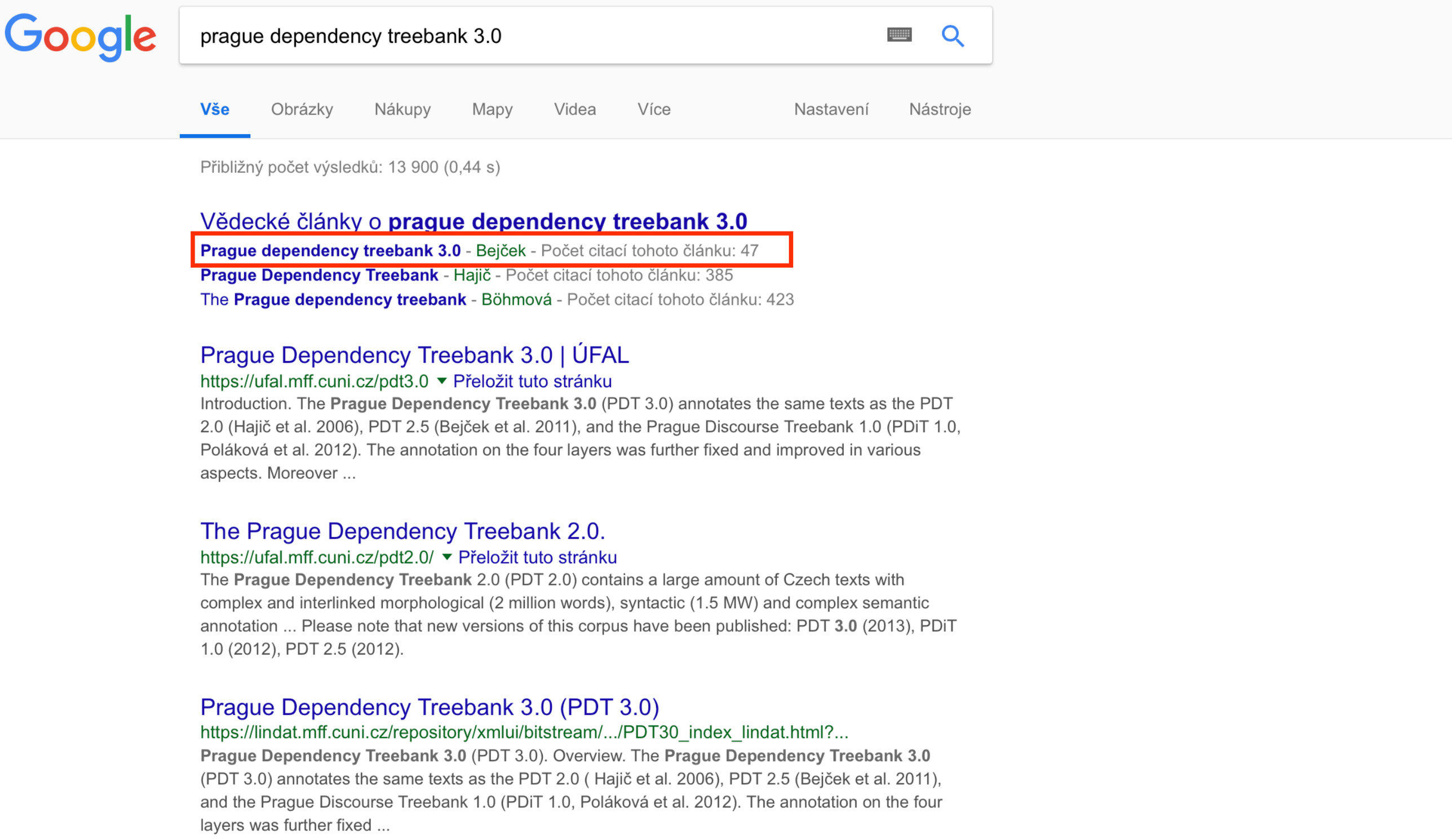
Prague Dependency Treebank 3.0 (PDT 3.0)

https://lindat.mff.cuni.cz/repository/xmlui/bitstream/.../PDT30_index_lindat.html?...

Prague Dependency Treebank 3.0 (PDT 3.0). Overview. The **Prague Dependency Treebank 3.0** (PDT 3.0) annotates the same texts as the PDT 2.0 (Hajič et al. 2006), PDT 2.5 (Bejček et al. 2011), and the Prague Discourse Treebank 1.0 (PDiT 1.0, Poláková et al. 2012). The annotation on the four layers was further fixed ...

DIRECT DATA CITATIONS

CREDIT FOR DATA



DIRECT DATA CITATIONS

CREDIT FOR DATA

LINDAT CLARIN

enTenTen

Please use the following text to cite this item or export to a predefined format: BIBTEX CMDI

Masaryk University, NLP Centre, 2011, *enTenTen*, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8.

This resource is also integrated in following services: KonText

Share:

Item identifier	http://hdl.handle.net/11858/00-097C-0000-0001-CCDF-8
Date issued	2011-12-16
Type	corpus
Language(s)	English
Description	Very large English web corpus enTenTEn, comprising 3,268,798,627 tokens.
Publisher	Masaryk University, NLP Centre
Acknowledgement	Lexical Computing Ltd.
Subject(s)	English large corpus
Collection(s)	LINDAT / CLARIN Data & Tools

Show full item record

Pavel Straňák | Lo

LINDAT CLARIN

CLARIN

What can you do?

DEPOSIT

CITE

Browse

› All of the Repository

My Account

Logout

Profile

Submissions

Context

› Edit this item

› Export Item

› Export Metadata

Administrative

Control Panel

LICENSING

AS OPEN AS POSSIBLE

Choose a License

Answer the questions or use the search to find the license you want

Start again

What do you want to deposit?

Software | Data

Search for a license...

Public Domain Mark (PD)

The work identified as being free of known restrictions under copyright law, including all related and neighboring rights.

Publicly Available

Public Domain Dedication (CC Zero)

CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Publicly Available

LICENSING

AS OPEN AS POSSIBLE (NOT MORE)

Publisher Faculty of Arts, Institute of the Czech National Corpus, Charles University in Prague

Acknowledgement Ministerstvo školství, mládeže a tělovýchovy

Project code: LM2011023

Project name: Český národní korpus

Subject(s) representative corpus written language

Collection(s) LINDAT / CLARIN Data & Tools

Show full item record

Files in this item

Download instructions for command line

This item is Academic Use and licensed under:
Czech National Corpus (Shuffled Corpus Data)

Name	syn2015.gz
Size	1.48 GB
Format	application/x-gzip
Description	corpus
MD5	e0242cc77e999794af6cfaf57f843c12

Download file

LICENSING

AS OPEN AS POSSIBLE (NOT MORE)

Publisher Faculty of Arts, Institute of the Czech National Corpus, Charles University in Prague

Acknowledgement Ministerstvo školství, mládeže a tělovýchovy

Project code: LM2011023

Project name: Český národní korpus

Subject(s) representative corpus written language

Collection(s) LINDAT / CLARIN Data & Tools

Show full item record

Files in this item

Download instructions for command line

This item is Academic Use and licensed under:
Czech National Corpus (Shuffled Corpus Data)

Name	syn2015.gz
Size	1.48 GB
Format	application/x-gzip
Description	corpus
MD5	e0242cc77e999794af6cfaf57f843c12

Download file

CLEAR RULES
CUSTOM LICENSES
LICENSE SIGNING

ANY LICENSE (OPEN SOURCE / OPEN DATA PREFERRED)

LICENSING FRAMEWORK

Manage Licenses

All Licenses | Define License | Define License Label

	License Name	Definition (URL)	Confirmation	Required user info	License Label	Extended Labels	Used by Bitstreams
☐	Universal Derivations v0.5 License Agreement	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UDer-0.5	Not required		PUB	CC	1
☐	Licence Universal Dependencies v2.4	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.4	Not required		PUB	CC GPLv3	4
☐	License agreement for The Multilingual corpus of literal occurrences of multiword expressions	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-literal	Not required		PUB	CC GPLv3	5
☐	Licence Universal Dependencies v2.3	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.3	Not required		PUB	CC GPLv3 GPLv2	4
☐	PARSEME Shared Task Data (v. 1.1) Agreement	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-1.1	Not required		PUB	CC GPLv3	22
☐	Licence Universal Dependencies v2.2	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.2	Not required		PUB	CC GPLv3 GPLv2	7
☐	Licence Universal Dependencies v2.1	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.1	Not required		PUB	CC GPLv3 GPLv2	4
☐	PARSEME Shared Task Data (v. 1.0) Agreement	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-mwe-1.0	Not required		PUB	CC GPLv3	21
☐	Licence Universal Dependencies v2.0	https://lindat.mff.cuni.cz/repository/xmlui/page/licence-UD-2.0	Not required		PUB	CC GPLv3 GPLv2	14
☐	Licence Universal Dependencies v1.4	https://lindat.mff.cuni.cz/re	Not		PUB	CC	4

Pavel Straňák | Logout

LINDAT CLARIN

CLARIN

What can you do?

DEPOSIT

CITE

Browse

› All of the Repository

My Account

Logout

Profile

Submissions

Administrative

Control Panel

Access Control

› Content Administration

Registries

Collections & Communities

Statistics

Curation Tasks

Licenses

DEFINING A NEW LICENSE

LICENSING FRAMEWORK

Manage Licenses

All Licenses | Define License | Define License Label

Define new licence

License name	
License definition URL	
License requires confirmation	Not required
License Labels	○ Publicly Available (PUB) ○ Academic Use (ACA) ○ Restricted Use (RES)
License Labels	☐ Attribution Required (BY) ☐ Share Alike (SA) ☐ Noncommercial (NC) ☐ Redeposit Modified (ReD) ☐ No Derivative Works (ND) ☐ Inform Before Use (Inf) ☐ Distributed under Creative Commons (CC) ☐ No Copyright (ZERO) ☐ GNU General Public License, version 3.0 (GPLv3) ☐ GNU General Public License, version 2.0 (GPLv2) ☐ BSD (BSD) ☐ The MIT License (MIT) ☐ The Open Source Initiative (OSI)
Additional required user info	☐ The user will receive an email with download instructions. ☐ User Name ☐ Date of Birth ☐ Address ☐ Country

Pavel Straňák | Logout

LINDAT CLARIN

CLARIN

What can you do?

DEPOSIT

CITE

Browse

› All of the Repository

My Account

Logout

Profile

Submissions

Administrative

Control Panel

Access Control

› Content Administration

Registries

Collections & Communities

Statistics

Curation Tasks

Licenses

DEFINING A LICENSING LABEL / ATTRIBUTE

LICENSING FRAMEWORK

Manage Licenses

All Licenses | Define License | Define License Label

Define New Label

Short Label	
Label Title	
Is extended	Yes
Icon image	Choose File no file selected
	Save Return

LINDAT CLARIN

CLARIN

Pavel Straňák | Logout

What can you do?

DEPOSIT

CITE

Browse

> All of the Repository

My Account

Logout

Profile

Submissions

Administrative

Control Panel

Access Control

> Content Administration

Registries

Collections & Communities

Statistics

Curation Tasks

Licenses

Project name: Internet jako jazykový korpus

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: LN00A063

Project name: Centrum komputační lingvistiky

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: MSM 0021620838

Project name: Moderní metody, struktury a systémy informatiky

Subject(s) MorphoDiTa | Czech | morphological analysis | morphological generation | PoS tagging

Collection(s) LINDAT / CLARIN Data & Tools

This item is replaced by a newer submission:
http://hdl.handle.net/11234/1-1836

Please refer to the submission above for the latest available data. If you nevertheless need the original data, please click here.

List all versions

Project name: Internet jako jazykový korpus

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: LN00A063

Project name: Centrum komputační lingvistiky

Ministerstvo školství, mládeže a tělovýchovy České republiky

Project code: MSM 0021620838

Project name: Moderní metody, struktury a systémy informatiky

Subject(s) MorphoDiTa | Czech | morphological analysis | morphological generation | PoS tagging

Collection(s) LINDAT / CLARIN Data & Tools

VERSIONING

PREFER LATEST, PRESERVE ALL

LINDAT CLARIN

Collection(s) LINDAT / CLARIN Data & Tools

Other versions List all versions

Show full item record

Files in this item

Download instructions for command line

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Name	czech-morfflex-pdt-161115.zip
Size	69.18 MB
Format	application/zip
Description	Czech Models (MorfFlex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
MD5	adde38cd363219759e19165b06baa4ce

Download file Preview

VERSIONING

PREFER LATEST, PRESERVE ALL

Collection(s) LINDAT / CLARIN Data & Tools

Other versions List all versions

Show full item record

Files in this item

Download instructions for command line

This item is Publicly Available and licensed under:
Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Name	czech-morfflex-pdt-161115.zip
Size	69.18 MB
Format	application/zip
Description	Czech Models (MorfFlex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
MD5	adde38cd363219759e19165b06baa4ce

Download file Preview

VERSIONING

PREFER LATEST, PRESERVE ALL

Collection(s) LINDAT / CLARIN Data & Tools

Other versions List all versions

- Czech Models (MorfFlex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
- Czech Models (MorfFlex CZ 160310 + PDT 3.0) for MorphoDiTa 160310
- Czech Models (MorfFlex CZ + PDT) for MorphoDiTa

Show full item record

Files in this item

Download instructions for command line

This item is **Publicly Available** and licensed under:
Creative Commons - Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Name	czech-morfflex-pdt-161115.zip
Size	69.18 MB
Format	application/zip
Description	Czech Models (MorfFlex CZ 161115 + PDT 3.0) for MorphoDiTa 161115
MD5	adde38cd363219759e19165b06baa4ce

Download file **Preview**

VERSIONING – TECHNICAL

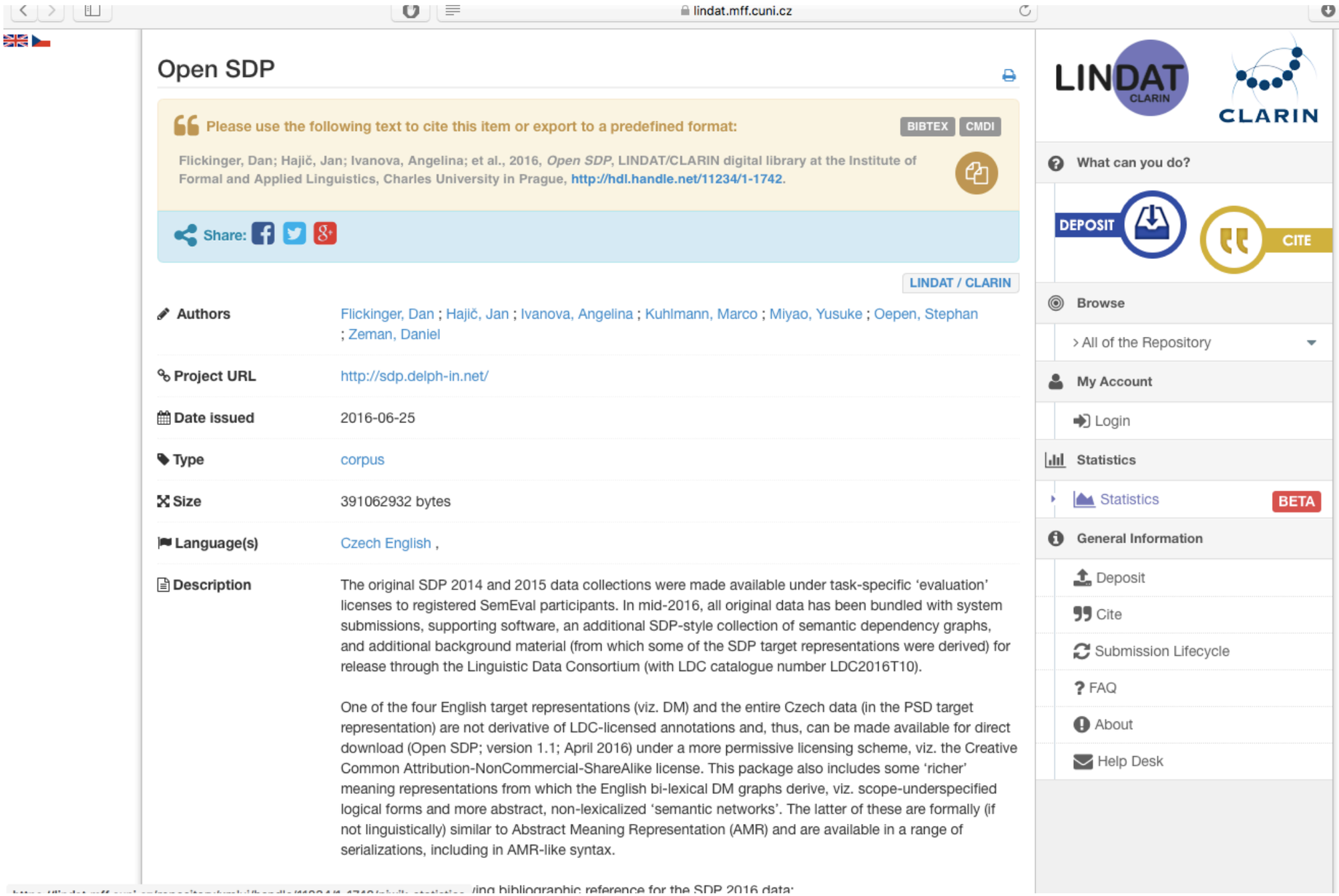
- Clone and modify previous version
  - auto-fill linking metadata:
    `dc.relation.replaces`
    `dc.relation.isreplacedby`
- hide files of older versions and point to the newest
- Promote the latest in search
- No “all-versions PID”

WHY CLARIN-DSPACE

- Modifications for external assignment of PIDs
- Licensing framework (CC is not everything)
- User Experience (search-centric)
- Admin experience: better control panel
- Citations (Force11 – direct data citations)
- Statistics (global – Matomo; item: graphs, reports)
- Integrations: CLARIN VLO, Clarivate DCI, OpenAIRE, EUDAT, …

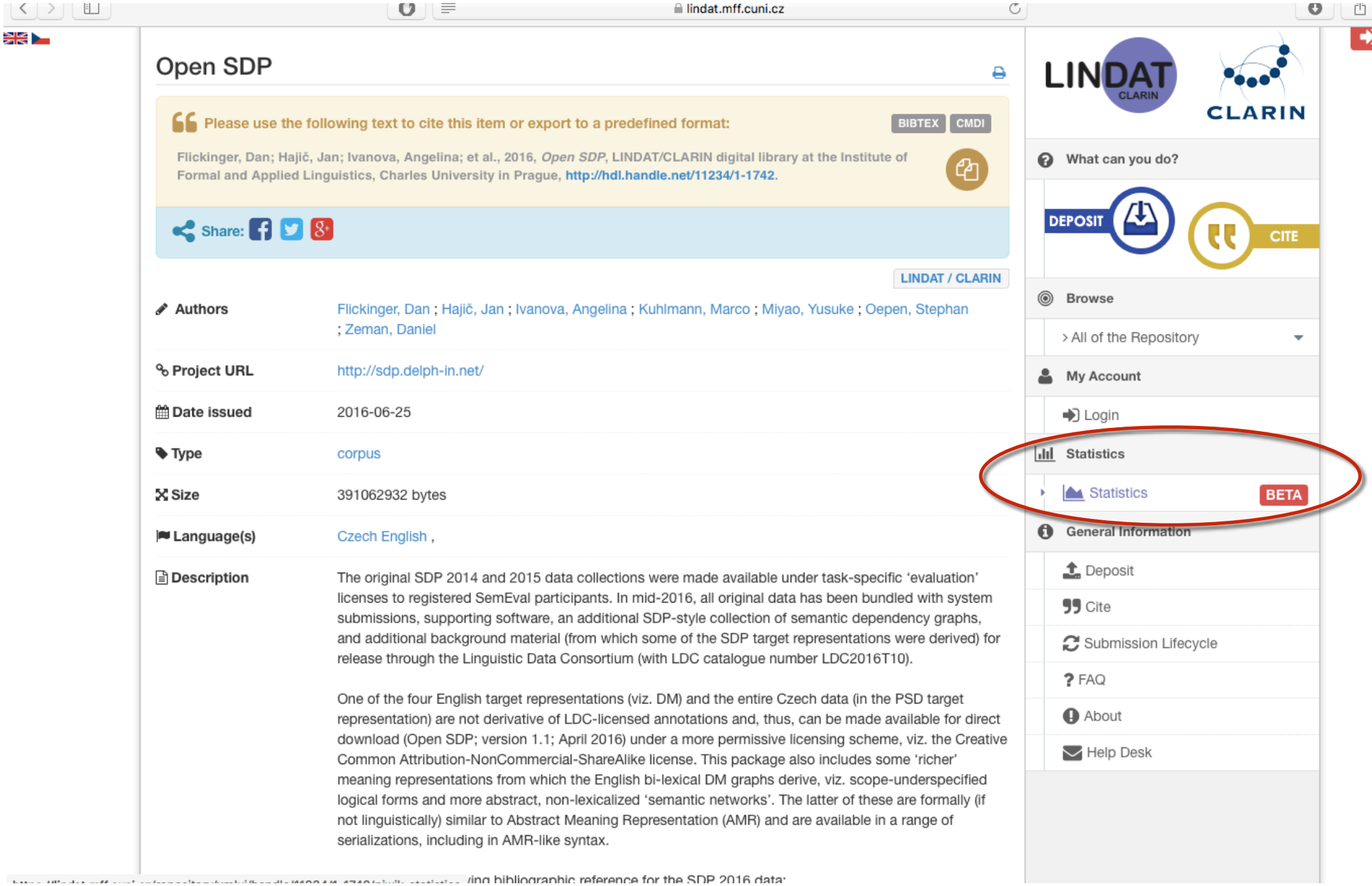
MUCH MORE

- content negotiation for XML metadata or HTML page
- EUDAT B2SAFE replication
  - external tool linked to DSpace
- summary statistics
- bibtex format for bibliography (for typesetting in LaTeX)
- optional request for user information
  (custom form before download)



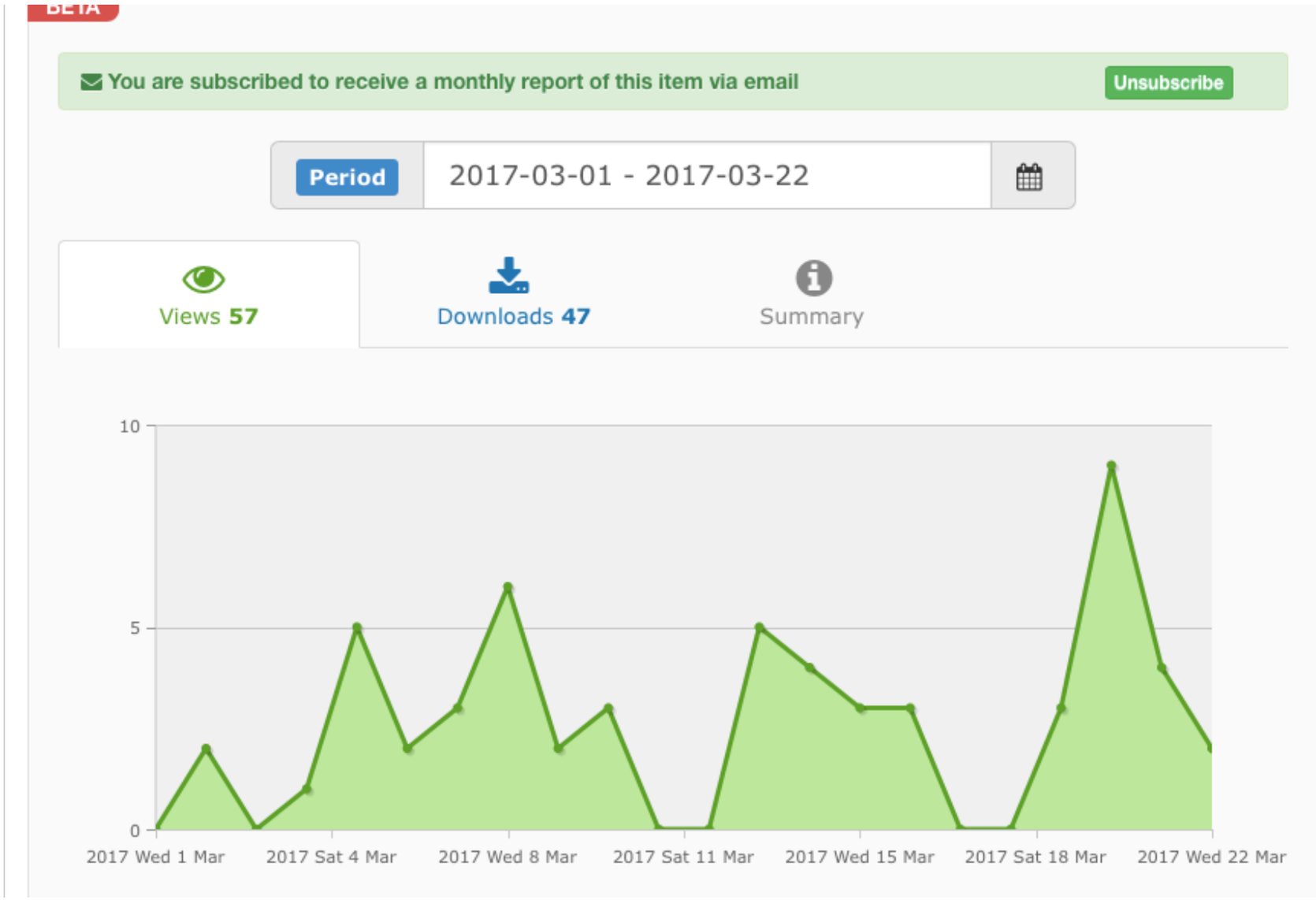
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



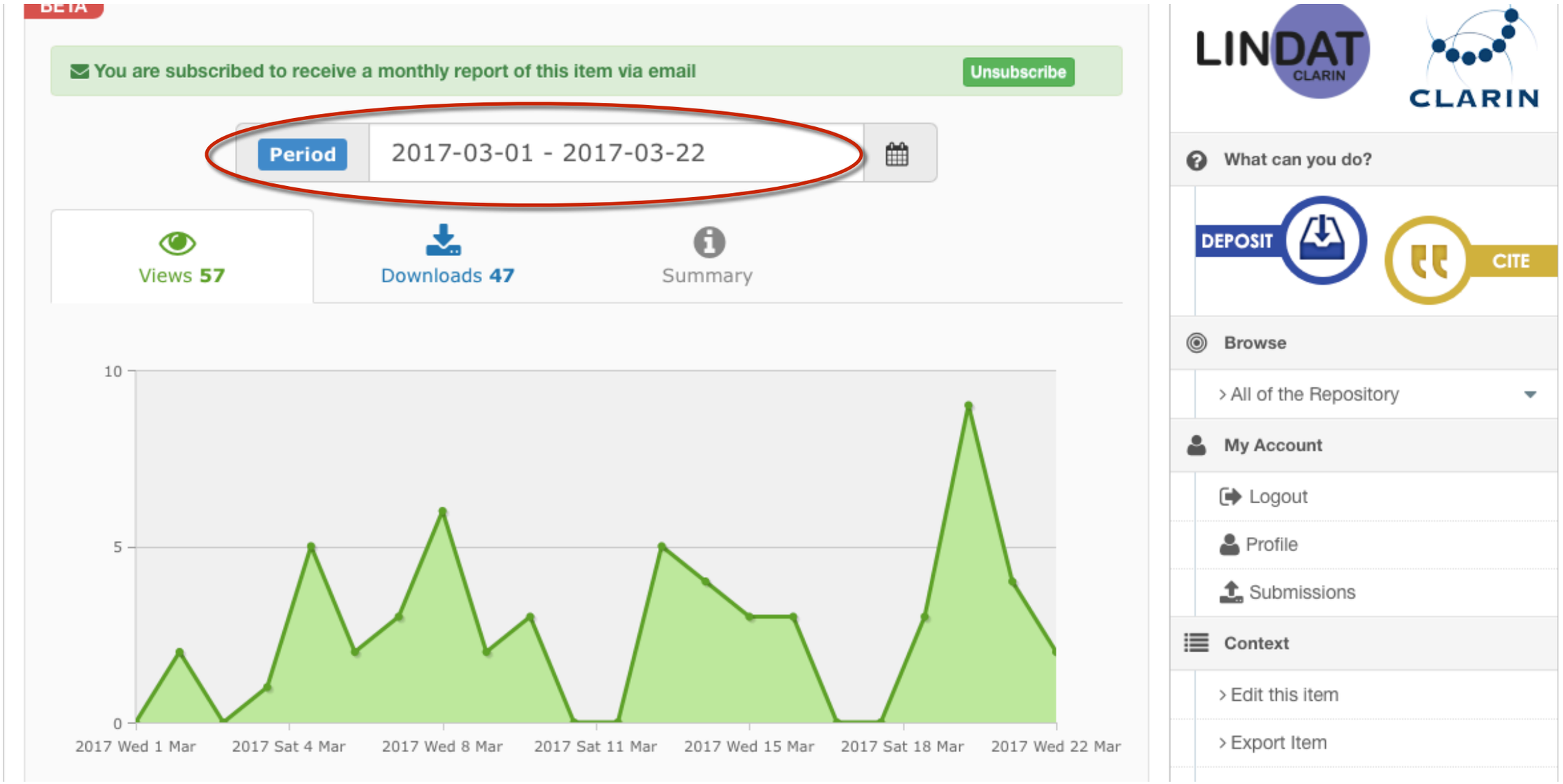
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



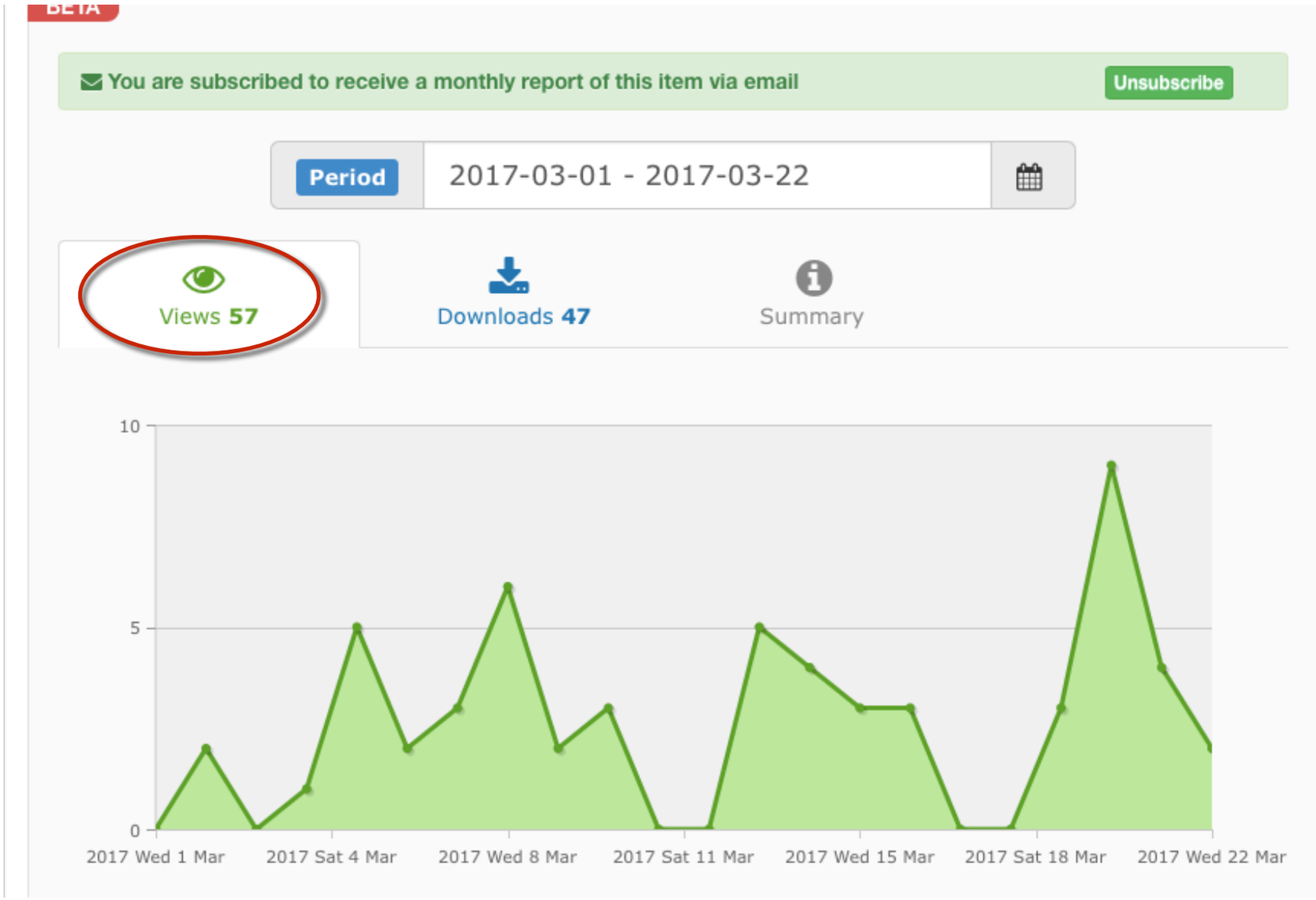
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



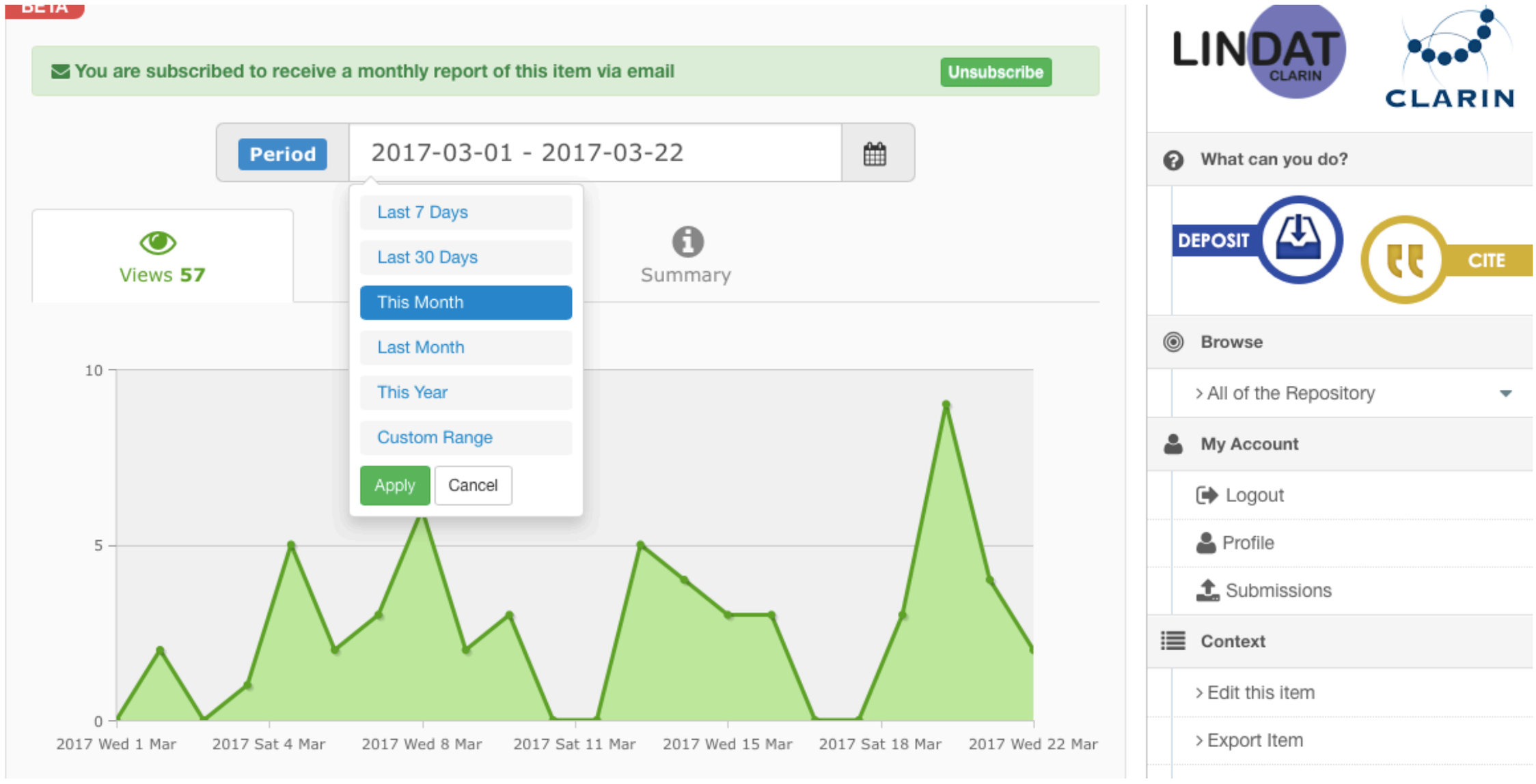
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



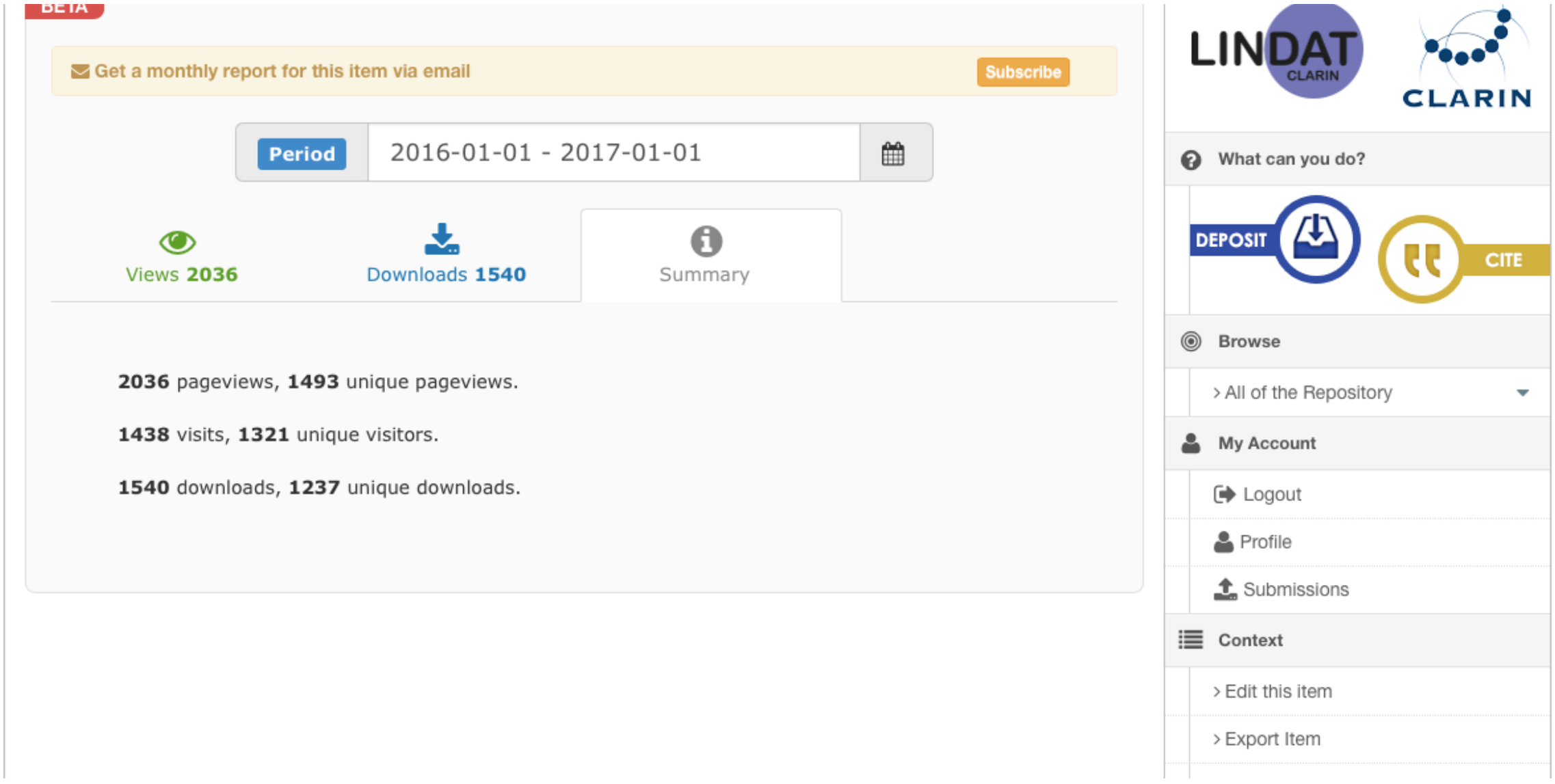
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



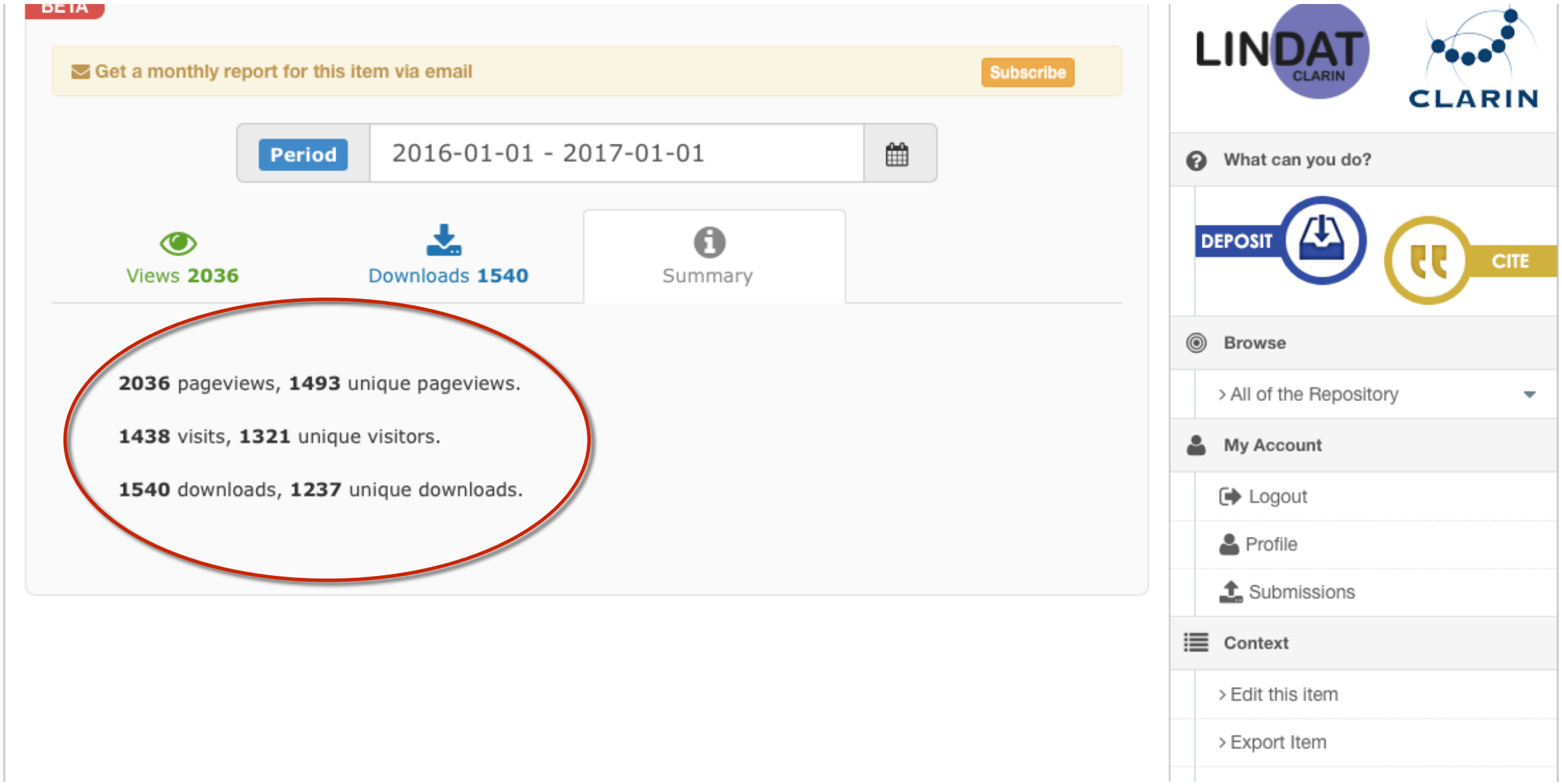
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



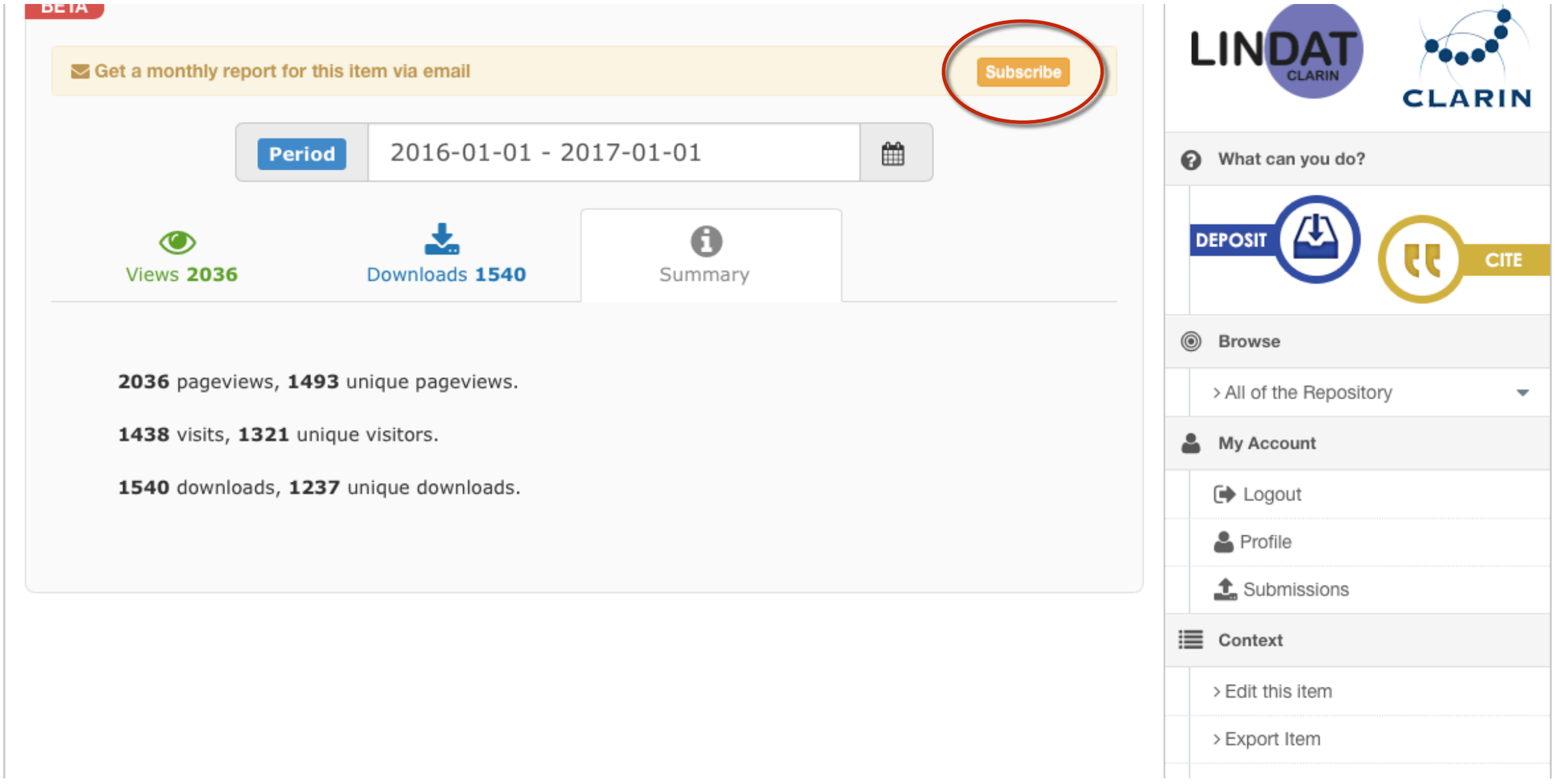
Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item



Item Statistics

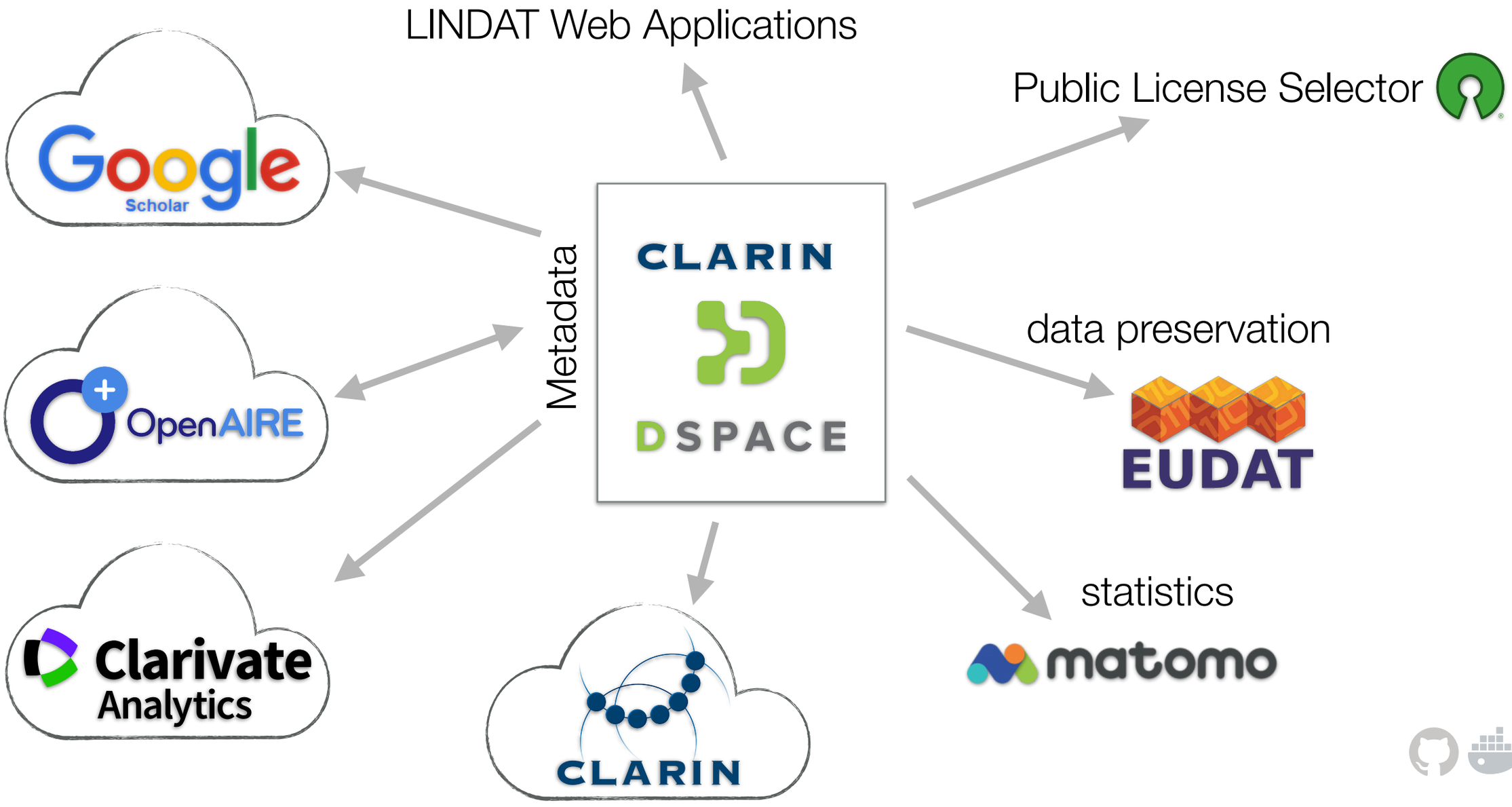
- View item visits and downloads in time
- Subscribe to monthly statistics of the item



Item Statistics

- View item visits and downloads in time
- Subscribe to monthly statistics of the item

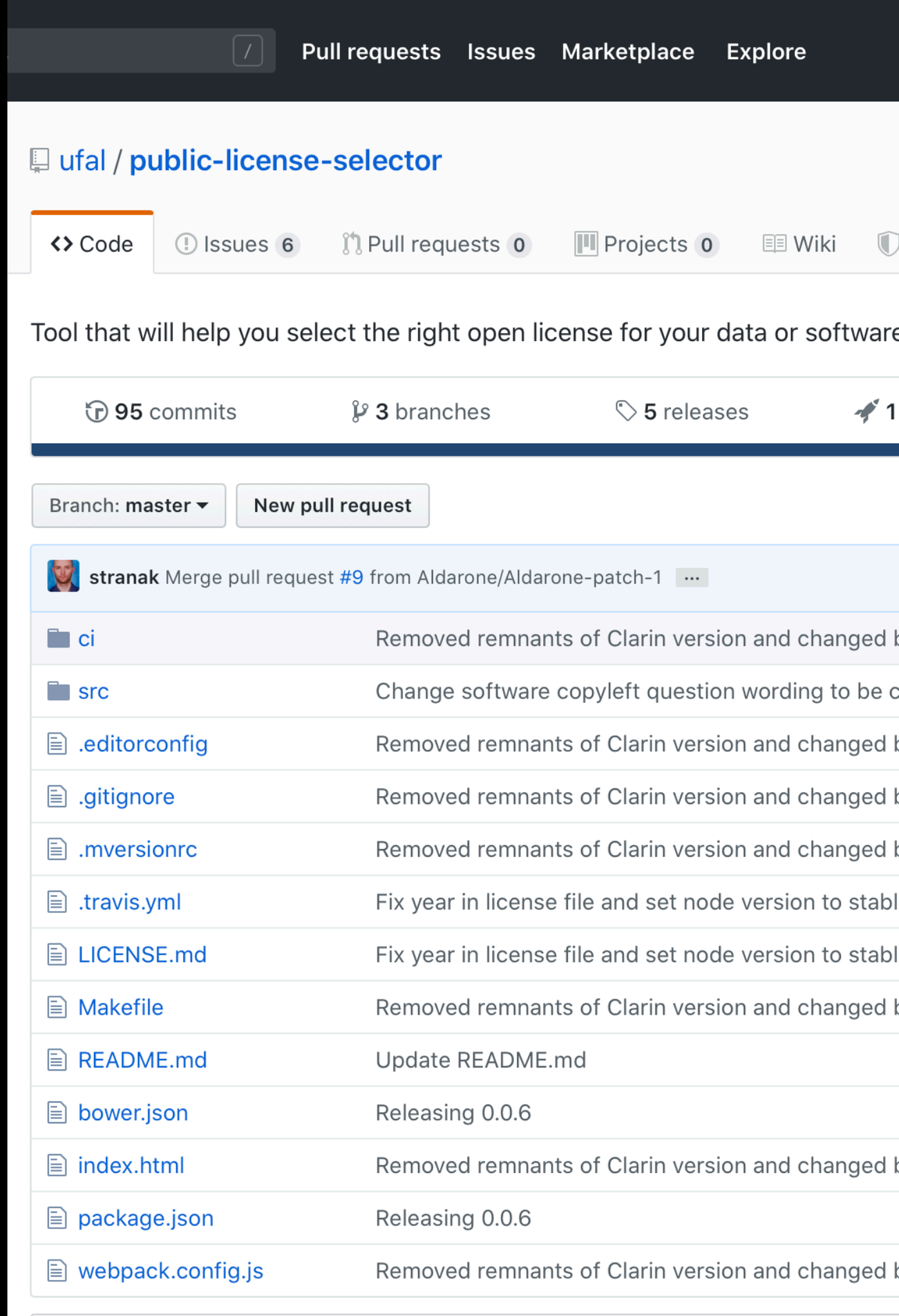
Integrations



PUBLIC LICENSE SELECTOR

https://github.com/ufal/public-license-selector

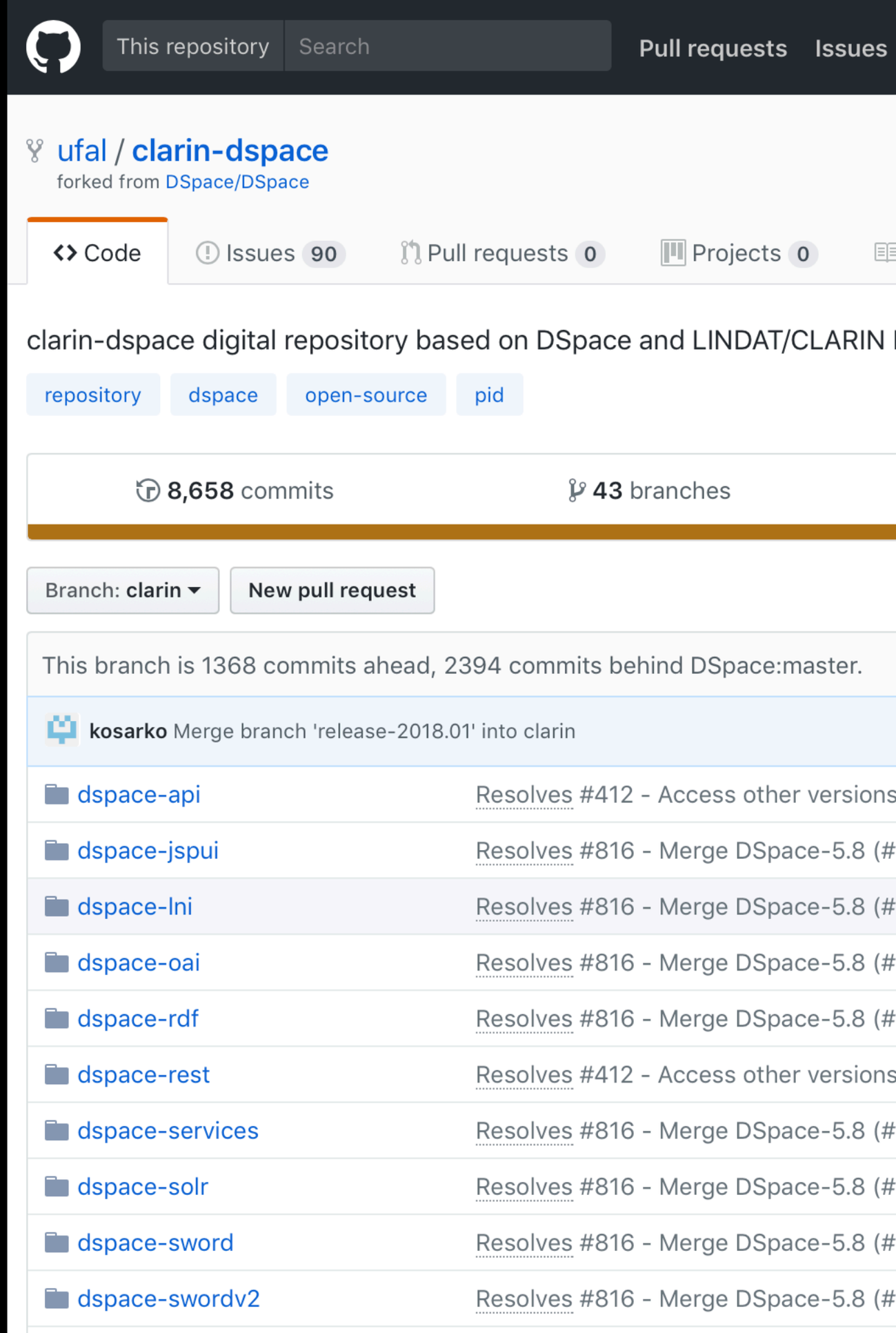
- public license: no signatures, public distribution
- data / software
- explanations provided
- choose as open as possible
- open source / open data
  - best licenses chosen



CLARIN-DSPACE

https://github.com/ufal/clarin-dspace

- MIT license
- DSpace + licensing, versioning and more
- LINDAT's project converting to community
- Issues, Documentation
- 15+ deployments
  10+ countries
- since 24 October 2009



FAIR SUMMARY

F*indable*: Google, Google Scholar, Data Citation Index, CLARIN VLO, OLAC… and the repository itself

A*ccessible*: records with data (even when restricted), complete licensing, Open Access (Public License Selector), login only when needed, CESNET, EUDAT

I*nteroperable*: common data formats, full documentation (enhanced metadata, documentation bitstreams)

R*eusable*: records with data, complete licensing, full versioning, direct data citations, maximal OA

Thank you!

http://lindat.cz

LINDAT CLARIN

MINISTRY OF EDUCATION
YOUTH AND SPORTS