

SAFE AND EASY REPLICATION OF DSPACE REPOSITORIES

B2SAFE WITH DSPACE

PAVEL STRAŇÁK

SAFE AND EASY REPLICATION OF DSPACE REPOSITORIES

B2SAFE WITH DSPACE

PAVEL STRAŇÁK



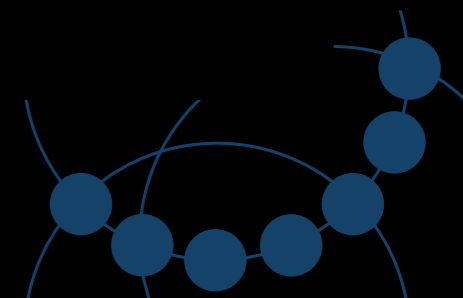
B2SAFE
Replicate Research Data Safely

LANGUAGE DATA IN HUMANITIES: HERITAGE AND PROGRESS

CLARIN – LANGUAGE DATA



B2SAFE
Replicate Research Data Safely



B2SAFE – PRESERVATION

WE NEED BACKUPS



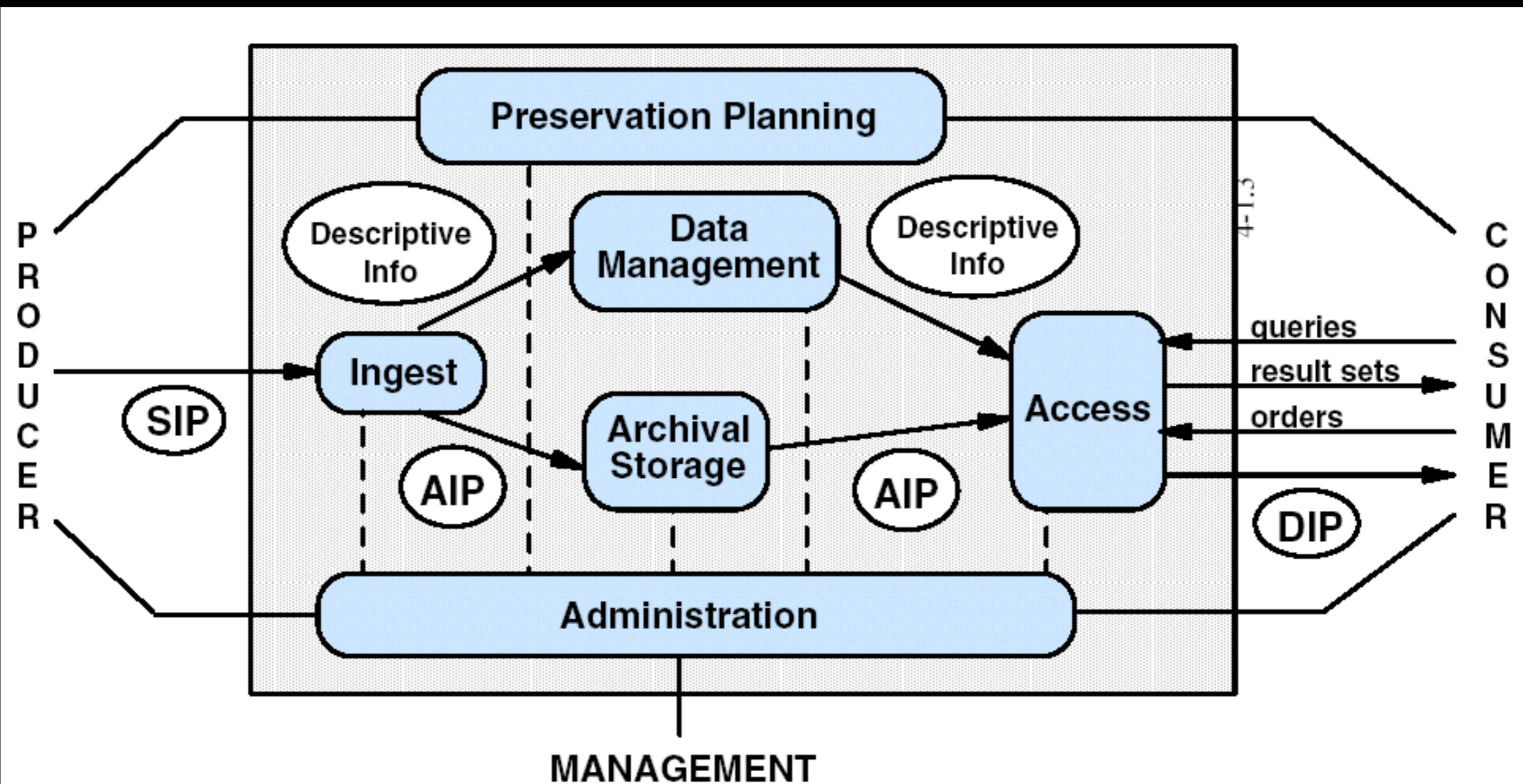
B2SAFE – PRESERVATION

GOOD BACKUPS



WHAT TO PRESERVE AND HOW?

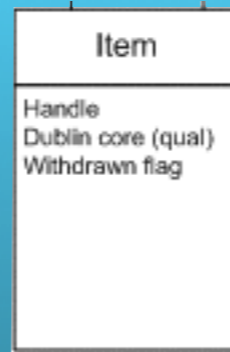
OAIIS REPOSITORY MODEL



ARCHIVE IT ALL, ATOMICALLY – ARCHIVAL INFORMATION PACKAGE

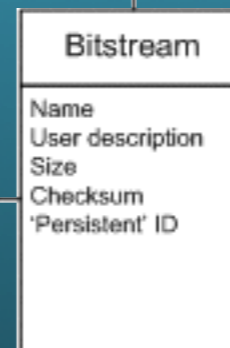
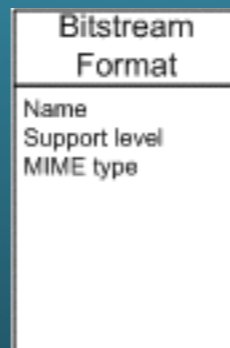
AIP

METS



STRUCTURE

LICENSE



CONTENT FILES



OUR DEVELOPMENT IS OPEN

LINDAT-CLARIN B2SAFE SETUP

Home Projects Help

Sign In

LINDAT Repository

Search:

Overview Activity Roadmap Issues New issue Documents **Wiki** Repository

WikiStart »

Eudat Replication

What do we replicate

Implementation

UI

Todo

Test it

Related tasks

Links

Wiki

[Start page](#)

[Index by title](#)

[Index by date](#)

Eudat Replication

iRods (<https://www.irods.org>) technology is used for replication in EUDAT project. We use the replication for submissions. Our requirement is to have this module as an integral part of DSpace. The original DSpace module ([link](#)) cannot be used because it is too outdated; however, there is another implementation at <https://www.irods.org/index.php/DSpace> but has its own problems. The default jargon irods client library ([link](#)) shipped with DSpace had to be updated as well.

The workflow from the client point-of-view is simple: file replicas are uploaded somewhere* and you access the replicas using handles.

For now, we obtained credentials to cines (ariane.cines.fr) from Stephane.

Finally, after upgrading and reimplementing the plugin we have a working solution, see Control panel's IRODs replication tab (only on <https://ufal-point-dev.ms.mff.cuni.cz/jm/repository/> for now)

What do we replicate

We automatically replicate each submission after it has been approved. The submission is converted to AIP format which is uploaded to the iRods server. We use our PID in the name of each AIP e.g.,

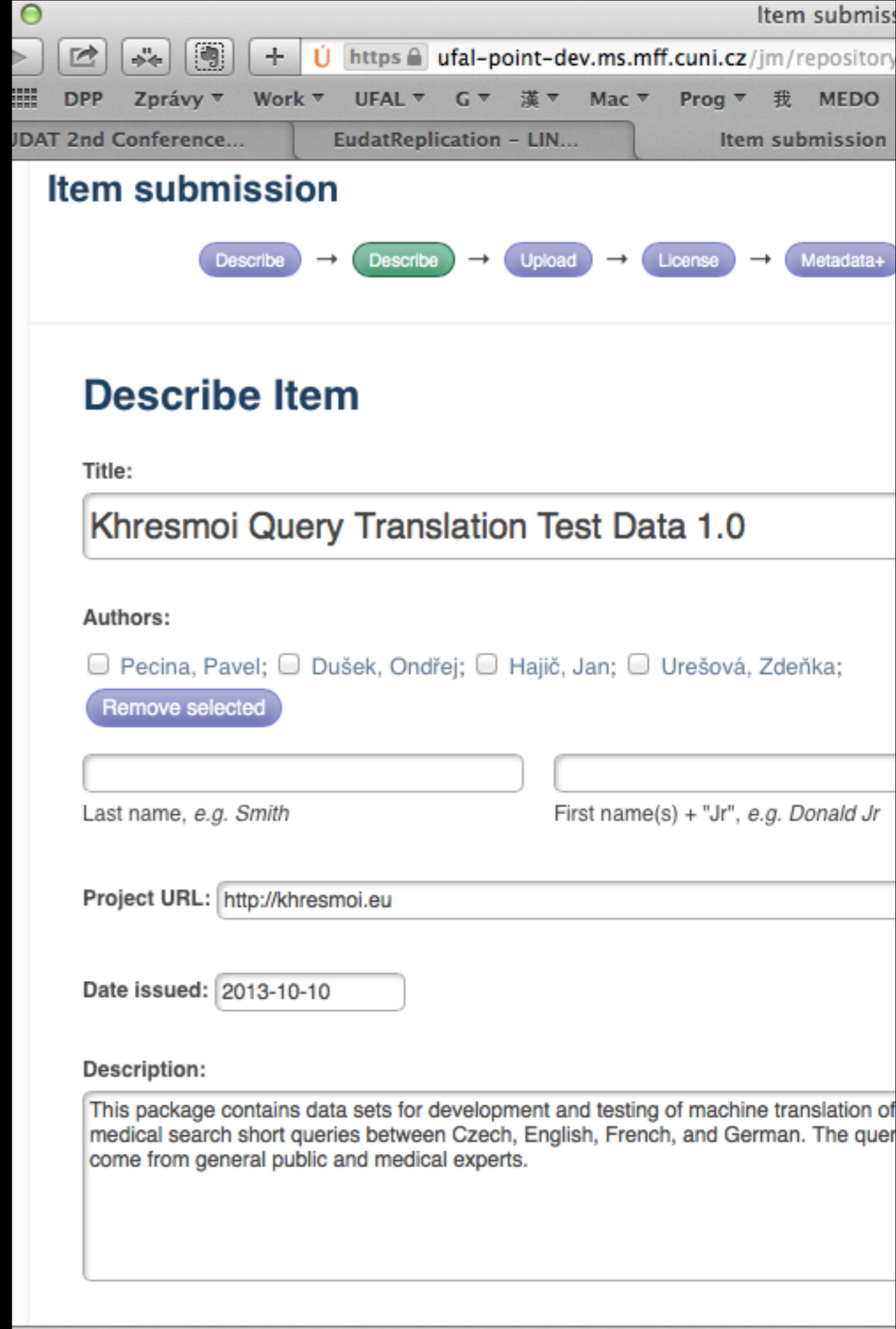
```
irods://XXX@ariane.cines.fr:XX/CINESZone/home/jmisutka/dspace_1.8.2/11858_00-097C-0000-0001-487A-4-6451959456007
irods://XXX@ariane.cines.fr:XX/CINESZone/home/jmisutka/dspace_1.8.2/11858_00-097C-0000-0001-487E-B-6544474914476
```

B2SAFE WITH
DSPACE



B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)



Item submission

Describe → Describe → Upload → License → Metadata+

Describe Item

Title:

Authors:

Pecina, Pavel; Dušek, Ondřej; Hajič, Jan; Urešová, Zdeňka;

Last name, e.g. Smith

First name(s) + "Jr", e.g. Donald Jr

Project URL:

Date issued:

Description:

This package contains data sets for development and testing of machine translation of medical search short queries between Czech, English, French, and German. The queries come from general public and medical experts.

B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)

Item submission


https://ufal-point-dev.ms.mff.cuni.cz/jm/repository/handle/11858/00-09


DPP Zprávy Work UFAL G 漢 Mac Prog 我 MEDO Corpus.BYE W@SiH

2nd Conference... EudatReplication - LIN... Item submission Hand

This item is **PUB** and licensed under:

[Creative Commons - Attribution 3.0 Unported \(CC BY 3.0\)](#)



 **Name:** test_upload.txt
Size: 20bytes
Format: Text file
Description: Unknown

[Show full item record](#)

Actions you may perform on this task:

If you have reviewed the item and it is suitable for inclusion in the collection, select "Approve". [Approve item](#)

Validate this item with the curation framework. [Curate item](#)

If you have reviewed the item and found it is **not** suitable for inclusion in the collection, select "Reject". You will then be asked to enter a message indicating why the item is unsuitable, and whether the submitter should change something and resubmit. [Reject item](#)

Select this option to change the item's metadata. [Edit metadata](#)

Return the task to the pool so that another user may perform the task. [Return task to pool](#)

[Cancel](#)

B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published



Home

https://ufal-point.mff.cuni.cz/repository/xmlui/

Zprávy Work UFAL G 漢 Mac Prog 我 MEDO Corpus.BYE W@SiH

EudatReplica... Item submis... Handle Proxy Home

[Browse](#) | [Advanced Search](#)

Author	Subject	Language (ISO)
Hajič, Jan (15)	Machine Translation (5)	Ces (37)
Bojar, Ondřej (11)	Treebank (5)	Eng (12)
Pajas, Petr (10)	Corpus (4)	Sik (8)
Straňák, Pavel (8)	Morphology (4)	Deu (3)
Štěpánek, Jan (7)	Parallel Corpus (4)	Hin (3)
... View More	... View More	... View More

What's New

<p>Khresmoi Query Translation Test Data 1.0</p> <p>Author(s): Pecina, Pavel; Dušek, Ondřej; Hajič, Jan; Urešová, Zdeňka</p> <p>Description: This package contains data sets for development and testing of machine translation of medical search short queries between Czech, English, French, and German. The queries come from general public and medical experts.</p> <p>This item contains 1 file.</p>	<p>PMLTQ::Web</p> <p>Author(s): Sedlák, Michal</p> <p>Description: Simple web build on the top of the PML Tree Query service.</p> <p>This item contains 1 file.</p>
--	---

B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published



Khresmoi Query Translation Test Data

ufal-point.mff.cuni.cz/repository/xmlui/handle/11858/00-097C

Home Repository Treebank Search Tools & Services

Home → View Item

Khresmoi Query Translation Test Data 1.0

[Pecina, Pavel](#) ; [Dušek, Ondřej](#) ; [Hajič, Jan](#) ; [Urešová, Zdeňka](#)

Project URL : <http://khresmoi.eu>

Date issued : 2013-10-10

Please use the following URI to cite or link to this item. You can also access this item by scanning the provided QR Code.

<http://hdl.handle.net/11858/00-097C-0000-0022-D9BF-5>

Description :

This package contains data sets for development and testing of machine translation of medical queries between Czech, English, French, and German. The queries come from general public experts.

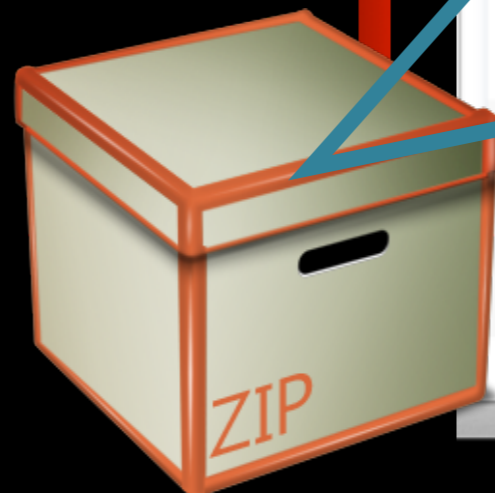
[Show full item record](#) Export to:

Files in this item

This item is **PUB** and licensed under:
[Attribution-NonCommercial 3.0 Unported \(CC BY-NC 3.0\)](#)

B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published
- Save AIP to disk



Khresmoi Query Translation Test Data

ufal-point.mff.cuni.cz/repository/xmlui/handle/11858/00-097C

Home Repository Treebank Search Tools & Services

Home → View Item

Khresmoi Query Translation Test Data 1.0

[Pecina, Pavel](#) ; [Dušek, Ondřej](#) ; [Hajič, Jan](#) ; [Urešová, Zdeňka](#)

Project URL : <http://khresmoi.eu>

Date issued : 2013-10-10

Please use the following URI to cite or link to this item. You can also access this item by scanning the provided QR Code.

<http://hdl.handle.net/11858/00-097C-0000-0022-D9BF-5>

Description :

This package contains data sets for development and testing of machine translation of medical queries between Czech, English, French, and German. The queries come from general public experts.

[Show full item record](#) [Export to...](#)

Files in this item

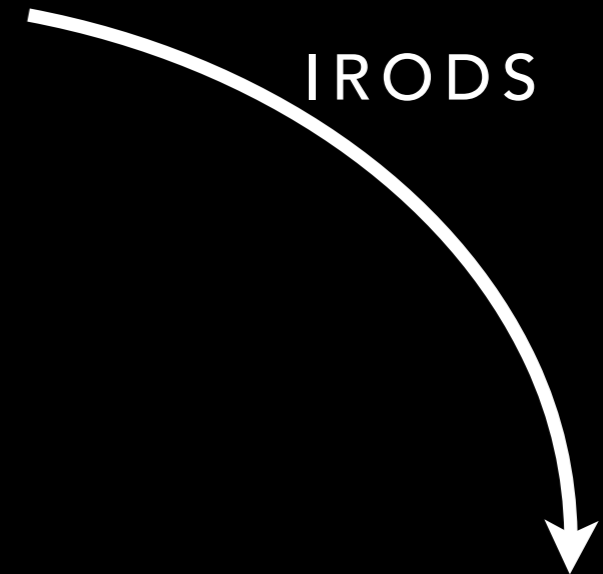
This item is **PUB** and licensed under:
[Attribution-NonCommercial 3.0 Unported \(CC BY-NC 3.0\)](#)

B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published
- Save AIP to disk
- Trigger iRODS copy



iRODS

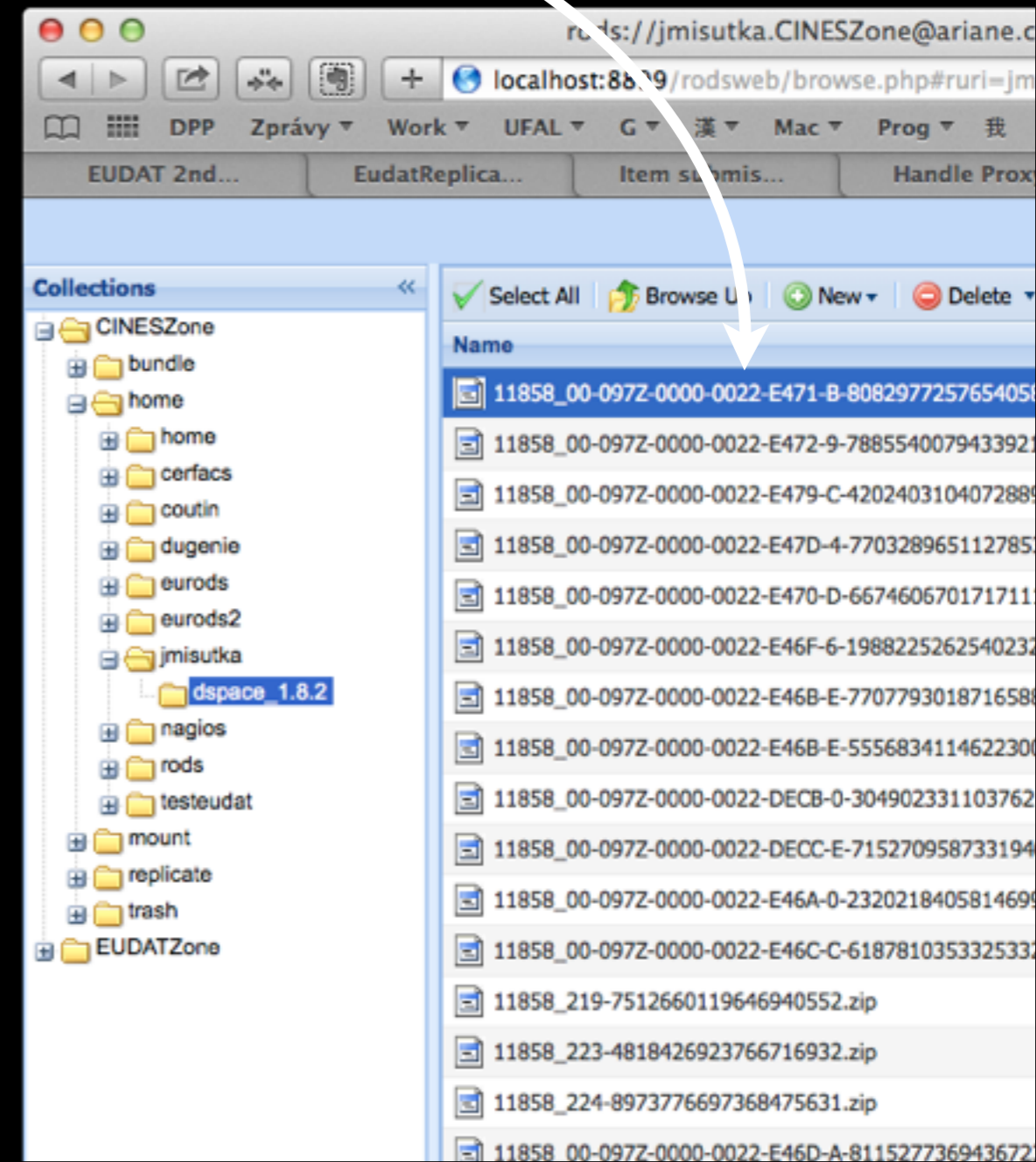


B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published
- Save AIP to disk
- Trigger iRODS copy
- replica in Destination created,
Eudat PID assigned (project internal)

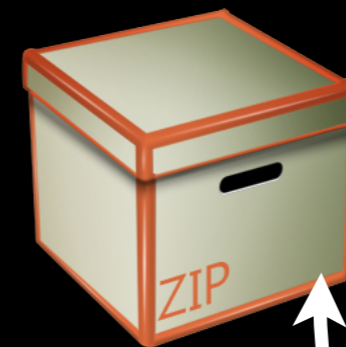


iRODS



B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published
- Save AIP to disk
- Trigger iRODS copy
- replica in Destination created, Eudat PID assigned (project internal)
- Confirmation of transfer, replica PID returned, log the replication



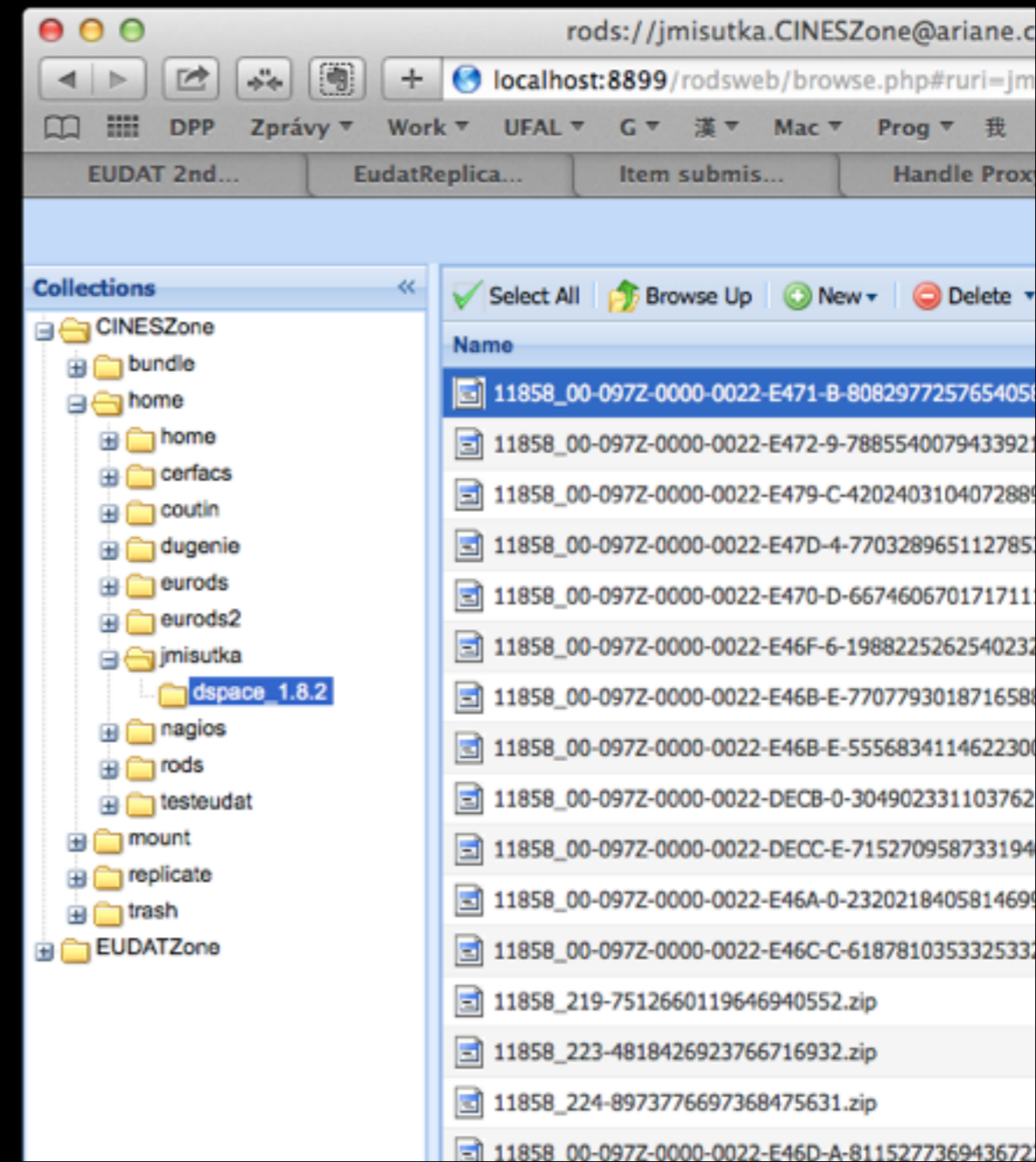
IRODS

EUDAT PID
REPLICA STATUS

Name
11858_00-097Z-0000-0022-E471-B-8082977257654058
11858_00-097Z-0000-0022-E472-9-7885540079433921
11858_00-097Z-0000-0022-E479-C-4202403104072889
11858_00-097Z-0000-0022-E47D-4-770328965112785
11858_00-097Z-0000-0022-E470-D-667460670171711
11858_00-097Z-0000-0022-E46F-6-1988225262540232
11858_00-097Z-0000-0022-E46B-E-770779301871658
11858_00-097Z-0000-0022-E46B-E-555683411462230
11858_00-097Z-0000-0022-DECB-0-304902331103762
11858_00-097Z-0000-0022-DECC-E-715270958733194
11858_00-097Z-0000-0022-E46A-0-2320218405814699
11858_00-097Z-0000-0022-E46C-C-618781035332533
11858_219-7512660119646940552.zip
11858_223-4818426923766716932.zip
11858_224-8973776697368475631.zip
11858_00-097Z-0000-0022-E46D-A-811527736943672

B2SAFE WITH DSPACE

- Submit Data and Metadata (SIP)
- Approve the Record (Dspace Editor)
- The record with a PID published
- Save AIP to disk
- Trigger iRODS copy
- replica in Destination created, Eudat PID assigned (project internal)
- Confirmation of transfer, replica PID returned, log the replication
- Remove the local AIP copy



EUDAT B2SAFE SOLUTION FOR DSPACE

- Easy to setup
- Fully integrated in Dspace Control Panel
- Completely transparent in day-to-day Dspace use
- Well documented
- Open Source and Free



B2SAFE
Replicate Research Data Safely

LINKS + CREDITS

- LINDAT-Clarin B2SAFE replication documentation:
 - <https://svn.ms.mff.cuni.cz/redmine/projects/dspace-modifications/wiki/EudatReplication>
- Dspace documentation:
 - <https://wiki.duraspace.org/display/DSDOC18/Functional+Overview>
 - <https://wiki.duraspace.org/display/DSDOC18/AIP+Backup+and+Restore>

Zip box icon: <http://www.iconarchive.com/artist/babasse.html>, Fire: <http://www.flickr.com/photos/mattbraga/4044016187/>, Flood: <http://www.flickr.com/photos/jewishwomensarchive/4920004375/>