

# Data Discovery and Reuse with EUDAT Services

## Introduction

---

This document shows how EUDAT services facilitate data discovery and reuse according to FAIR principles. In particular we show the integration between B2FIND and B2SHARE for data discovery. We also present a demo on integration of community-specific tools with certain EUDAT services which facilitates access and reuse of research data. In what follows we give a general introduction of the EUDAT services is being used in the demos:

**B2FIND** is the EUDAT metadata catalogue and provides a data discovery portal, which allows users to quickly search and find research data. All EUDAT internal metadata is harvested by B2FIND. Meta data from external providers will also be harvested in B2FIND (via OAI-PMH protocol). B2FIND service can be accessed through <http://b2find.eudat.eu/> and no login credentials is required.

**B2SHARE** is the EUDAT data repository service to store, preserve and publish research data. All metadata that is published in B2SHARE is open access. You can choose for the data to be private, but the metadata is always public. To access the service you can login with your B2ACCESS credentials at <https://b2share.eudat.eu>.

**B2DROP** is the EUDAT personalized cloud storage service. It can be used as a secure and trusted data exchange service. The service is meant to store data in the early stage of research data life cycle, and to exchange and share data with team members. The data can be shared data with fine grained access controls. You can also synchronize the data over different devices and platforms. B2DROP users are offered 20 GB of storage for free. The service can be accessed through <https://b2drop.eudat.eu>. You can login with your B2ACCESS credentials.

## Demo: B2FIND-B2SHARE integration

---

**Step 1:** The metadata published in B2SHARE will be harvested by B2FIND. In the landing page of B2FIND you can quickly and easily search for datasets.

GO TO EUDAT WEBSITE



GUIDELINES ▾

COMMUNITIES

FACETED SEARCH

ABOUT ▾


### Search Your Data

eg. IPCC



**Step 2:** B2FIND also offers faceted search, which you can filter your searches based on time, publication year, community or discipline, language, publisher, ...

**Filter by location** Clear



Map data © OpenStreetMap contributors  
Tiles by Stamen Design (CC BY 3.0)

**785,826 datasets found**

Order by: Relevance ⌵

---

**Filter by time** Clear

Start:

End:

**Publication Year** Clear

to

**Communities** ⌵

**Tags** ⌵

**Creator** ⌵

**Discipline** ⌵

**Language** ⌵

**TRIQS/DFTTools: A TRIQS application for ab initio calculations of correlated ...**

This program has been imported from the CPC Program Library held at Queen's University Belfast (1969-2018) Abstract We present the TRIQS/DFTTools package, an application based...

---

**HYPERgeometric functions Differential REDuction (HYPERDIRE): MATHEMATICA base...**

This program has been imported from the CPC Program Library held at Queen's University Belfast (1969-2018) Abstract HYPERDIRE is a project devoted to the creation of a set of...

---

**HYPERDIRE HYPERgeometric functions Differential REDuction: Mathematica-based ...**

This program has been imported from the CPC Program Library held at Queen's University Belfast (1969-2018) Abstract HYPERDIRE is a project devoted to the creation of a set of...

---

**HYPERDIRE, HYPERgeometric functions Differential REDuction: MATHEMATICA-based...**

This program has been imported from the CPC Program Library held at Queen's University Belfast (1969-2018) Abstract HYPERDIRE is a project devoted to the creation of a set of...

---

**HYPERDIRE—HYPERgeometric functions Differential REDuction: Mathematica-based ...**

This program has been imported from the CPC Program Library held at Queen's University Belfast (1969-2018) Abstract We present a further extension of the HYPERDIRE project,...

---

**A finite difference Hartree–Fock program for atoms and diatomic molecules**

This program has been imported from the CPC Program Library held at Queen's University Belfast (1969-2018) Abstract The newest version of the two-dimensional finite difference...

**Step 3:** As an example workflow, we want to find data records published by a scientist named "Dieter" in the Linguistics discipline. The normal search hits a lot of data records. But in the Faceted Search, you can search for keyword "Dieter" in the Linguistics discipline which narrows down your search.

**Filter by location** [Clear](#)

Map data © [OpenStreetMap](#) contributors  
Tiles by [Stamen Design](#) (CC BY 3.0)

**Filter by time** [Clear](#)

Start: -0342-06-13

End: 1504-12-31 18:20:41

**Publication Year** [Clear](#)

to

**Communities** ▾

**Tags** ▾

**Creator** ▾

**Discipline** ▾

Filter 9-1 ▾

Linguistics (1) ✕

**dieter** 🔍

**1 dataset found for "dieter"** **Order by:** Relevance ▾

Discipline: **Linguistics** ✕

**Orthography-based dating and localisation of Middle Dutch charters**

In this study we build models for the localisation and dating of Middle Dutch charters. First, we extract character trigrams and use these to train a machine learner (K Nearest...

**Step 4:** By clicking on the data record, you can see more detailed information about the data identifiers (such as PID and DOI), provenance (data creator, publication year and rights), representation (language and resource type) and discipline.

B2FIND only contains information about the metadata. To get access to the data itself, you need to visit the relevant data repository that contains the data. The DOI (data object identifier) or the PID (persistent identifier) links in the metadata section contains that information.

**Social**

- Google+
- Twitter
- Facebook

Dataset
Communities

## Orthography-based dating and localisation of Middle Dutch charters

DOI
PID

In this study we build models for the localisation and dating of Middle Dutch charters. First, we extract character trigrams and use these to train a machine learner (K Nearest Neighbours) and an author verification algorithm (Linguistic Profiling). Both approaches work quite well, especially for the localisation task. Afterwards, an attempt is made to derive features that capture the orthographic variation between the charters more precisely. These are then used as input for the earlier tested classification algorithms. Again good results (at least as good as using the trigrams) are attained, even though proper nouns were ignored during the feature extraction. We can conclude that the localisation, and to a lesser extent the dating, is feasible. Moreover, the orthographic features we derive from the charters are an efficient basis for such a classification task.

One file (PDF) contains the text of the master thesis, the other file (.tar.gz) contains all the used data sets and analysis scripts.

Middle Dutch
machine learning
orthography
text mining

Identifier	
DOI	<a href="http://dx.doi.org/doi:10.23728/b2share.b1092be3cd4844e0bffd7b669521ba3c">http://dx.doi.org/doi:10.23728/b2share.b1092be3cd4844e0bffd7b669521ba3c</a>
PID	<a href="http://hdl.handle.net/11304/3720bb44-831c-48f3-9847-6988a41236e1">http://hdl.handle.net/11304/3720bb44-831c-48f3-9847-6988a41236e1</a>
Source	<a href="https://b2share.eudat.eu/api/records/b1092be3cd4844e0bffd7b669521ba3c">https://b2share.eudat.eu/api/records/b1092be3cd4844e0bffd7b669521ba3c</a>
Metadata Access	<a href="http://www.clarin.si/repository/oai/request?verb=GetRecord&amp;metadataPrefix=oai_dc&amp;identifier=oai:b2share.eudat.eu:b2rec/b1092be3cd4844e0bffd7b669521ba3c">http://www.clarin.si/repository/oai/request?verb=GetRecord&amp;metadataPrefix=oai_dc&amp;identifier=oai:b2share.eudat.eu:b2rec/b1092be3cd4844e0bffd7b669521ba3c</a>

Provenance	
Creator	Dieter Van Uytvanck
Publication Year	2017
Rights	info:eu-repo/semantics/openAccess

Representation	
Language	English
Resource Type	Text

**Step 5:** If you click on the PID link in the metadata, you will be forwarded to the data repository that the data object is stored, in this case B2SHARE. In the data record page in B2SHARE, you can find more detailed information and metadata about the data object such as a summary, keywords, Basic metadata and community specific metadata (CLARIN Metadata). If the record is open access, you can also download the data in the Files section.

# Orthography-based dating and localisation of Middle Dutch charters

by [Dieter Van Uytvanck](#)

Jan 13, 2017

Last updated at Jan 11, 2018



**Abstract:** In this study we build models for the localisation and dating of Middle Dutch charters. First, we extract character trigrams and use these to train a machine learner (K Nearest Neighbours) and an author verification algorithm (Linguistic Profiling). Both approaches work quite well, especially for the localisation task. Afterwards, an attempt is made to derive features that capture the orthographic variation between the charters more precisely. These are then used as input for the earlier tested classification algorithms. Again good results (at least as good as using the trigrams) are attained, even though proper nouns were ignored during the feature extraction. We can conclude that the localisation, and to a lesser extent the dating, is feasible. Moreover, the orthographic features we derive from the charters are an efficient basis for such a classification task. One file (PDF) contains the text of the master thesis, the other file (tar.gz) contains all the used data sets and analysis scripts.

**Keywords:** machine learning; text mining; orthography; Middle Dutch;

DOI: [10.23728/b2share.b1092be3cd4844e0bfd7b669521ba3c](https://doi.org/10.23728/b2share.b1092be3cd4844e0bfd7b669521ba3c) Copy

PID: [11304/3720bb44-831c-48f3-9847-6988a41236e1](https://purl.org/urn:nbn:nl:ui:2f-11304-3720bb44-831c-48f3-9847-6988a41236e1) Copy

Files		
Name		Size
<a href="#">ab8d38f2-835b-11e3-b283-005056943408.gz</a>	2 downloads	1.57GB
<a href="#">c617828a-8357-11e3-8ef2-005056943408.pdf</a>	1 download	5.90MB

Basic metadata	
Open Access	True ✓
License	
Contact Email	
Publication Date	
Contributors	
Resource Type	Category Text