



Account ▼

<https://readcoop.eu>[Transkribus \(/transkribus/\) >](#)[How-To Guides \(https://readcoop.eu/transkribus/resources/how-to-guides/\) >](https://readcoop.eu/transkribus/resources/how-to-guides/)

How To Enrich Transcribed Documents with Mark-up

# How To Enrich Transcribed Documents with Mark-up

💡 **Metadata** 📋 Transkribus Expert Client*Last update 1 year ago*[« How-To's overview \(https://readcoop.eu/transkribus/resources/how-to-guides/\)](https://readcoop.eu/transkribus/resources/how-to-guides/)

## ► About Transkribus ()

### Table of Contents



How To Enrich Transcribed Documents with Mark-up

Introduction

[Tagging interface](#)

Create your own tags

Adding tags

Historical letters and abbreviation signs

Illegible text

Deletions

Black out text



Searching for tags

Metadata

Editorial Declaration

Credits

This guide will show you how to add mark-up to documents which are already transcribed in Transkribus. This gives you the opportunity to define persons, places and abbreviations. You can add customized tagging categories and search for individual tags in your documents.

Additional tags can be exported in different formats. More information about the export of tags can be found in the How to Export Documents from Transkribus guide.    
(<https://readcoop.eu/transkribus/howto/how-to-export-documents-from-transkribus/>)

# Introduction

The tagging interface in Transkribus enables you to

- Assign tags to important words or phrases in your document.
- Search for individual tags or tag categories.
- Export the tags you added in different file formats so that you can go on working with them outside of Transkribus.

## Tagging interface

- The tagging interface can be found by clicking the “Metadata” tab, and then the “Textual” tab.

READcoop Account (https://readcoop.eu)

Server Overview Layout **Metadata** Tools

Document Structural **Textual** Comments

### Tags of current Transcript

	Tag	Value	Text	Properties
1	organization	Kom̄it	erscheinenden Kom̄ittenscha	
2	abbrev	Kom̄it	erscheinenden Kom̄ittenscha	expansion: Kommittens
3	abbrev	ernsth	Kom̄ittenschaften zur ernsth	expansion: ernsthaften
4	abbrev	empfo	ernsthaft überlegung empfo	expansion: empfohlen
5	textStyle	n	überlegung empfahl werde	strikethrough: true
6	textStyle	möge	empfohl werden möge, da	strikethrough: true
7	abbrev	lezthin	jene beÿ beeden lezthinig k	expansion: lezthinigen
8	organization	Kongr	beeden lezthinig Kongresse	
9	organization	Städt	Zusage, lokern, den Städt, G	
10	abbrev	Städt	Zusage, lokern, den Städt, G	
11	organization	Gerich	lokern, den Städt, Gerichten	
12	abbrev	u.	Städt, Gerichten, u. Gemein	expansion: und
13	organization	Geme	Gerichtern, u. Gemeinden be	
14	abbrev	erheÿr	ledig - oder verheÿrath Stan	expansion: verheÿrather
15	abbrev	un	ein art einer un gezwunger	

< >

### Tags

**All Tags** ☒ Show all Customize...


Tag specification	Color	Shortcut
abbrev		
add		
Address		
blackening		
date		
...		

**Props for tag: no tag selected**


Previous
 Next
 Apply to selected



Figure 1 The "Textual" tab

-  Account ▼
- If you put a tick at "Show all" at the bottom of the "Textual" tab, all the predefined tags (<https://readcoop.eu>) will be shown. You can start working with these right away.

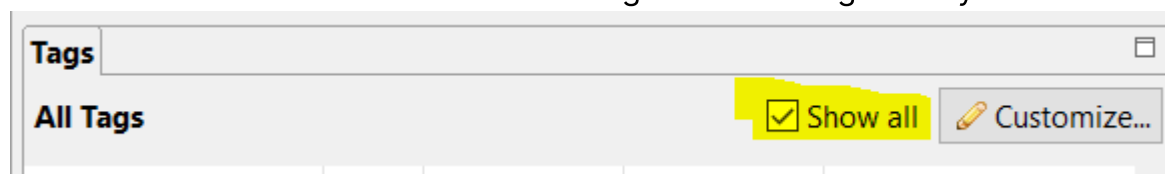


Figure 2 Show all predefined tags

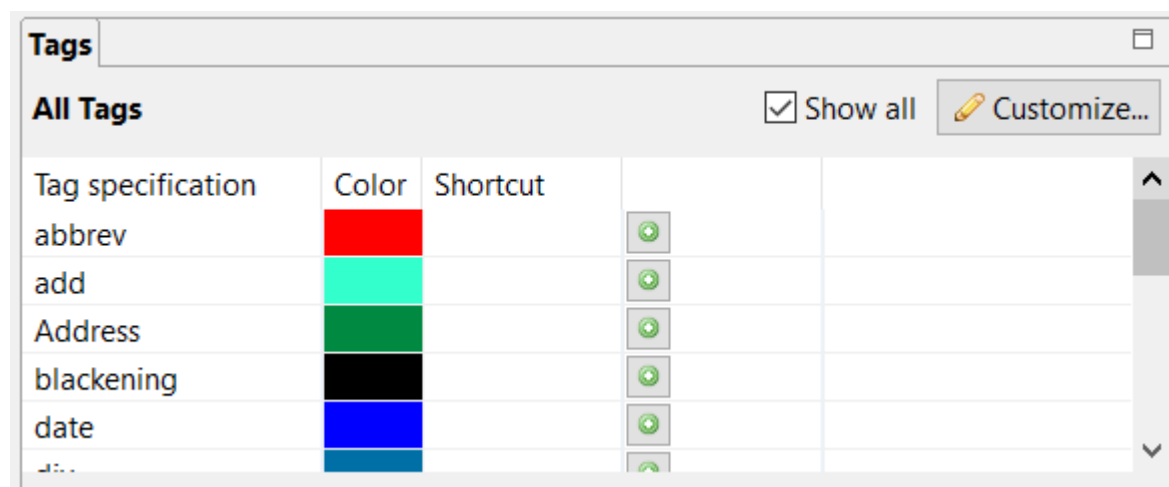


Figure 3 Predefined tags in Transkribus

## Create your own tags

- To create your own tag categories, click the "Customize" button in the "Tags" tab. The "Tag configuration" window will open up.

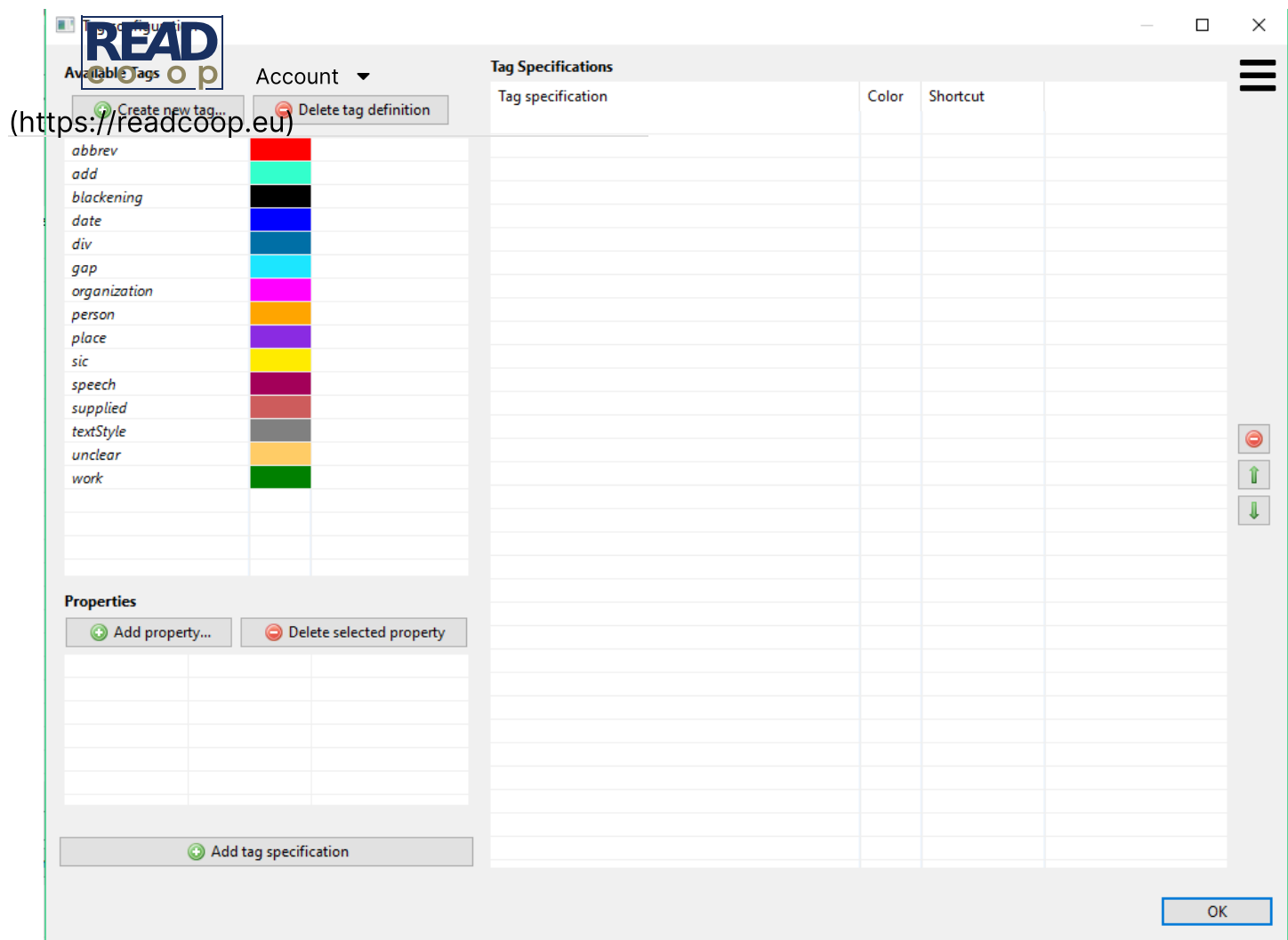


Figure 4 Create your own tags

- With the “Create new tag” button you can add your own tags.
- Once you have created a new tag, it will appear when you click the “Show all” button.
- In the “Tag configuration” window predefined tags are shown in italics, customized ones are shown without italicisation.

## Adding tags

- If you want to tag a word or phrase there are three ways (at least) to do it:
  - Highlight the text in the Text Editor field and afterwards click on the green + button of the tag you want to apply.

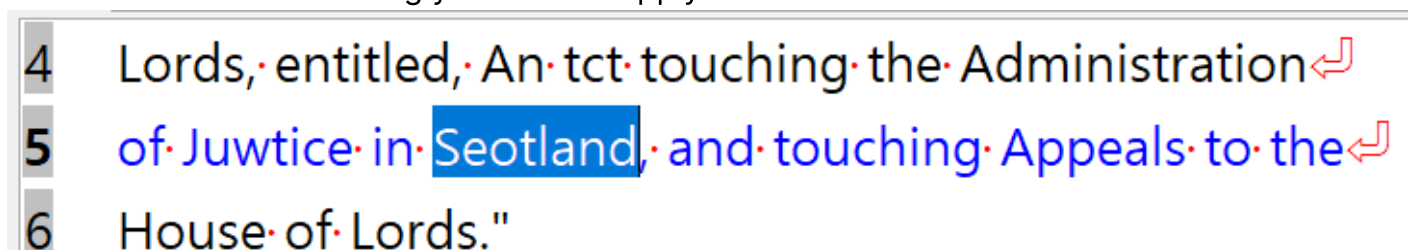


Figure 5 Highlight the word to be tagged

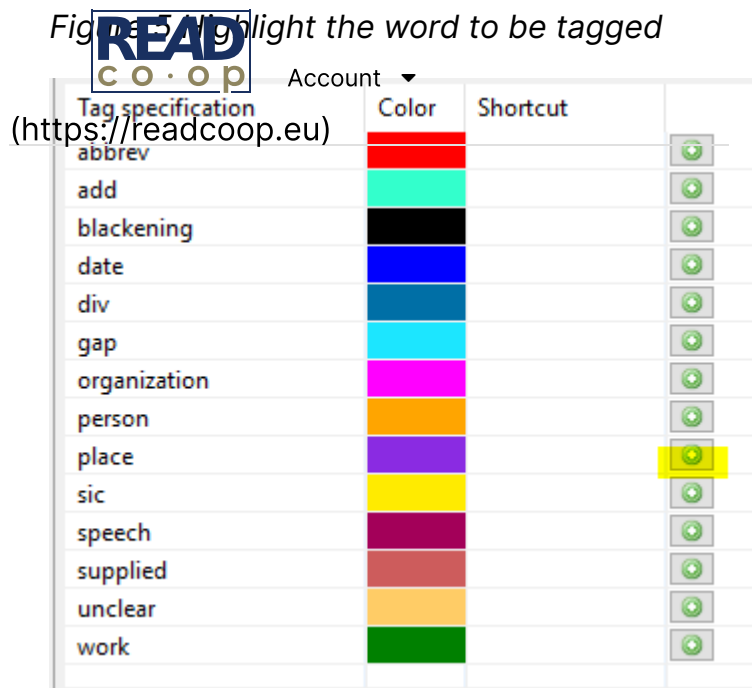


Figure 6 Choosing the right tag

- Alternatively, you can highlight the word or phrase and then make a right click with your mouse. Under “All tags” the suitable one can then be chosen.

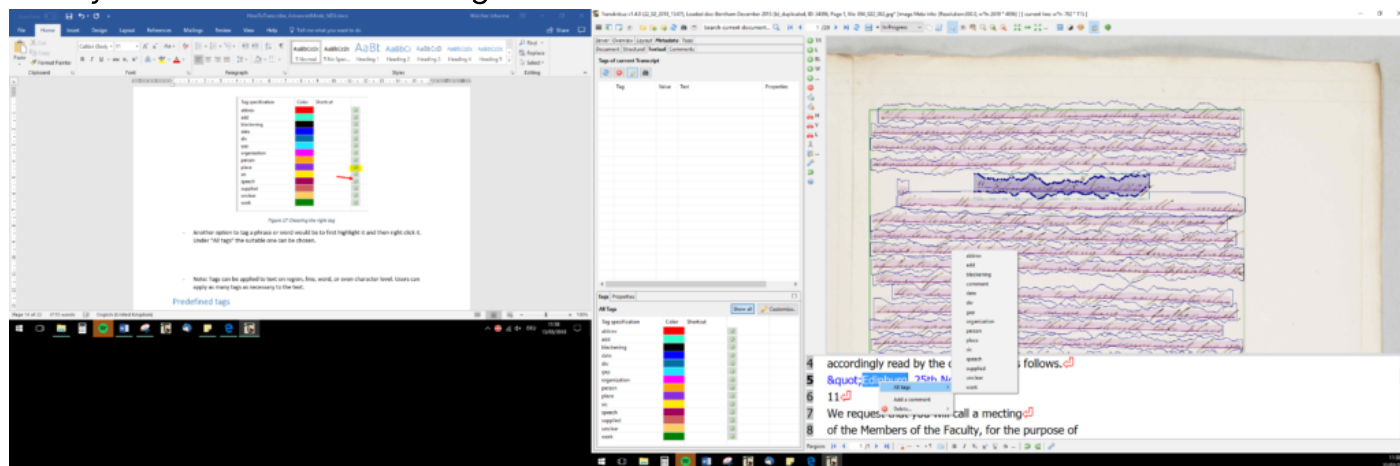


Figure 7 Tag a word or phrase with right mouse click

- Finally, if there are tag categories you use frequently, you can create a shortcut for them in order to speed up your work. To do so, within the “Textual” tab, click the “Customize” button in the “Tags” tab. In the “Tag Specifications” section, you can now add your preferred shortcut in the “Shortcut” column.

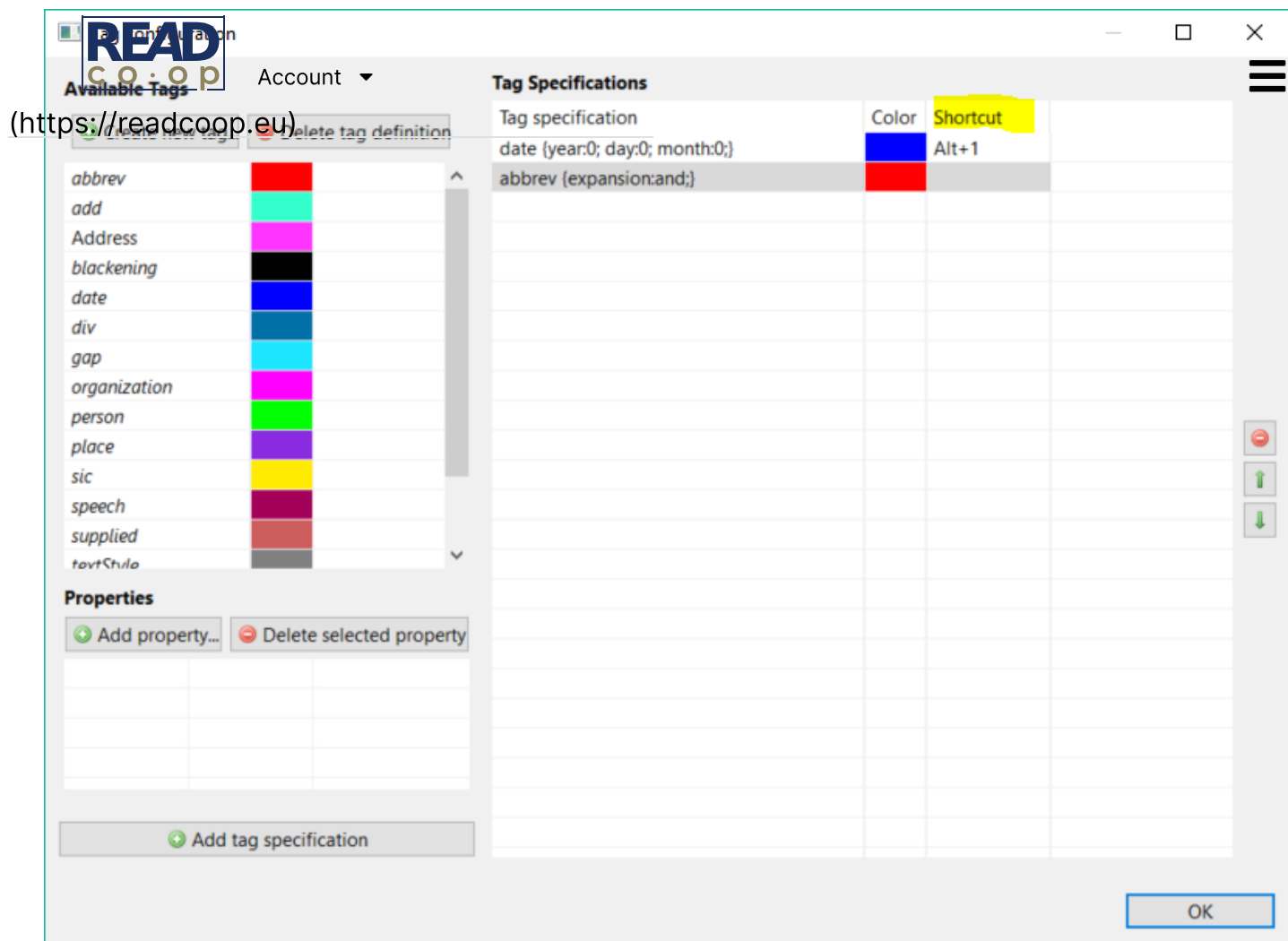


Figure 8 Add shortcuts for frequently used tags

- You can also add a shortcut relating to the properties of your tags, e.g. for expanding abbreviations or adding a standardised country name to a place tag.
  - Click the “Customize” button in the “Tags” tab.
  - In the “Tag configuration” window click the desired tag. The details relating to that tag will appear in the “Properties” section.
  - Click “Add property” to add the property you would like.
  - Then click “Add tag specification”
  - Now your tag and its property (e.g. an expansion for an abbreviation) will appear in the “Tag Specification” section of the window.
  - Add the shortcut you would like to use.
  - Now you can add the tag and its property by simply highlighting the word or phrase in the Text Editor field and then pressing the short cut.

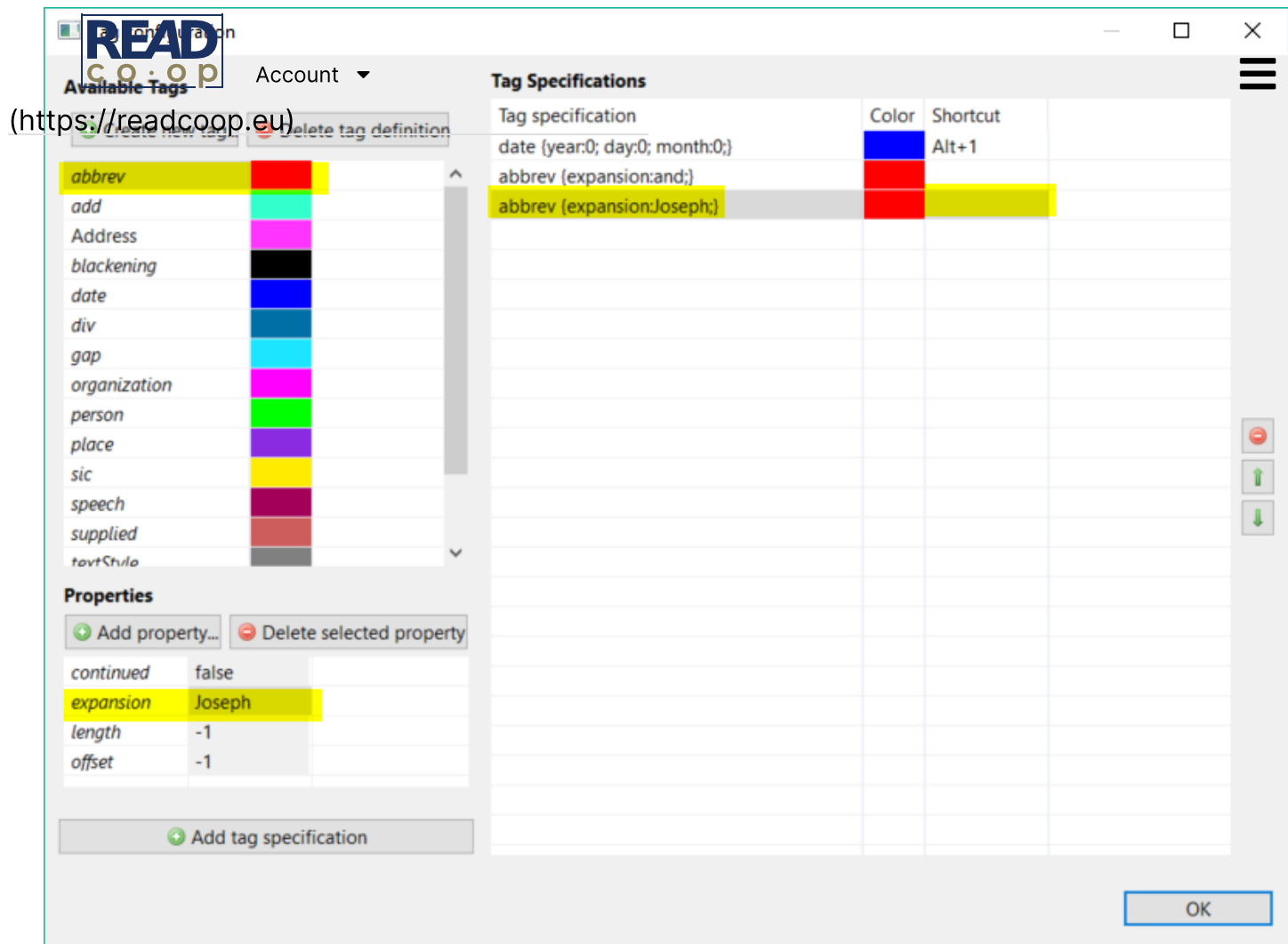


Figure 9 How to add a fixed abbreviation

- If you tagged something by mistake you can undo it by highlighting the word or phrase again, right clicking with your mouse and then pressing the “Delete” button. The program will give you two options:
  - Delete only the highlighted tag
  - Delete all the tags for the current selection
- Note: Tags can be applied to text on region, line, word, or even character level. To apply tags to a segmentation element, click on a text or line region in the Canvas image viewer and follow the above instructions.
- Users can apply as many tags as necessary to the text.
- In the “Textual” tab Transkribus will give you an overview of the tags you have put in your document.



READ Account (https://readcoop.eu)

Tags of current Transcript

	Tag	Value	Text	Properties
11	abbrev	ds	Bertolini selbst für ds Pferd 2 f 30 Xr	
12	abbrev	f	für ds Pferd 2 f 30 Xr an- gebothen	
13	abbrev	Xr	ds Pferd 2 f 30 Xr an- gebothen hat	
14	abbrev	f:	habe, und daß um 20 f: zu wenig Ge	
15	organization	Magistr	satze, daß der Magistrat von von dei	
16	abbrev	löß.	hoffe, massen von löb. Kreisamt un	
17	organization	Kreisam	massen von löb. Kreisamt unterm 1	
18	date	14t 9br	Kreisamt unterm 14t 9br abhin und	
19	textStyle	t	Kreisamt unterm 14t 9br abhin und	superscript:
20	abbrev	t	Kreisamt unterm 14t 9br abhin und	
21	abbrev	9br	Kreisamt unterm 14t 9br abhin und	
22	organization	Militar	durchgehends von dem Militar Ver	
23	abbrev	u.	Vertheilungsmagazin besorget, u. vi	
24	abbrev	u	für den Staab u. für die Kompagnien	
25	abbrev	werd.	Kompagnien gefodert werd.	
26	abbrev	H'r	H'r v Mayrl. No 1459.	
27	person	v Mayrl	H'r v Mayrl. No 1459.	
28	abbrev	v	H'r v Mayrl. No 1459.	
29	abbrev	No	H'r v Mayrl. No 1459.	
30	textStyle	o	H'r v Mayrl. No 1459.	superscript:
31	organization	Gubern	Gubernial-Circulare d. d.	
32	abbrev	d. d	Gubernial-Circulare d. d. 4t, præ. 2t	
33	date	4t	Gubernial-Circulare d. d. 4t, præ. 2t	

Figure 10 Overview of tags

## Historical letters and abbreviation signs

- In modern documents the handling of abbreviations is less important, but in historical documents it is a complex and challenging task.
- In earlier time periods words were often heavily abbreviated, in the hope of writing faster or saving paper. In some documents more than 20 or 30% of all words are abbreviated as shown in the figure below:

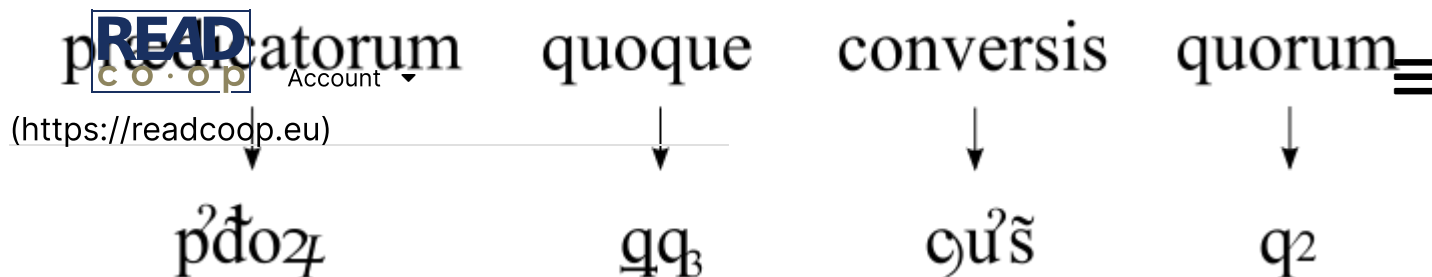


Figure 11 Examples of typical abbreviations in Latin text of the Middle Ages  
(cf. Wikipedia: [https://en.wikipedia.org/wiki/Scribal\\_abbreviation](https://en.wikipedia.org/wiki/Scribal_abbreviation))

- Again there are two main options to transcribe abbreviated text:
  - **Option 1:** Expand abbreviations in the usual way. Neural networks are often able to learn to recognise and reproduce expansions. E.g. Latin prefixes and suffices such as “cum”, “con” or “us” and “orum” are learned easily by the machine. This means that you just need to provide an expanded version of the text in your transcription.
  - **Option 2:** Keep to the rule mentioned above– as long as you can recognize the base character – **transcribe the base character**. This rule is especially suited to historians and people interested in the “content” of a document and those who want to provide training data for the HTR engine.
    - Note: When it comes to HTR training, tags are not relevant yet. Developments in Named Entity Recognition technology should make the automated recognition of tags possible in the future.
  - Therefore the correct transcription for the examples above would be simple:
    - **pdor – qq – cus – qr**
    - Note: In the future HTR engines may also learn to automatically expand these abbreviations (or to supply the correct abbreviation for an expansion) so that computer assisted transcription may be supported.
  - **Option 3:** If you are also interested in using Unicode characters which are near to the **special graphemes** of the original document, then you can transcribe the text by utilizing the full power of Unicode.
  - In this case the transcription of above could look like the following:
    - p̃: LATIN SMALL LETTER P COMBINING OGONEK ABOVE
    - đ: LATIN SMALL LETTER D WITH MIDDLE TILDE
    - o: LATIN SMALL LETTER O
    - Ț : LATIN SMALL LETTER RUM ROTUNDA. Also LATIN SMALL LETTER R ROTUNDA may be used to represent this letter.
  - **Note:** In real-world cases it is often hard to decide which diacritic, modifier letter or Unicode character may be the right one. You may consult the MUFI website to



For more information on this issue (cf. section “References”):

<http://folk.uib.no/hnooh/mufi/> (<http://folk.uib.no/hnooh/mufi/>)

(<https://readcoop.eu>) Unicode and other special characters can be found in the “Virtual keyboards” button in the Text Editor menu.



Figure 12 “Virtual” keyboards button



Figure 13 “Virtual keyboards” window

- Of course **mixed models will often be useful**. E.g. frequently occurring historical characters may be transcribed with their correct Unicode letter, whereas characters which were used just by a specific writer may be transcribed with their base character. You should note such editorial decisions in the “**Editorial Declaration**” in the “Document” tab, within the “Metadata” tab so that your transcription rules are transparent to other users.

- **Example:** LATIN SMALL LETTER RUM ROTUNDA **ŕ** is regularly used in medieval and early modern texts. Therefore it might be useful to introduce this letter to an HTR model which deals exclusively with medieval documents and is dedicated to processing large amounts of such documents.

(<https://readcoop.eu>) cannot be transcribed since it is illegible can be marked with the tags “unclear” or “gap”.

- If the text is unclear, highlight it in the text editor field and tag it as “unclear”.
- If text is impossible to read, click your cursor where the text appears in the text editor field and add the “gap” tag.
- You may also add alternatives or suggestions for the illegible word in the “Properties” section of the tag.

## Deletions

- If you discover deleted text you have several options:
  - **Option 1:** The text which is deleted is **still readable**, or at least large portions are readable. In this case transcribe the text as well as possible and mark it as strike through. You can find the “strike through” button in the Text Editor menu.

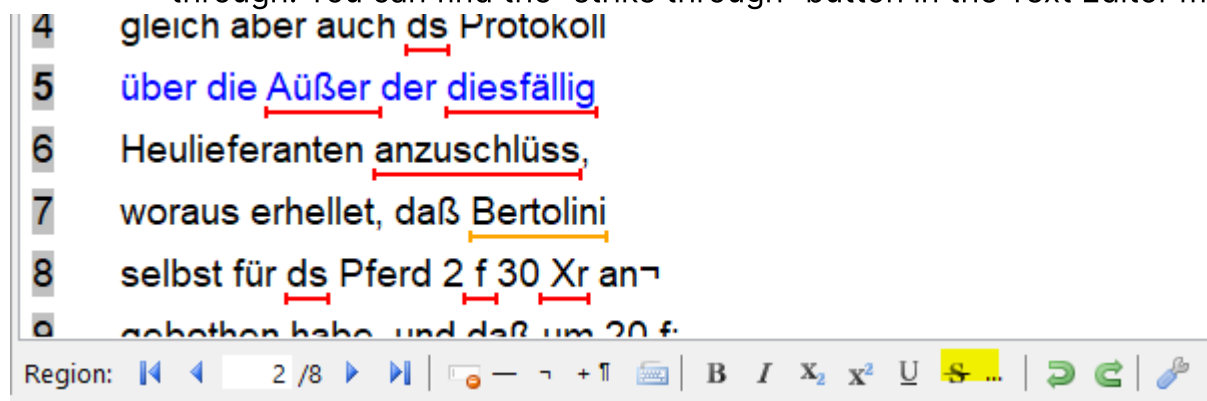


Figure 14 Strike through button

- Note: HTR engines are able to decipher strike through text and the more examples they have, the better.

## Black out text

- The “blackening” tag can be used to redact sensitive information in the export formats. Typically this is used to hide personal data in a document which is made publicly available.
- The blackening tag is used in conjunction with the “blackening” region which must be added with the segmentation tools.
- To blacken part of your text:



Use the drop down menu on the “+...” segmentation element button on the Main menu and select “Blackening”. Use the “Blackening” region to mark the word or section that you want to hide.

- Note: Click the “Item visibility” button on the Main menu and select “Render blackenings” to display the blackened sections on a page.
- Highlight the corresponding word in the Text Editor field and select the “Blackening” tag. In the export of the document the text will be replaced by: [...].
- When you export your document, make sure that “Do blackening” is selected.
- Note: In METS and TEI files the word or phrase is blacked out but the information behind the blackened section is kept. In other file formats, the text behind the blacked out section is completely obscured.

Figure 15 Select “Do blackening” to hide image regions and text in exported files

SEARCHING for tags

Account ▼



(<https://readcoop.eu/>) you need to search for distinct tags click the binoculars button in the “Textual” tab.

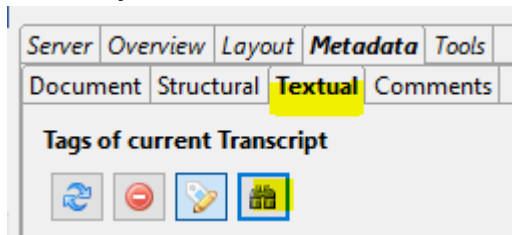


Figure 16 Binoculars button for tag search

- In the window which will open up you can define your search
  - Choose where you would like to search (current collection, current page...)
  - Line or word level
  - In the “Name” field put the name of the tag
  - In the “Text” field put the written text
  - Press the “Search!” button
  - The search results will appear at the bottom of the window.

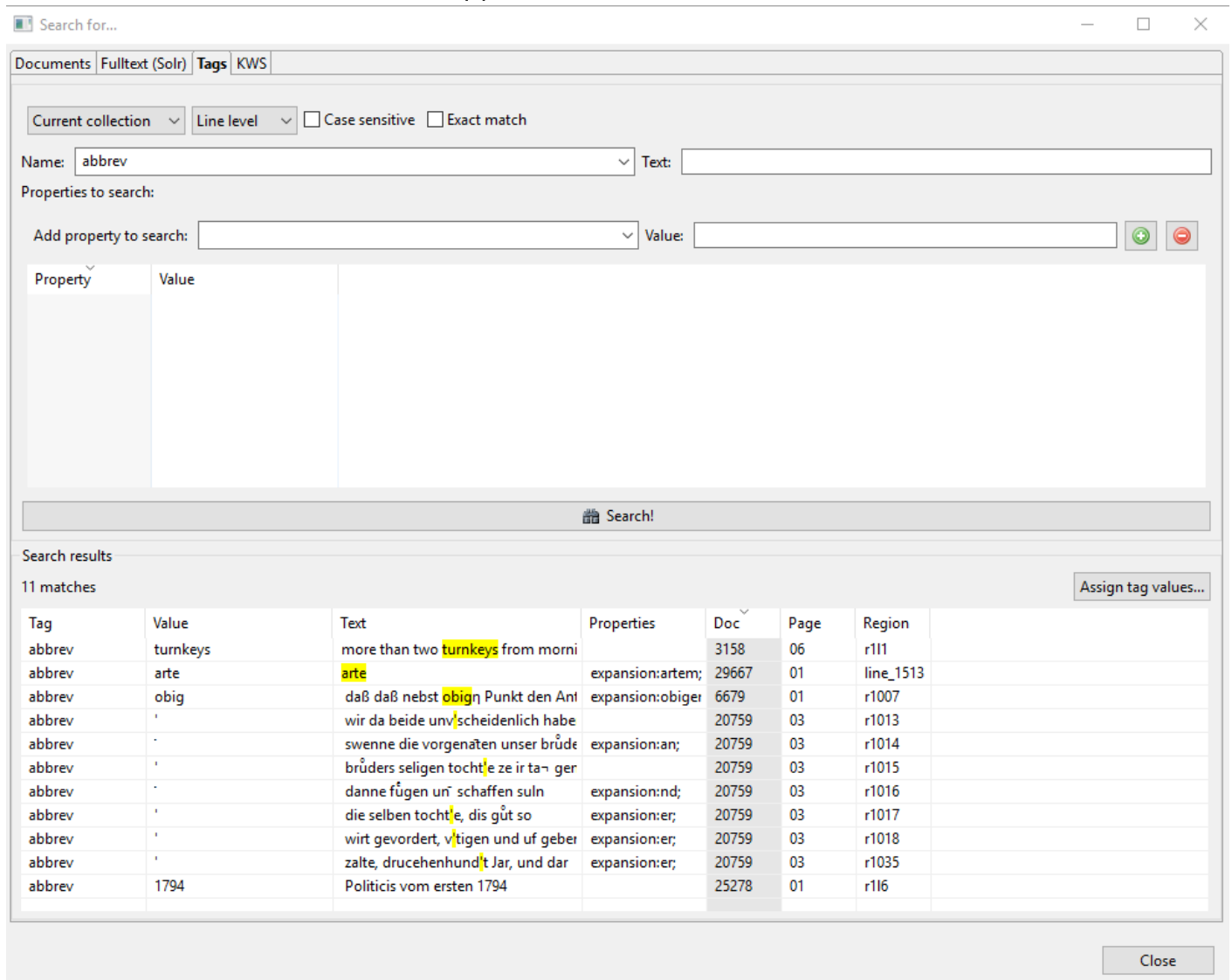


Figure 17 "Search for.." window for tag search

READcoop Account ▼  
 • To quickly add an expansion or another property to a word which appears several times in the text:

- Sort the searching results by "Value". This is done by simply clicking on "Value".
- Mark the similar words by clicking them while holding the "Control" button on your keyboard.
- Then click the "Assign tag values..." button and type in the property that should be added.

Search for...

Documents Fulltext (Solr) **Tags** KWS

Current collection ▼ Line level ▼ ☐ Case sensitive ☐ Exact match

Name: abbrev ▼ Text:

Properties to search:

Add property to search:  Value:

Property	Value

Search results

11 matches

Tag	Value	Text	Properties
abbrev	*	wir da beide unv'scheidenlic	
abbrev	*	brüders seligen tocht'e ze ir	
abbrev	*	die selben tocht'e, dis güt sc	expansion:er;
abbrev	*	wirt gevordert, v'tigen und u	expansion:er;
abbrev	*	zalte, drucehenhund't Jar, un	expansion:er;
abbrev	1794	Politicis vom ersten 1794	
abbrev	arte	arte	expansion:artem;

Properties for selected tags:

Property	Value
continued	false
expansion	
length	-1
offset	-1

Figure 18 Speeding up your work by adding properties to more words or phrases at the same time

(<https://readcoop.eu/>) We are currently supporting only a very simple description of documents since we assume that in a Digital Edition most of the metadata would reside on an external server and be linked to the document. Every document has its unique ID and can be accessed also via the REST services provided by the Transkribus platform (<https://transkribus.eu/wiki/> (<https://transkribus.eu/wiki/>)).

- The following fields are currently available in the “Document” tab, within the “Metadata” tab:
  - Title
  - Author
  - Uploaded
  - Genre
  - Writer
  - Language
  - Script type
  - Date of writing
  - Description

## Editorial Declaration

- Since there are always several ways to produce a correct transcript of a text it is important to be transparent about the way in which the transcription was undertaken.
- For this purpose we have included a special feature in Transkribus, called “Editorial Declaration”. This is found in the “Document” tab, within the “Metadata” tab.
- As with the tagging system, the “Editorial Declaration” offers a set of predefined features and options. Moreover you are able to create your own descriptions and to store them together with your document.
- It is especially important to list special characters and their use in the Editorial Declaration using the form:
  - Character Set Extension: LATIN SMALL LETTER LONG S (U+017F)



**READ** **COOP**

Layout **Metadata** Tools

Document Structural Textual Comments

Account

(<https://readcoop.eu>)

Title:

Author:

Uploaded:

Genre:


Writer:


Language:

- ☐ Arabic
- ☐ Basque
- ☐ Bulgarian
- ☐ Catalan
- ☐ Croatian

Script type:

Date of writing:

☒ From:  

☒ To:  

**Editorial Declaration...**

Description:

Figure 19 Create your Editorial Declaration

## Credits

We would like to thank the many users who have contributed their feedback to help improve the Transkribus software.

## The COOP

About us (<https://readcoop.eu/about/>)

Join us! (<https://readcoop.eu/join/>)

Our Members (<https://readcoop.eu/members/>)

Success Stories (<https://readcoop.eu/success-stories/>)

Work with us (<https://readcoop.eu/work-with-us/>)



Account ▼



(<https://readcoop.eu>)

## Products & Services

Transkribus (<https://readcoop.eu/transkribus/>)

Transkribus lite (<https://transkribus.eu/lite>)

Read&search (<https://readcoop.eu/readsearch/>)

ScanTent (<https://readcoop.eu/scantent/>)

## Useful information

News (<https://readcoop.eu/news/>)

Download Transkribus (<https://readcoop.eu/transkribus/download/>)

Public Models (<https://readcoop.eu/transkribus/public-models/>)

Payment and shipping (<https://readcoop.eu/payment-and-shipping/>)

## Helpful resources

Resource center (<https://readcoop.eu/transkribus/resources/>)

How-to Guides (<https://readcoop.eu/transkribus/resources/how-to-guides/>)

Getting started with Transkribus (<https://readcoop.eu/transkribus/start/>)

FAQs (<https://readcoop.eu/transkribus/questions/>)

Videos (<https://readcoop.eu/transkribus/resources/videos/>)





Account ▼



[Terms & conditions \(https://readcoop.eu/terms-and-conditions/\)](https://readcoop.eu/terms-and-conditions/)  
[Privacy Policy \(https://readcoop.eu/privacy-policy/\)](https://readcoop.eu/privacy-policy/)  
[Contact \(https://readcoop.eu/contact/\)](https://readcoop.eu/contact/)  
[Imprint \(https://readcoop.eu/imprint/\)](https://readcoop.eu/imprint/)

EN