



HAL
open science

What can you do next? Choice of output and reuse of your transcription

Alix Chagué, Floriane Chiffolleau

► To cite this version:

Alix Chagué, Floriane Chiffolleau. What can you do next? Choice of output and reuse of your transcription. 2023. hal-04247966

HAL Id: hal-04247966

<https://hal.science/hal-04247966>

Preprint submitted on 18 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

What can you do next? Choice of output and reuse of your transcription

Alix Chagué, Floriane Chiffoleau
From Source to Full Text: Workshop on Using Automatic Text Recognition (ATR)
September 8th 2023



ALMAnaCH project-team

Inria



Summary of the presentation

1. Which **output** for your transcription?
2. **Exercise: Export** your transcription
3. **Why and when** do we export?
4. How can your data be **reused**?
 - a. **Sharing** your (gold) data
 - b. **Transforming** your transcription
 - c. **Publishing** your transcription
5. **Resources**

Which output for your transcription?

Which output for your transcription?

- ❑ Three different output formats in eScriptorium
 - ❑ PAGE XML
 - ❑ ALTO XML
 - ❑ Text
- ❑ Two kinds of output
 - ❑ Reproduction of the structure of the transcribed document with information such as coordinates (of the regions or the lines), annotations and text → PAGE XML or ALTO XML
 - ❑ Plain text containing only what was recognized in the segmented lines of the document → Text

→ **Your choice of output will depend on what you want to do with your data**

ALTO XML

- ❑ ALTO XML (Analyzed Layout and Text Object)
- ❑ XML Schema that details technical metadata for describing the layout and content of physical text resources

```
<TextLine ID="eSc_line_3b2a0fba"
  BASELINE="703 102 1507 108" HPOS="701.0" VPOS="73.0" WIDTH="806.0" HEIGHT="69.0">
  <Shape><Polygon POINTS="703 102 701 135 1052 129 1080 139 1082 139 1084 139 1085 139 1106 129 1124 140 1126 140 1128 140
  1129 140 1144 140 1267 133 1413 142 1432 142 1433 142 1439 142 1481 133 1501 142 1507 108 1505 76 703 73 703 102"/></Shape>
  <String CONTENT="Clermont-Créans (Sarthe) 15 Aout 1914" HPOS="701.0" VPOS="73.0" WIDTH="806.0" HEIGHT="69.0"></String>
</TextLine>
<TextLine ID="eSc_line_9f920a1a"
  BASELINE="717 276 827 276" HPOS="717.0" VPOS="245.0" WIDTH="110.0" HEIGHT="82.0">
  <Shape><Polygon POINTS="717 276 717 318 825 327 827 276 825 245 717 247 717 276"/></Shape>
<String CONTENT="N° 1" HPOS="717.0" VPOS="245.0" WIDTH="110.0" HEIGHT="82.0"></String>
</TextLine>
<TextLine ID="eSc_line_db3e5d2e"
  BASELINE="653 411 985 408" HPOS="651.0" VPOS="378.0" WIDTH="334.0" HEIGHT="59.0">
  <Shape><Polygon POINTS="653 411 653 437 985 435 985 408 979 378 651 384 653 411"/></Shape>
<String CONTENT="Mon cher Butler" HPOS="651.0" VPOS="378.0" WIDTH="334.0" HEIGHT="59.0"></String>
</TextLine>
```

PAGE XML

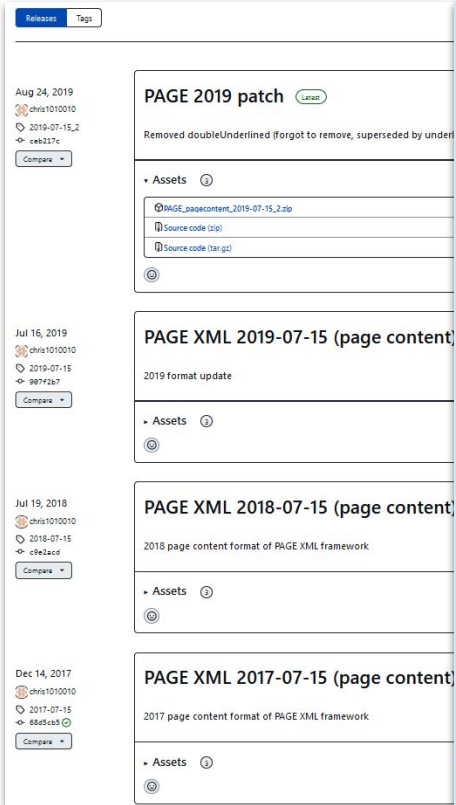
- ❑ PAGE (Page Analysis and Ground-Truth Elements) XML
- ❑ XML-based page image representation framework that records information on image characteristics (image borders, geometric distortions and corresponding corrections, binarisation etc.) in addition to layout structure and page content.

```
<TextLine id="eSc_line_3b2a0fba" >
  <Coords points="703,102 701,135 1052,129 1080,139 1082,139 1084,139 1085,139 1106,129 1124,140 1126,140 1128,140 1129,140
  1144,140 1267,133 1413,142 1432,142 1433,142 1439,142 1481,133 1501,142 1507,108 1505,76 703,73 703,102"/>
  <Baseline points="703,102 1507,108"/>
  <TextEquiv>
    <Unicode>Clermont-Créans (Sarthe) 15 Aout 1914</Unicode>
  </TextEquiv>
</TextLine>
<TextLine id="eSc_line_9f920a1a" >
  <Coords points="717,276 717,318 825,327 827,276 825,245 717,247 717,276"/>
  <Baseline points="717,276 827,276"/>
  <TextEquiv>
    <Unicode>N° 1</Unicode>
  </TextEquiv>
</TextLine>
```

ALTO XML or PAGE XML ?

- ❑ Difference
 - ❑ Pretty similar content and structure
 - ❑ Transformation *alto2page* and *page2alto* are possible
- ❑ Evolution
 - ❑ ALTO changes often
 - ❑ PAGE hasn't changed in a long time

→ **Your choice will depend on the other tools you use.**



The screenshot displays a GitHub repository page for 'PAGE XML' releases. It shows a list of releases with their dates and descriptions. The releases are as follows:

- PAGE 2019 patch** (Latest): Released on Aug 24, 2019. Description: 'Removed doubleUnderlined (forgot to remove, superseded by underlined)'. Assets include 'PAGE_pagecontent_2019-07-15_2.zip', 'Source code (zip)', and 'Source code (tar.gz)'.
- PAGE XML 2019-07-15 (page content)**: Released on Jul 16, 2019. Description: '2019 format update'. Assets include 'PAGE XML 2019-07-15 (page content)'.
- PAGE XML 2018-07-15 (page content)**: Released on Jul 19, 2018. Description: '2018 page content format of PAGE XML framework'. Assets include 'PAGE XML 2018-07-15 (page content)'.
- PAGE XML 2017-07-15 (page content)**: Released on Dec 14, 2017. Description: '2017 page content format of PAGE XML framework'. Assets include 'PAGE XML 2017-07-15 (page content)'.

On the right side of the screenshot, there is a list of schema versions with their release dates:

- > [Schema Version 4.4 released](#) **2023-04-07**
- > [Schema Version 4.3 released](#) **2022-06-08**
- > [Schema Version 4.2 released](#) **2020-07-15**
- > [Schema Version 4.1 released](#) **2019-05-07**
- > [Schema Version 4.0 released](#) **2018-04-11**
- > [Schema Version 3.1 released](#) **2016-01-25**
- > [Schema Version 3.0 released \(version 2.2 renamed\)](#) **2014-08-25**
- > [Schema Version 2.2 Draft Released](#) **2014-07-08**
- > [Schema Version 2.1 Officially Released](#) **2014-02-23**
- > [Schema Version 2.1 Draft Released](#) **2014-01-23**
- > **Schema** **Current version (4.3)**

TEXT

- ❑ Plain text output
- ❑ Content of the (manually or automatically) transcribed lines of your images
- ❑ **Warning:** While the two previous export download directly a ZIP file, the “Text” option does not. You can either click right on “Download” and choose “Save target as” or click left to let the text open in another tab, and, from there, save the text file manually.

```
Clermont-Créans (Sarthe) 15 Aout 1914
N° 1
Mon cher Butler
Je reçois de notre cher Président de la CONCILIATION
ALLEMANDE, le Dr. Nippold, la triste et touchante lettre ci-
jointe. Je ne cesse de penser à lui depuis que la terrible
guerre a éclaté malgré tout ; je n'oublierai jamais son aver-
tissement prophétique, que vous avez entendu comme moi, sur
l'effrayant progrès de la "démagogie militariste allemande"
Quelle pitié que tous ces hommes d'élite réunis
par nous, après tant d'années d'expérience et tant de généreu-
efforts, aient été impuissants contre la folie du militaris-
me et les excitations de la presse! Mais aussi quelle leçon
pour nous, et comme elle nous dicte notre devoir : plus que ja-
mais poursuivre notre oeuvre d'éducation. Nous n'avons pu, jus-
qu'ici, que le préparer, certes avec succès inespéré, mais les
```

Export with eScriptorium's API

- ❑ It's possible to use the API for multiple exports
- ❑ There is a wrapper for the API here:
https://gitlab.com/sofer_mahir/escriptorium_python_connector
- ❑ The API is not yet very documented
- ❑ (No demo of the API, we just want you to know it exists)

Exercise: Export your transcription

Exercise: Export your transcription (ALTO version)

1. Select the images you want to export, then click on “Export”
2. A window will open where you will be able to choose the parameters for your export
3. Choose the right transcription (the one where you made your corrections and that contains the ‘gold’ transcription of your image)
4. Then, choose the “ALTO” option
5. Leave the “Include images” unchecked
6. Click on “Export”
7. A green popup will appear with a link to download the result
8. In your “Downloads” folder, unzip the file; you should have in it the ALTO XML of your image as well as a METS file

Drop images here or click to upload.

Select all Unselect all Selected 0/69

Import Export

Train

Binarize Segment Transcribe

1 X 2 X 3 X 4 X 5 X 6 X 7 X 8 X 9 X

100% 100% 100% 100% 100% 100% 100% 100% 100%

Why and when do we export?

Why and when do we export?

- ❑ To create back-ups of the data (?)
- ❑ To migrate from one server/application to another
- ❑ To move to the next step of the pipeline (the transcription is over)
- ❑ To publish the whole dataset or a sample
- ❑ To feed the data to another tool (like Kraken in command line)

How can your data be reused?



How can your data be reused?

- ❑ By making it **available for others**: **sharing** your training data
- ❑ By making it a more **enriched version**: **transforming** your transcription
- ❑ By making it into a **digital scholarly edition**: **publishing** your transcription

Sharing your (gold) data



Sharing your (gold) data

- ❑ Sharing **ground truth** is essential for the community of users and developers of HTR software
- ❑ A specific type of research output
 - ❑ Gold data: *perfect* transcriptions (according to the project's guidelines)
 - ❑ Sort-of gold data: not 100% perfect (very common)
 - ❑ Silver data: automatically generated
- ❑ You can publish a sample of your corpus as a potential ground truth even if you don't train a model
- ❑ It's not enough to post the data online: it's better when they are **FAIR**

Sharing your (gold) data

Making data FAIR:

- ❑ Follow standards (transcription guidelines, dataset organization, file formats)
- ❑ Describe thoroughly (what do other people need to know to decide the usefulness and trustworthiness of the data?)
- ❑ Write *data papers* ?
- ❑ Reference in catalogs of ground truth for HTR



Findable



Accessible



Interoperable

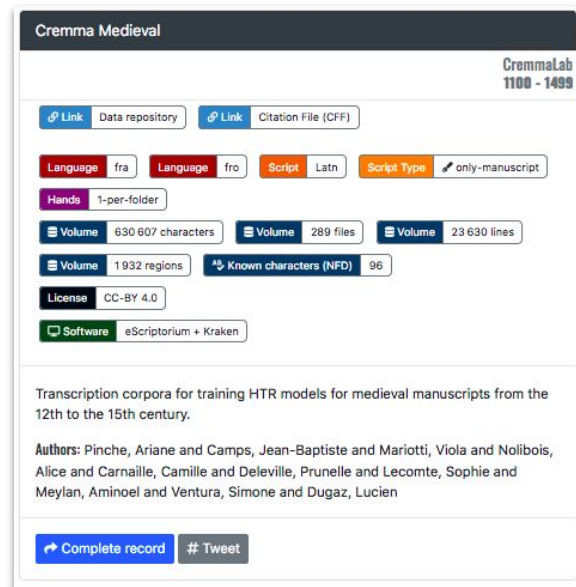


Reusable

Sharing your (gold) data: HTR-United

HTR-United is accessible here: <https://htr-united.github.io/>

- ❑ A catalog to find useful datasets
- ❑ A metadata schema to standardize descriptions
- ❑ A series of tools to control quality of data
- ❑ Recommendations for repository setup

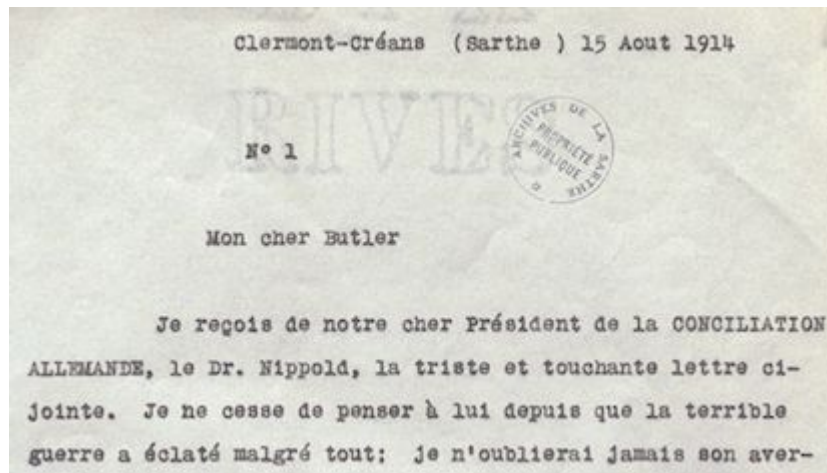
The logo consists of the words 'HTR' and 'United' stacked vertically. 'HTR' is written in a bold, black, sans-serif font, while 'United' is written in a black, cursive script. Both words are set against a light green, textured background that resembles a piece of paper or a brushstroke.A screenshot of the 'Cremma Medieval' web interface. The header shows 'Cremma Medieval' and 'CremmaLab 1100 - 1499'. Below the header are two buttons: 'Link Data repository' and 'Link Citation File (CFF)'. The main content area features several filterable metadata fields: 'Language' (fra, fro), 'Script' (Latn), 'Script Type' (only-manuscript), 'Hands' (1-per-folder), 'Volume' (630 607 characters, 289 files, 23 630 lines), 'Volume' (1932 regions), 'Known characters (NFD)' (96), 'License' (CC-BY 4.0), and 'Software' (eScriptorium + Kraken). A descriptive paragraph follows: 'Transcription corpora for training HTR models for medieval manuscripts from the 12th to the 15th century.' Below this, the authors are listed: 'Authors: Pinche, Ariane and Camps, Jean-Baptiste and Mariotti, Viola and Nollbois, Alice and Carnaille, Camille and Deleville, Prunelle and Lecomte, Sophie and Meylan, Aminoel and Ventura, Simone and Dugaz, Lucien'. At the bottom, there are two buttons: 'Complete record' and '# Tweet'.

Transforming your transcription



Transforming your transcription : TEI XML encoding

- ❑ TEI (Text Encoding Initiative) → standard for the representation of texts in digital form
- ❑ Ensemble of tags, attributes, and values for the metadata and the text of your documents → [TEI Guidelines](#)
- ❑ Several advantages:
 - ❑ Link directly your text to your images (*pb facs*)
 - ❑ Reproduce the structure of the facsimile via the attributes (*rend, place*, etc.)
 - ❑ Enrich your texts with complementary data such as named entities linking, contextual notes, critical editions comments, etc.
 - ❑ Linking one text to another or a whole collection



Facsimile and its transcription

Clermont-Créans (Sarthe) 15 Aout 1914
N° 1
Mon cher Butler
Je reçois de notre cher Président de la CONCILIATION ALLEMANDE, le Dr. Nippold, la triste et touchante lettre ci-jointe. Je ne cesse de penser à lui depuis que la terrible guerre a éclaté malgré tout ; je n'oublierai jamais son aver-

```
<dateline rend="align(right)">  
  <placeName ref="#l0006" type="pec">Clermont-Créans</placeName> (<placeName ref="#l0456" type="pec">Sarthe  
  </placeName>) 15  
  Aout 1914</dateline>  
<title rend="align(center)">N° 1</title>  
<salute rend="indent">Mon cher <persName ref="#p0002" type="pec">Butler</persName>  
</salute>  
</opener>  
<p rend="indent">Je reçois de notre cher Président de la CONCILIATION<lb/> ALLEMANDE, le Dr. <persName ref="#  
#p0475" type="pec"  
>Nippold</persName>, la triste et touchante lettre ci-<lb break="no"/>jointe. Je ne cesse de penser à  
  lui depuis que la  
  terrible<lb/> guerre a éclaté malgré tout ; je n'oublierai jamais son aver-<lb break="no"/>tissement
```


Transforming your transcription : TEI XML encoding

Some tools/scripts are available to speed up and/or ease the process

- ❑ XSLT transformation from ALTO to XML-TEI:
 - ❑ <https://github.com/TEI4HTR/alto2tei>
 - ❑ <https://github.com/kat-kel/alto2tei>
- ❑ XSLT transformation from PAGE to XML-TEI:
<https://github.com/TEI4HTR/page2tei>
- ❑ Text tagging from plain text:
https://github.com/DiScholEd/.../text_tagging.py

Publishing your transcription

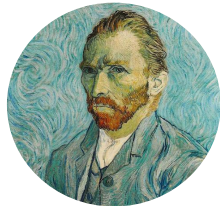


Publishing your transcription: TEI Publisher

- TEI Publisher is ...
 - an easy-to-use tool to publish your TEI XML files
 - a *prêt-à-porter* application customizable with few tweaks
- Few platforms created with TEI Publisher



[When the Wall Came Down](#)



[Van Gogh Letters](#)



[Digital Scholarly Editions](#)



[Antonmaz](#)

The Instant Publishing Toolbox

Log in as **tei-demo** with password **demo** to experiment with your own files and ODDs

Documents

Sort by **Title**

Query scope **Search content**

Enter Query


SEARCH

Genre

- Documentation 2
- Drama 1
- Letters and Correspondence 9
- Lexicons and Dictionaries 2
- Prose 5


Language

- German 4
- deu 3
- English 13
- Spanish 2
- French 1
- grc 1
- Latin 3
- lat 4
- Polish 3
- Ukrainian 1




TEI Publisher Demo Collection

Documents demonstrating various features. The selection is in no way comprehensive but represents a subset of what is possible. Please contribute your own use cases to complement this.




Playground

A writable collection into which you can upload your own documents for testing.




Documentation

Documentation (using Docbook XML).



Annotation Samples











Another playground for annotating documents.



Markdown Sample

An example of using markdown for static content.

ODD Files

- [Serafin Letters](#) 
- 15th c. manuscript correspondence with parallel translation
- [DOCX Output Preview](#) 
- A test ODD to preview TEI imported from Word docx
- [Graves' Letters](#) 
- Modern correspondence with visualization of prosopographical and geographical information
- [ODD_TIMEUS](#) 
- [Dantiscus' Letters](#) 
- 16th c. correspondence with facsimile and encoding of transcriptional features and prosopographical information
- [Erasmus Adagia](#) 
- [Potocki Moralia](#) 
- [Deutsches Textarchiv](#) 
- [EBBE](#) 
- [DOCX Import](#) 

Lettre N° 1 (15 Août 1914)

La Démagogie militariste allemande.
La mobilisation - La France récolte ce qu'elle a semé.
Aux funérailles de Jaurès.

Annexe:

Texte de mon discours aux funérailles de Jaurès.

Clermont-Créans (Sarthe) 15 Aout 1914

N° 1

Mon cher Butler

Je reçois de notre cher Président de la CONCILIATION ALLEMANDE, le Dr. Nippold, la triste et touchante lettre ci-jointe. Je ne cesse de penser à lui depuis que la terrible guerre a éclaté malgré tout ; je n'oublierai jamais son avertissement prophétique, que vous avez entendu comme moi, sur l'effrayant progrès de la "démagogie militariste allemande"

Quelle pitié que tous ces hommes d'élite réunis par nous, après tant d'années d'expérience et tant de généreux efforts, aient été impuissants contre la folie du militarisme et les excitations de la presse! Mais aussi quelle leçon pour nous, et comme elle nous dicte notre devoir : plus que jamais poursuivre notre oeuvre d'éducation. Nous n'avons pu, jusqu'ici, que le préparer, certes avec un succès inespéré, mais les résultats importants que nous avons obtenus si vite et sur tant de points ne sont que des débuts, des germes ; c'est maintenant seulement que le monde civilisé, brutalement éveillé de son ignorance, va comprendre ce que nous avons commencé et sur-

Lettre N° 1 (15 Août 1914)

La Démagogie militariste allemande. La mobilisation - La France récolte ce qu'elle a semé. Aux funérailles de Jaurès.

Annexe:

Texte de mon discours aux funérailles de Jaurès.

Clermont-Créans (Sarthe) 15 Aout 1914

N° 1

Mon cher Butler

Je reçois de notre cher Président de la CONCILIATION ALLEMANDE, le Dr. Nippold, la triste et touchante lettre ci-jointe. Je ne cesse de penser à lui depuis que la terrible guerre a éclaté malgré tout ; je n'oublierai jamais son avertissement prophétique, que vous avez entendu comme moi, sur l'effrayant progrès de la "démagogie militariste allemande"

Quelle pitié que tous ces hommes d'élite réunis par nous, après tant d'années d'expérience et tant de généreux efforts, aient été impuissants contre la folie du militarisme et les excitations de la presse! Mais aussi quelle leçon pour nous, et comme elle nous dicte notre devoir : plus que jamais poursuivre notre oeuvre d'éducation. Nous n'avons pu, jusqu'ici, que le préparer, certes avec un succès inespéré, mais les résultats importants que nous avons obtenus si vite et sur tant de points ne sont que des débuts, des germes ; c'est maintenant seulement que le monde civilisé, brutalement éveillé de son ignorance, va comprendre ce que nous avons commencé et sur-tout ce qui reste à faire.

Je vous écrirai par un prochain courrier spécialement à ce sujet, qui sera pour nous, avant peu, le sujet essentiel. Aujourd'hui je suis encore tout à la [... caractère...] douleur de me sentir im-

– 2 – puissant devant les combats meurtriers engagés en Belgique et sur nombre de points de notre frontière de l'Est et du Nord, en attendant les grandes batailles et les hécatombes qui s'annoncent. Je suis tout à la pensée de ces calamités effroyables et volontairement déchaînées : tant d'existences précieuses et déjà fauchées ou condamnées ; tant de deuils évi-tables qu'on ne peut plus, qu'on ne veut plus éviter ! Il n'y a pas en France une seule famille qui n'ait donné sans hésiter tous ses enfants en âge de combattre pour repousser l'envahis-seur. Notre ami Charles Richet, à lui

Lettre N° 1 (15 Août 1914)

La Démagogie militariste allemande.
La mobilisation - La France récolte ce qu'elle a semé.
Aux funérailles de Jaurès.

Annexe:

Texte de mon discours aux funérailles de Jaurès.

N° 1

Mon cher Butler

Je reçois de notre cher Président de la CONCILIATION ALLEMANDE, le Dr. Nippold, la triste et touchante lettre ci-jointe. Je ne cesse de penser à lui depuis que la terrible guerre a éclaté malgré tout ; je n'oublierai jamais son avertissement prophétique, que vous avez entendu comme moi, sur l'effrayant progrès de la "démagogie militariste allemande"

Quelle pitié que tous ces hommes d'élite réunis par nous, après tant d'années d'expérience et tant de généreux efforts, aient été impuissants contre la folie du militarisme et les excitations de la presse! Mais aussi quelle leçon pour nous, et comme elle nous dicte notre devoir : plus que jamais poursuivre notre oeuvre d'éducation. Nous n'avons pu, jusqu'ici, que le préparer, certes avec un succès inespéré, mais les résultats importants que nous avons obtenus si vite et sur tant de points ne sont que des débuts, des germes ; c'est maintenant seulement que le monde civilisé, brutalement éveillé de son ignorance, va comprendre ce que nous avons commencé et sur-tout ce qui reste à faire.

Clermont-Créans (Sarthe) 15 Aout 1914

teipublisher.odd <>

▼ TEI Publisher ODD ,

Add Element

teiHeader × note × pb ×

+

Jump to element ...

Element Specs

TEI

ab

add

anchor

argument

author

back

bibl

 pb MODE CHANGE


▼ model [webcomponent]

 Use the url from the facs attribute to link with IIIF image
 @facs

▼ model [webcomponent]

 If facs attribute starts with iiif prefix, use the trailing part as a link to the IIIF image
 starts-with(@facs, 'iiif:')


▼ model [break]

[Introduction](#)[Quickstart](#)[Supported XML vocabularies](#)[Processing Model transformations](#)[Custom CSS styling](#)[Page templates and pb-components](#)[Server-side API](#)[Language versions](#)[Creating applications with the App Generator](#)[Best Practice Recommendations](#)[Updating Applications](#)[Data](#)[Facet Search Configuration](#)[Embedding TEI Publisher in other systems](#)[Creating Custom Web Components](#)[Adding a custom vocabulary](#)[Annotating Documents](#)[Configuring the Annotation Editor](#)[TEI Publisher in production](#)[> Introduction](#)

Introduction

What TEI Publisher does ...

The motivation behind TEI Publisher was to provide a tool which enables scholars and editors to publish their materials without becoming programmers, but also does not force them into a one-size-fits-all framework. Experienced developers will benefit as well by writing less code, avoiding redundancy, improve maintenance and interoperability - to just name a few. TEI Publisher is all about *standards*, *modularity*, *reusability* and *sustainability*!

In Publisher, it all starts with your source documents - regardless if they are in TEI or other form of XML: DocBook, MS Word (DOCX) or JATS. No matter how the source material has been encoded, it can be easily transformed into a range of output formats for publication - from a modern web page that you can open on your laptop or mobile device, to an ebook, a PDF file or its LaTeX source.

TEI Publisher derives its name from TEI and the TEI Processing Model (PM). Processing Model is a part of the TEI vocabulary and TEI ODD specification format, described in the [TEI P5 guidelines](#) as well as further chapters here. It defines how a TEI document should be rendered in different output formats and lies at the heart of TEI Publisher.

However, online editions require more than just a text transformation: the text needs to be embedded into an application context, adding navigation, pagination, search, facsimile display and so on. The larger part of TEI Publisher deals with those aspects, providing all the necessary building blocks for an online edition.

Staying true to the spirit of code reuse and interoperability, TEI Publisher implements its functionalities as small "lego" blocks to be freely arranged and recombined. The technology making this possible is called *Web Components*. It is part of the HTML5 specification and natively implemented by many browsers. Users don't need to dive into the details of this standard though: all you need to modify the example pages is some basic HTML knowledge.

Only where the [available components](#) are not enough, a new use case needs to be described and suitable new components implemented but then they can be incorporated into existing component pool for everyone else to use. After all, our mantra is reuse, reuse, reuse and we want to turn TEI Publisher into a box of tools the entire community can benefit from.

Resources

Resources

- ❑ eScriptorium: <https://escriptorium.inria.fr/>
- ❑ HTR-United: <https://htr-United.github.io/>
- ❑ Kraken: <https://kraken.re/main/index.html>
- ❑ PAGE XML:
http://www.primaresearch.org/publications/ICP_R2010_Pletschacher_PAGE
- ❑ TEI Guidelines:
<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- ❑ TEI Publisher:
<https://teipublisher.com/index.html>
- ❑ XML ALTO:
<https://www.loc.gov/standards/alto/>

Thank you for your attention

Any questions ?

Contact: *alix.chague[at]inria.fr*
floriane.chiffoleau[at]inria.fr