

ÉCOLE NATIONALE DES CHARTES  
UNIVERSITÉ PARIS, SCIENCES & LETTRES

---

**Sarah Bénière**

*licenciée ès lettres*

*diplômée de master*

## De l'encodage à la publication :

Les éditions en ligne de la *European  
Holocaust Research Infrastructure* et leur  
chaîne éditoriale



Mémoire pour le diplôme de master

« Technologies numériques appliquées à l'histoire »

2023



# Résumé

Le projet EHRI (*European Holocaust Research Infrastructure*) a pour objectif de promouvoir la recherche sur la Shoah, notamment grâce à la publication en ligne de documents d'archives sous la forme d'éditions thématiques. Le présent mémoire rend compte du développement du modèle d'encodage de ces éditions, à partir des collections déjà publiées, et propose un aperçu de leur publication avec l'outil TEI Publisher.

**Mots-clés :** Édition numérique ; XML-TEI ; ODD ; Modèle d'encodage ; TEI Publisher ; Histoire de la Shoah.

**Informations bibliographiques :** Sarah Bénére, *De l'encodage à la publication : les éditions en ligne de la European Holocaust Research Infrastructure et leur chaîne éditoriale*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Floriane Chiffolleau et Jean-Damien Géro, École nationale des chartes, 2023.



# Remerciements

Je souhaite remercier toute l'équipe d'ALMA<sub>na</sub>CH pour leur accueil, et auprès de qui j'ai beaucoup appris. Je tiens à remercier en particulier Floriane Chiffolleau, ma tutrice, pour son encadrement, ses conseils et son soutien. J'adresse également mes remerciements à mon directeur de mémoire, Jean-Damien Généro, pour son aide précieuse et ses conseils.

Je remercie chaleureusement mes camarades et amies Jade et Maëva, pour toute leur aide, leur soutien et leur présence au cours de ces deux dernières années. J'adresse également mes remerciements à mon entourage d'Aix-en-Provence, en particulier à Prudence, Aurélia, Manon et Jules, pour m'avoir soutenue et encouragée pendant la rédaction de ce mémoire. Je remercie enfin Samuel, mon co-stagiaire, pour sa patience à mon égard et sa bienveillance.



# Liste des sigles et abréviations

## Institutions

**ALMAAnaCH** Automatic Language Modelling and Analysis & Computational Humanities

**CNRTL** Centre National de Ressources Textuelles et Lexicales

**EHRI** European Holocaust Research Infrastructure

**ERIC** European Research Infrastructure Consortium

**Iana** Internet Assigned Number Authority

**Inria** Institut National de Recherche en Informatique et Automatique

## Standards, langages et formats

**HTML** HyperText Markup Language

**IIIF** International Image Interoperability Framework

**ISO** International Organization for Standardization

**Relax NG** Regular Language for XML Next Generation

**TEI** Text Encoding Initiative

**XML** eXtensible Markup Language

**XSLT** eXtensible Stylesheet Language Transformation

## Vocabulaire technique

**CMS** Content Management System (Système de gestion de contenu)

**HTR** Handwritten Text Recognition (Reconnaissance d'écriture manuscrite)

**OCR** Optical Character Recognition (Reconnaissance optique de caractères)

**URL** Uniform Resource Locator





# Sommaire

Résumé	i
Remerciements	iii
Liste des sigles et abréviations	v
Sommaire	v
Bibliographie	ix
Introduction	xiii
I Édition scientifique et numérique	1
1 Le processus d'édition	3
2 Les éditions du projet EHRI	9
3 La diplomatique à l'épreuve du numérique	15
II Modèle d'encodage	19
4 Encodage des éditions EHRI	21
5 Définition du nouveau modèle d'encodage	25
6 Semi-automatisation du modèle	29

<b>III</b>	<b>Publication et valorisation</b>	<b>31</b>
<b>7</b>	<b>Choix de l'outil de publication</b>	<b>33</b>
<b>8</b>	<b>EHRI sur l'application DiScholEd</b>	<b>39</b>
<b>9</b>	<b>Transition vers une application dédiée</b>	<b>43</b>
	<b>Conclusion</b>	<b>47</b>
<b>A</b>	<b>Glossaire</b>	<b>49</b>
<b>B</b>	<b>Chronologie de la persécution des Juifs</b>	<b>53</b>
<b>C</b>	<b>Tableau récapitulatif d'encodage</b>	<b>59</b>
<b>D</b>	<b>Encodage des métadonnées</b>	<b>63</b>
<b>E</b>	<b>Extrait du script de recherche</b>	<b>67</b>

# Bibliographie

## Généralités sur l'édition de textes

MOUNIER (Pierre), « L'édition en ligne : un nouvel eldorado pour les sciences humaines », dans *Read/Write Book, Le livre inscriptible*, dir. Marin Dacos, 2<sup>e</sup> éd., Marseille, 2009, p. 167-176, DOI : [10.4000/books.oep.169](https://doi.org/10.4000/books.oep.169).

NOUGARET (Christine) et PARINET (Élisabeth), *L'édition critique des textes contemporains, XIXe-XXIe siècle*, dir. Christine Nougaret et Élisabeth Parinet, Paris, 2015.

## Pratique de l'édition numérique

CHAGUÉ (Alix) et CHIFFOLEAU (Floriane), « An Accessible and Transparent Pipeline for Publishing Historical Egodocuments », dans The NewsEye International Conference, 2021, URL : <https://hal.science/hal-03180669> (visité le 14/04/2023).

CLAUDAUD (Florence), « L'édition numérique », dans *L'édition critique des textes contemporains, XIXe-XXIe siècle*, dir. Christine Nougaret et Élisabeth Parinet, Paris, 2015, p. 175-200.

FRANKL (Michal), BRYANT (Michael), GREEN (Jessica), SCHELLENBACHER (Wolfgang) et SEDLICKÁ (Magdalena), *Edition of documents*, EHRI GA no. 654164 D.12.2, European Holocaust Research Infrastructure, 2018, p. 1-21, URL : <https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D12%202%20Thematic%20approach%201%20Edition%20of%20documents.pdf> (visité le 14/06/2023).

IDMHAND (Fatima), RIFFARD (Claire) et WALTER (Richard), « L'édition électronique de manuscrits modernes », dans *Expérimenter les humanités numériques*, dir. Étienne Cavalié, *et al.*, Montréal, 2017, p. 105-123, DOI : [10.4000/books.pum.11118](https://doi.org/10.4000/books.pum.11118).

PIERAZZO (Elena), « What Future for Digital Scholarly Editions?, From Haute Couture to Prêt-à-Porter », *International Journal of Digital Humanities*—1 (2019), p. 209-220, DOI : [10.1007/s42803-019-00019-3](https://doi.org/10.1007/s42803-019-00019-3).

SAHLE (Patrick), « What is a Scholarly Digital Edition ? », dans *Digital Scholarly Editing : Theories and Practices*, dir. Elena Pierazzo et Matthew James Driscoll, Cambridge, 2016, p. 19-39, DOI : [10.11647/OBP.0095.02](https://doi.org/10.11647/OBP.0095.02).

## Publication électronique

BOULAIRE (Cécile) et CARABELLI (Roméo), « Du digital naïve au bricoleur numérique : les images et le logiciel Omeka », dans *Expérimenter les humanités numériques*, dir. Étienne Cavalié, *et al.*, Montréal, 2017, p. 81-103, DOI : [10.4000/books.pum.11115](https://doi.org/10.4000/books.pum.11115).

CHARTRON (Ghislaine), « Édition et publication des contenus : regard transversal sur la transformation des modèles », dans *Publier, éditer, éditorialiser, Nouveaux enjeux de la production numérique*, dir. Lisette Calderan, *et al.*, 1<sup>re</sup> éd., Louvain-la-Neuve, 2016, p. 9-35, DOI : [10.3917/dbu.calde.2016.01.0009](https://doi.org/10.3917/dbu.calde.2016.01.0009).

CHIFFOLEAU (Floriane), *Publication of my Digital Edition – Working with TEI Publisher*, Digital Intellectuals, 4 déc. 2020, URL : <https://digitalintellectuals.hypotheses.org/3912> (visité le 17/05/2023).

— *Publication of my Digital Edition – Developing my TEI Publisher Application*, Digital Intellectuals, 16 juin 2021, URL : <https://digitalintellectuals.hypotheses.org/4173> (visité le 17/05/2023).

DAUSSIN (Christine), *Présentation d’Omeka*, Association des usagers francophones d’Omeka, 11 avr. 2019, URL : <https://omeka.fr/document/38> (visité le 30/06/2023).

LEBLANC (Elina), « Omeka Classic, Un environnement de recherche pour les éditions scientifiques numériques », *RIDE*–11 (2020), p. 1-28, DOI : [10.18716/ride.a.11.3](https://doi.org/10.18716/ride.a.11.3).

## Environnement XML et standard TEI

BÉNIÈRE (Sarah), *Writing an ODD for the EHRI Online Editions — Preparatory Work*, Digital Intellectuals, 5 juin 2023, URL : <https://digitalintellectuals.hypotheses.org/4765> (visité le 05/06/2023).

— *Writing an ODD for the EHRI Online Editions — Specifications and Documentation*, Digital Intellectuals, 28 juill. 2023, URL : <https://digitalintellectuals.hypotheses.org/5034> (visité le 28/07/2023).

BURNARD (Lou), *Qu’est-ce que la Text Encoding Initiative ?*, trad. par Marjorie Burghart, 2015 (Encyclopédie numérique), DOI : [10.4000/books.oep.1237](https://doi.org/10.4000/books.oep.1237).

— « What is TEI Conformance, and Why Should You Care ? », *Journal of the Text Encoding Initiative*–12 (2020), *Selected Papers from the 2017 TEI Conference*, p. 1-22, DOI : [10.4000/jtei.1777](https://doi.org/10.4000/jtei.1777).

HAROLD (Elliotte Rusty) et MEANS (W. Scott), *XML in a Nutshell*, 2<sup>e</sup> éd., 2002, URL : <https://docstore.mik.ua/oreilly/xml/xmlnut/index.htm> (visité le 28/08/2023).

## Archivistique et diplomatique

- ALTED VIGIL (Alicia), « Les ego-documents comme source pour l'histoire », trad. par Allison Taillot, *Matériaux pour l'histoire de notre temps*–143-144 (2022), p. 45-51, DOI : [10.3917/mate.143.0045](https://doi.org/10.3917/mate.143.0045).
- BLACHON (Jérôme) et GUEIT-MONTCHAL (Lydiane), « Valorisation des archives », dans *Abrégé d'archivistique*, dir. Lydiane Gueit-Montchal et Dominique Naud, 4<sup>e</sup> éd., Paris, 2020, p. 295-325.
- Maria Milagros Cárcel Ortí (éd.), *Vocabulaire international de la diplomatie*, Commission internationale de diplomatie, 1997, URL : <http://www.cei.lmu.de/VID/> (visité le 19/08/2023).
- DURANTI (Luciana), « Pour une diplomatie des documents électroniques », *Bibliothèque de l'École des chartes*, 161–2 (2003), p. 603-623, DOI : [10.3406/bec.2003.463631](https://doi.org/10.3406/bec.2003.463631).

## Histoire de la Shoah

- BENSOUSSAN (Georges), *Histoire de la Shoah*, 7<sup>e</sup> éd., Paris, 2020 (Que sais-je?), URL : <https://www.cairn.info/histoire-de-la-shoah--9782715404557.htm> (visité le 12/07/2023).
- BRUTTMANN (Tal), *Les 100 mots de la Shoah*, avec la coll. de Christophe Tarricone, 3<sup>e</sup> éd., Paris, 2020 (Que sais-je?), URL : <https://www.cairn.info/les-100-mots-de-la-shoah--9782715404588.htm> (visité le 12/07/2023).
- SAFRIAN (Hans), « L'accélération de la spoliation et de l'émigration forcée, Le “modèle viennois” et son influence sur la politique antijuive du Troisième Reich en 1938 », trad. par Valérie Nguyen, *Revue d'Histoire de la Shoah*, 1–186 (2007), p. 131-163, DOI : [10.3917/rhsho.186.0131](https://doi.org/10.3917/rhsho.186.0131).
- SALLÉE (Frédéric), *Anatomie du nazisme*, 2018 (Idées reçues), DOI : [10.3917/lcb.salle.2018.01](https://doi.org/10.3917/lcb.salle.2018.01).

## Webographie

- DiScholEd, *Éditions scientifiques numériques*, URL : <https://discholed.huma-num.fr/exist/apps/discholed/index.html>.
- EHRI Online Editions, European Holocaust Research Infrastructure, URL : <https://www.ehri-project.eu/ehri-online-editions>.
- Mémorial de la Shoah, URL : <https://www.memorialdelashoah.org/>.
- P5 : *Guidelines for Electronic Text Encoding and Interchange*, version 4.6.0, TEI Consortium, 4 avr. 2023, URL : <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.



# Introduction

« Le monde semble être divisé en deux parties : les endroits où les Juifs ne peuvent pas vivre et ceux où ils ne peuvent pas entrer<sup>1</sup>. »

Le projet EHRI<sup>2</sup> a été lancé en 2010 grâce au financement du septième programme-cadre de l'Union Européenne pour la période 2007-2013 (FP7)<sup>3</sup> dédié à la recherche et à l'innovation. Depuis 2014, le projet bénéficie du financement des programmes-cadres Horizon 2020 (2014-2020)<sup>4</sup> et Horizon Europe (2021-2027).<sup>5</sup> En 2018, l'EHRI est devenue une organisation permanente et souhaite acquérir le statut d'ERIC<sup>6</sup> d'ici à janvier 2025, commémorant ainsi le 80<sup>e</sup> anniversaire de la libération d'Auschwitz.

L'EHRI est une organisation transnationale comptant actuellement vingt-sept partenaires à travers l'Europe, Israël et les États-Unis. L'infrastructure est coordonnée par le *NIOD Institute for War, Holocaust and Genocides Studies*<sup>7</sup>, basé à Amsterdam aux Pays-Bas, et regroupe des centres d'archives, des bibliothèques, des musées et des instituts de recherche. En tant qu'infrastructure de recherche, la principale mission de l'EHRI est de promouvoir la recherche sur la Shoah<sup>8</sup>. En effet, après la Seconde Guerre mondiale, les sources<sup>9</sup> concernant son histoire se sont dispersées et certaines sont devenues difficiles d'accès. L'EHRI a donc créé un portail<sup>10</sup> permettant d'accéder aux informations concernant la conservation des documents liés à la Shoah.

---

1. Georges Bensoussan, *Histoire de la Shoah*, 7<sup>e</sup> éd., Paris, 2020 (Que sais-je?), URL : <https://www.cairn.info/histoire-de-la-shoah--9782715404557.htm> (visité le 12/07/2023), p. 27, citation de Haïm Weizmann.

2. Site de l'EHRI : <https://www.ehri-project.eu/>.

3. Évaluation de la Commission européenne du 25/01/2016 sur le septième programme-cadre : [https://ec.europa.eu/commission/presscorner/detail/en/MEMO\\_16\\_146](https://ec.europa.eu/commission/presscorner/detail/en/MEMO_16_146).

4. Présentation du programme Horizon 2020 sur le site du Ministère de l'Enseignement Supérieur et de la Recherche : <https://www.enseignementsup-recherche.gouv.fr/fr/horizon-2020>.

5. Présentation du programme Horizon Europe : <https://www.horizon-europe.gouv.fr/>.

6. Pour plus d'informations : <https://research-and-innovation.ec.europa.eu/strategy/>.

7. Site du NIOD : <https://www.niod.nl/en>.

8. Pour une définition du terme « Shoah » et l'utilisation que nous en faisons, voir notre glossaire en Annexe A.

9. Sur l'usage du terme « source », l'EHRI rassemble des documents que nous considérons comme des sources dites « primaires », constituant l'objet d'étude des historien-ne-s. En tant qu'ingénieur-e, et dans un souci de neutralité, nous lui préférons le terme « document », qui nous semble plus adapté.

10. Portail de l'EHRI : <https://portal.ehri-project.eu>.

Depuis mai 2015, dans le cadre du programme Horizon 2020, l'équipe-projet ALMAAnaCH<sup>11</sup> d'Inria, apporte son soutien à l'EHRI. Coordonnés par Laurent Romary, les membres de l'équipe affiliés au projet EHRI s'attachent à développer des méthodes et des outils informatiques permettant de faciliter la recherche archivistique sur la Shoah.

Le présent mémoire rend compte de notre expérience au sein de l'équipe ALMAAnaCH sur le projet EHRI. Nous nous intéresserons à la question de l'édition scientifique numérique, en nous appuyant sur les éditions en ligne de l'EHRI. Notre réflexion s'articule autour de trois axes, constituant les grandes parties de ce travail. Nous nous attacherons dans un premier temps à définir l'édition en tant que concept et la chaîne éditoriale. Nous présenterons également notre corpus de travail et son positionnement d'un point de vue diplomatique. Dans un second temps, nous nous pencherons sur les étapes et la réflexion nécessaires à la création d'un modèle d'encodage. Enfin, nous nous intéresserons à la publication des éditions numériques, et plus particulièrement à l'outil TEI Publisher.

---

11. Projets de l'équipe ALMAAnaCH : <https://almanach.inria.fr/projects-fr.html>.



# Première partie

Édition scientifique et numérique



# Chapitre 1

## Le processus d'édition

### 1.1 Édition et édition scientifique

#### 1.1.1 Définition de la notion d'édition

Le substantif « édition » est polysémique, mais tous ses sens se rattachent à la notion de diffusion. Le Trésor de la Langue Française informatisé (TLFi) distingue trois grandes significations : la « reproduction, publication et diffusion » du livre, la « préparation du texte d'une œuvre en vue de sa publication » et, par métonymie des deux premières, le « texte d'une œuvre tel qu'il a été fixé par un éditeur » ou le « texte d'une œuvre tel qu'il a été publié dans une édition <sup>1</sup> ».

L'industrie du livre et de l'édition peut être divisée en deux grands ensembles : l'édition commerciale et l'édition scientifique. Leur distinction se situe dans le public visé. L'édition commerciale s'adresse, de manière générale, au grand public, tandis que l'édition scientifique s'adresse à un public plus ou moins spécialisé dans un domaine en particulier. Nous ne nous intéresserons dans ce mémoire qu'au volet de l'édition scientifique.

#### 1.1.2 Caractéristiques de l'édition scientifique

L'historien allemand Patrick Sahle donne la définition suivante d'une édition scientifique : « *Edition ist die erschließende Wiedergabe historischer Dokumente* <sup>2</sup> ». Pour lui, l'essence de l'édition scientifique se trouve dans le sémantisme du verbe « *erschließen* » (au sens littéral, « ouvrir » ou « développer »), mais dont il n'existe pas d'équivalent satisfaisant en anglais (ni, nous le supposons, en français).

---

1. Entrée complète sur le site du CNRTL : <https://www.cnrtl.fr/definition/edition>.

2. Patrick Sahle, « What is a Scholarly Digital Edition ? », dans *Digital Scholarly Editing : Theories and Practices*, dir. Elena Pierazzo et Matthew James Driscoll, Cambridge, 2016, p. 19-39, DOI : [10.11647/OBP.0095.02](https://doi.org/10.11647/OBP.0095.02), p. 23.

Il propose une traduction, en anglais, de sa définition : « *A scholarly edition is the critical representation of historic documents*<sup>3</sup> ». Il s'attache ensuite à en définir les termes clés. Il revient, d'abord, sur l'utilisation du terme « document » plutôt que « texte », en déclarant qu'il préfère parler de « documents » car tous ne contiennent pas que du texte. Nous le rejoignons sur ce point, en particulier dans la mesure où la notion de « document » nous semble englober celle de « texte ». Parmi les termes clés, il y a la notion de « représentation ». Ainsi, la représentation d'un document peut prendre plusieurs formes, telles que sa numérisation ou la transcription de son contenu ; toutefois, pour parler de « représentation critique », l'édition se doit d'apporter une information supplémentaire à la simple représentation. L'identification de la structure du document et des entités nommées apporte donc une dimension critique à la représentation d'un document.

En outre, une édition scientifique a un objectif particulier, celui de « fournir à la recherche des sources fiables, lisibles, explicitées par une mise en contexte et l'annotation appropriée » et de « rassembler des corpus éclatés<sup>4</sup> ». C'est précisément cet objectif qui caractérise les éditions en ligne de l'EHRI.

## 1.2 Édition et édition numérique

### 1.2.1 Caractéristiques d'une édition numérique

Nous l'avons vu, la distinction entre une édition commerciale et une édition scientifique tient en partie au public auquel elle s'adresse. L'édition numérique, elle, peut être commerciale ou scientifique. Elle s'oppose à l'édition imprimée, notamment au niveau de sa portée.

Il existe plusieurs façons de publier une édition numérique : on peut, par exemple, choisir de publier en ligne un ouvrage imprimé sous la forme d'un livre numérique (ou *ebook*), ou bien essayer une forme de publication différente avec un site web. Il a été montré qu'une édition numérique disponible sur Internet sous la forme d'un site web permet une « démultiplication du public<sup>5</sup> ». En effet, la fréquentation d'un site web dépasse de loin celle d'une salle de lecture, et une édition numérique publiée en ligne permet de toucher un public international, qui ne pourrait autrement peut-être pas accéder à la ressource qu'il recherche.

---

3. *Ibid.*

4. Christine Nougaret et Élisabeth Parinet, *L'édition critique des textes contemporains, XIXe-XXIe siècle*, dir. Christine Nougaret et Élisabeth Parinet, Paris, 2015, p. 19.

5. Jérôme Blachon et Lydiane Gueit-Montchal, « Valorisation des archives », dans *Abrégé d'archivistique*, dir. Lydiane Gueit-Montchal et Dominique Naud, 4<sup>e</sup> éd., Paris, 2020, p. 295-325, p. 314.

L'édition numérique est particulièrement adaptée pour l'édition scientifique puisqu'elle facilite une lecture non linéaire<sup>6</sup>. Les nombreux renvois au sein d'une édition, notamment grâce à des liens hypertextes, facilitent le travail du·de la chercheur·euse en lui donnant accès d'un simple « clic » à une référence ou une définition, par exemple.

Enfin, lorsque l'édition numérique est présentée sous la forme d'un site web, plusieurs « vues » du document peuvent être proposées. Il est ainsi possible, et grandement pratique, d'afficher à la fois la numérisation du document et sa transcription. Il est également possible de choisir de n'afficher que l'une ou l'autre, ou encore d'opter pour la transcription du texte uniquement ou sa traduction. Les possibilités d'enrichissement pour une édition numérique en ligne sous la forme d'un site web sont nombreuses.

### 1.2.2 Limites de l'édition imprimée

Chaque mode de publication présente des avantages et des inconvénients. Il faut toutefois reconnaître que certaines caractéristiques de l'édition numérique découlent des limites de l'édition imprimée. Nous nous concentrerons sur deux points qui nous semblent importants : l'évolution du contenu et le prix.

La mise en vente d'un livre imprimé marque souvent la fin du travail de l'éditeur·ice, à moins de préparer une nouvelle édition. Une édition imprimée représente donc « une expression figée du travail de l'éditeur<sup>7</sup> ». La limite de l'édition imprimée se trouve dans la définition du projet éditorial. Une édition d'un texte en particulier bénéficierait d'une édition imprimée, de même que d'une édition numérique sous la forme d'un *ebook*. Au contraire, une édition qui chercherait à mettre à disposition des chercheur·euse·s de nombreux documents d'archives bénéficierait grandement, voire même uniquement, d'une édition numérique sous la forme d'un site web. En effet, au fur et à mesure des recherches et des découvertes, une édition en ligne peut être augmentée et complétée :

« Un site web suffisamment robuste, s'il a été rendu capable d'accueillir à la fois texte et images des documents retenus, pourra en intégrer un nombre très important, ce qui permet notamment d'étendre le périmètre du projet à l'ensemble des reproductions numériques des textes, et finalement à un plus grand nombre de textes qu'une publication imprimée ne pourra en présenter.<sup>8</sup> »

Cela est particulièrement vrai pour un projet comme les éditions en ligne de l'EHRI, qui sont thématiques ; des archives appartenant à des pays qui rejoindraient l'EHRI dans quelques années pourront intégrer sans problème les éditions déjà publiées.

---

6. Florence Clavaud, « L'édition numérique », dans *L'édition critique des textes contemporains, XIXe-XXIe siècle*, dir. Christine Nougaret et Élisabeth Parinet, Paris, 2015, p. 175-200, p. 179.

7. *Ibid.*, p. 177.

8. *Ibid.*, p. 176-177.

Le second aspect qu'il nous paraît important d'évoquer est le prix d'une édition imprimée, notamment pour le·la lecteur·ice. Se procurer une édition scientifique imprimée peut avoir un coût très élevé<sup>9</sup>, et parfois relever de l'impossible en fonction du nombre de tirages et/ou de l'exhaustivité du catalogue de sa bibliothèque. En revanche, une édition numérique peut s'inscrire dans une démarche de science ouverte<sup>10</sup>, définie ainsi :

« La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche<sup>11</sup>. »

D'un point de vue scientifique, nous estimons qu'il est primordial, surtout lorsque l'on édite des documents d'archives qui serviront de « sources primaires » aux historien·ne·s, que le résultat de cette édition soit accessible à tou·te·s.

## 1.3 La chaîne éditoriale

### 1.3.1 Définition de la chaîne éditoriale

Pour fonctionner, l'industrie du livre et de l'édition a besoin d'une « chaîne éditoriale ». Qu'il s'agisse d'une édition commerciale ou scientifique, imprimée ou numérique, la chaîne éditoriale est centrale.

« Une chaîne éditoriale ou chaîne d'édition est la suite d'opérations (ou procédé industriel) par lequel un document rédigé par un auteur est transformé en document publiable et publié. Elle consiste à formater le document écrit, à élaborer des modèles de documents, et à effectuer les conversions de fichiers nécessaires. Elle s'occupe également du stockage et de la diffusion des documents<sup>12</sup>. »

La chaîne éditoriale est donc le procédé par lequel est produite une version publiable, imprimée ou en ligne, d'un ou de plusieurs documents.

---

9. E. Pierazzo, « What Future for Digital Scholarly Editions?, From Haute Couture to Prêt-à-Porter », *International Journal of Digital Humanities*–1 (2019), p. 209-220, DOI : [10.1007/s42803-019-00019-3](https://doi.org/10.1007/s42803-019-00019-3), p. 210.

10. On pourra consulter à ce sujet : Laurent Romary, « Éléments d'une communication scientifique et ouverte », dans *Publier, éditer, éditorialiser, Nouveaux enjeux de la production numérique*, dir. Lisette Calderan, Pascale Laurent, Hélène Lowinger et Jacques Millet, 1<sup>re</sup> éd., Louvain-la-Neuve, 2016, p. 37-69, DOI : [10.3917/dbu.calde.2016.01.0037](https://doi.org/10.3917/dbu.calde.2016.01.0037).

11. Voir le deuxième plan national pour la science ouverte : <https://www.ouvrirlascience.fr/>.

12. Wikipédia, « chaîne éditoriale » : [https://fr.wikipedia.org/wiki/chaine\\_editoriale](https://fr.wikipedia.org/wiki/chaine_editoriale) (visité le 25/08/2023).

### 1.3.2 Un exemple : la *pipeline* DiScholEd

La publication d'une édition en ligne pouvant être significativement différente de la publication d'une édition imprimée, la chaîne éditoriale de cette dernière nécessite quelques ajustements. Nous prendrons l'exemple de la chaîne d'édition (également appelée *pipeline*) de l'application pour la publication en ligne d'egodocuments, DiScholEd<sup>13</sup> (pour « *Digital Scholarly Editions* »).

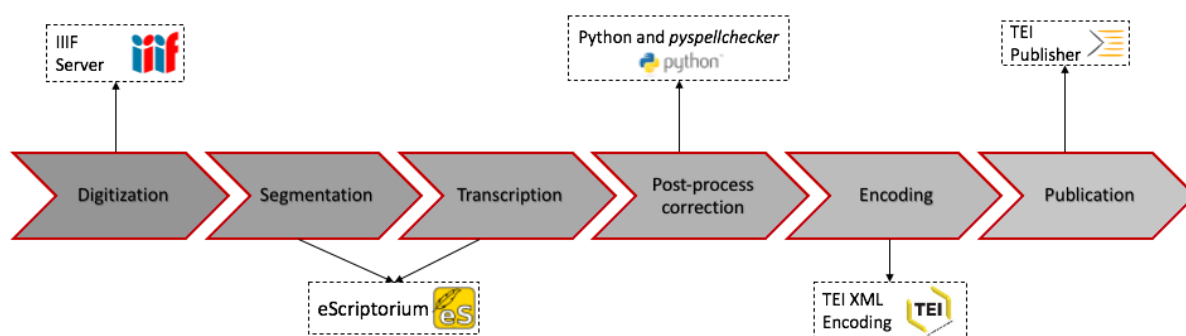


FIGURE 1.1 – Chaîne éditoriale de l'application DiScholEd (source : F. Chiffolleau)

La chaîne éditoriale de DiScholEd s'articule actuellement en six étapes. La première étape essentielle est la numérisation. DiScholEd recommande d'héberger les numérisations sur un serveur IIIF<sup>14</sup>. Le format IIIF est notamment apprécié en raison de son interopérabilité, et du fait qu'il permet la diffusion des images en haute résolution. La transcription d'un document numérisé peut être considérablement accélérée grâce à l'utilisation d'outils de segmentation et de transcription<sup>15</sup>, basés sur la reconnaissance optique de caractères (OCR) et la reconnaissance d'écritures manuscrites (HTR).

Une fois la transcription vérifiée, il est possible de télécharger un fichier XML de la transcription, prêt à être encodé. Pour cela, il est préférable de suivre les recommandations du standard TEI<sup>16</sup>, interopérable et ouvert. L'encodage terminé, le document peut être publié en ligne, sur une plateforme permettant l'affichage de l'encodage sémantique TEI. Notre travail sur les éditions de l'EHRI s'est articulé autour de ces deux dernières étapes de la chaîne éditoriale.

13. La chaîne d'édition de DiScholEd est accessible librement sur GitHub : <https://github.com/DiScholEd/pipeline-digital-scholarly-editions>.

14. Pour comprendre le fonctionnement de IIIF : <https://iiif.io/get-started/how-iiif-works/> (visité le 27/08/2023).

15. DiScholEd recommande l'utilisation d'*eScriptorium* : <https://escriptorium.paris.inria.fr/> (visité le 27/08/2023).

16. *P5 : Guidelines for Electronic Text Encoding and Interchange*, version 4.6.0, TEI Consortium, 4 avr. 2023, URL : <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.





# Chapitre 2

## Les éditions du projet EHRI

### 2.1 Une nouvelle approche des archives numériques

#### 2.1.1 Une initiative européenne

L'objectif principal de l'EHRI en tant qu'infrastructure est de promouvoir la recherche sur la Shoah en facilitant l'accès aux documents d'archives, dispersés à travers le monde. Elle doit établir un plan de travail et produire régulièrement des livrables<sup>1</sup>.

Lors de la deuxième phase du projet, nommée EHRI-2 et financée par le programme Horizon 2020, l'EHRI a dédié une partie de son travail au développement d'outils pour l'archivage numérique de documents concernant la Shoah. Cela a pris la forme d'un *Work Package* (WP12) intitulé « *New Views on Digital Archives* » (« Nouvelles perspectives sur les archives numériques »).

#### 2.1.2 Développer les éditions numériques sur la Shoah

L'équipe éditoriale du WP12 a défini les critères qu'elle estime essentiels à une édition numérique sur la Shoah. Parmi ces critères, nous retrouvons une interface claire, dotée d'une bonne fonctionnalité de recherche, l'intégration de leur vocabulaire contrôlé, ainsi que l'affichage des données géographiques sur une carte<sup>2</sup>.

---

1. Plan de travail et livrables pour EHRI-3 : <https://www.ehri-project.eu/division-work>.

2. Michal Frankl, Michael Bryant, Jessica Green, Wolfgang Schellenbacher et Magdalena Sedlická, *Edition of documents*, EHRI GA no. 654164 D.12.2, European Holocaust Research Infrastructure, 2018, p. 1-21, URL : <https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D12%202%20Thematic%20approach%201%20Edition%20of%20documents.pdf> (visité le 14/06/2023), p. 6.

## 2.2 Les éditions en ligne EHRI

Chaque édition est une collection<sup>3</sup> de documents conservés dans les institutions partenaires de l’EHRI. Ce sont des éditions thématiques<sup>4</sup> reflétant une période ou illustrant un phénomène.

### 2.2.1 BeGrenzte Flucht (2018)

La première édition publiée par l’EHRI, « *BeGrenzte Flucht : Die österreichischen Flüchtlinge an der Grenze zur Tschechoslowakei im Krisenjahr 1938* » (« Fuite entravée : les réfugié·e·s autrichien·ne·s à la frontière tchécoslovaque en 1938<sup>5</sup> »), a été publiée en 2018. Elle contient cent quatre documents.

Le nazisme est une idéologie basée sur l’antisémitisme : la population juive est considérée comme nuisible et sert de bouc-émissaire à la situation de l’Allemagne au sortir de la Première Guerre mondiale. Toutefois, le régime nazi ne semble pas souhaiter l’élimination massive et immédiate des Juifs allemands avant la guerre<sup>6</sup>. Sa politique cherche avant tout à exclure et humilier les Juifs, pour les forcer à quitter le territoire<sup>7</sup>. Si une grande partie de la population juive quitte l’Allemagne, le « problème » posé par la « question juive » a seulement été déplacé et se représente au fur et à mesure que le Reich étend son territoire.

Le 12 mars 1938, l’Autriche est annexée par l’Allemagne nazie, c’est l’« *Anschluss* », et la population juive résidant alors en Autriche est contrainte d’émigrer. La Tchécoslovaquie devient l’un des principaux refuge pour les Juif·ve·s autrichien·ne·s, mais la politique d’accueil de plus en plus restrictive de la Tchécoslovaquie conduit au blocage complet de la frontière<sup>8</sup>. L’édition « *BeGrenzte Flucht* » réunit des documents d’archives principalement tchèques et autrichiens, sous la forme de rapports gouvernementaux, articles de presse et de témoignages direct de victimes et de témoins.

---

3. Nous emploierons les termes « édition » et « collection » indistinctement pour faire référence aux éditions en ligne EHRI.

4. *EHRI Online Editions*, European Holocaust Research Infrastructure, URL : <https://www.ehri-project.eu/ehri-online-editions>.

5. Les traductions proposées pour les titres des éditions sont les nôtres.

6. Frédéric Sallée, *Anatomie du nazisme*, 2018 (Idées reçues), DOI : [10.3917/lcb.salle.2018.01](https://doi.org/10.3917/lcb.salle.2018.01), p. 123.

7. G. Bensoussan, *Histoire de la Shoah...*, p. 22.

8. À propos de l’édition « *BeGrenzte Flucht* » (en allemand) : <https://begrenzte-flucht.ehri-project.eu/exhibits/show/einleitung>.



FIGURE 2.1 – Frontières du Reich en mars 1938 (Source : Wikipédia)

### 2.2.2 Early Holocaust Testimony (2020)

La deuxième édition publiée par l’EHRI, « *Early Holocaust Testimony* » (« Premiers témoignages de la Shoah »), a été publiée en 2020. Elle contient cent dix-neuf documents.

Il convient tout d’abord de préciser le sens de « témoignage ». Le TLFi définit le témoignage comme une « déclaration qui confirme la véracité de ce que l’on a vu, entendu, perçu, vécu<sup>9</sup> ». La collection « *Early Holocaust Testimony* » se concentre sur le témoignage écrit, les témoignages oraux ayant été retranscrits. Nous reviendrons ensuite sur le terme « *early* », à propos duquel l’équipe éditoriale précise qu’il correspond à tout témoignage de la persécution de la communauté juive depuis le début du régime nazi en 1933 au procès d’Adolf Eichmann à Jérusalem en 1961<sup>10</sup>.

9. Entrée complète sur le site du CNRTL : <https://www.cnrtl.fr/definition/temoignage>

10. Notre propre traduction, à partir de l’introduction à l’édition (« *What is an early Holocaust testimony ?* ») : <https://early-testimony.ehri-project.eu/exhibits/show/about/what-is-testimony>.

### 2.2.3 Diplomatic Reports (2021)

La troisième édition publiée par l’EHRI, « *Diplomatic Reports* » (« Rapports diplomatiques »), a été publiée en 2021. Elle contient soixante-douze documents venant du Danemark, de l’Italie, du Japon et des États-Unis. De nouveaux rapports provenant de la Hongrie, de la Slovaquie et de la Suède devraient prochainement intégrer la collection.

La position du-de la diplomate est particulière : il-elle est amené-e à côtoyer les politicien-ne-s influent-e-s du pays où il-elle est stationné-e, mais doit garder une certaine distance en tant qu’observateur-ice. Les informations récoltées doivent être régulièrement communiquées au ministère des Affaires étrangères de leur état d’origine par le biais de rapports.

De nombreux-ses diplomates stationné-e-s en Allemagne nazie ont rapidement compris que les dirigeants nazis ne se contenteraient pas des lois discriminatoires à l’encontre de la population juive. L’étude de ces rapports diplomatiques permet d’aborder la Shoah comme un événement européen auquel de nombreux pays ont activement participé, notamment par la fermeture de leurs frontières<sup>11</sup>.

### 2.2.4 Nisko (2023)

La quatrième édition publiée par l’EHRI, « *Von Wien ins Nirgendwo : Die Nisko Deportationen 1939* » (« De Vienne à Nulle Part : les déportations vers Nisko de 1939 »), a été publiée en 2023. Elle contient quarante documents.

En avril 1938, le régime nazi ouvre son agence de l’Office central pour l’émigration juive à Vienne en Autriche<sup>12</sup>. Cette agence a pour but d’accélérer l’émigration forcée des Juif-ve-s autrichien-ne-s. Au moment de l’annexion de l’Autriche en mars 1938, « plus de 200 000 personnes étaient considérées comme juives selon les critères des lois raciales de Nuremberg. Les Juifs autrichiens représentaient 40 % des Juifs vivant sur tout le territoire du Grand Reich<sup>13</sup> ». L’agence de Vienne a été libre de « tester un certain nombre de mesures extrémistes contre les Juifs<sup>14</sup> », autonome de l’agence principale située à Berlin.

Le « Plan Nisko » débute en octobre 1939, un mois seulement après l’invasion de la Pologne. L’idée des dirigeants nazis est de relocaliser la population juive à Lublin et Nisko en Pologne, dans ce qu’ils appellent une « réserve juive<sup>15</sup> ».

---

11. Voir l’introduction à l’édition, rédigée par Frank Bajohr : <https://diplomatic-reports.ehri-project.eu/exhibits/show/about/introduction>.

12. *Ibid.*, p. 30.

13. Hans Safrian, « L’accélération de la spoliation et de l’émigration forcée, Le “modèle viennois” et son influence sur la politique antijuive du Troisième Reich en 1938 », trad. par Valérie Nguyen, *Revue d’Histoire de la Shoah*, 1–186 (2007), p. 131-163, DOI : [10.3917/rhsho.186.0131](https://doi.org/10.3917/rhsho.186.0131), p. 137.

14. *Ibid.*, p. 132.

15. G. Bensoussan, *Histoire de la Shoah...*, p. 34.

Jusqu'à présent, la recherche sur le « Plan Nisko » s'est concentrée sur le transport et l'instrumentalisation du train comme premier projet expérimental des futures déportations de masse. La collection « *Nisko* » se concentre sur la déportation d'environ 1 600 hommes juifs de Vienne les 20 et 26 octobre 1939, en seulement deux trains. Le « Plan Nisko » est abordé sous l'angle de la vie quotidienne des déportés à travers leur correspondance. Une partie de ces déportés a assuré la construction de la « réserve juive », certains ont été capturés et envoyés dans des camps de travail russes, tandis que d'autres encore ont été « chassés vers l'Est ». Les Nazis appelaient ce phénomène la « dispersion ».

Le « Plan Nisko » se termine en avril 1940 et des rapatriements vers Vienne sont organisés. Toutefois, le sort de ceux qui n'ont pas été rapatriés n'a pu être que partiellement reconstitué<sup>16</sup>.

---

16. Voir l'introduction à l'édition « *Nisko* », rédigée par Winfried Garscha : <https://nisko-transports.ehri-project.eu/exhibits/show/einleitung/nisko-transporte-wien>.



# Chapitre 3

## La diplomatie à l'épreuve du numérique

### 3.1 La diplomatie

#### 3.1.1 En tant que science

La diplomatie est « la science qui étudie la tradition, la forme et l'élaboration des actes écrits. Son objet est d'en faire la critique, de juger de leur sincérité, d'apprécier la qualité de leur texte, de dégager des formules tous les éléments du contenu susceptibles d'être utilisés par l'historien, de les dater, enfin de les éditer <sup>1</sup> ». À l'origine, la diplomatie ne servait qu'à analyser les actes médiévaux dans le cadre de contestations judiciaires <sup>2</sup>. Ceci explique pourquoi une partie du vocabulaire utilisé pour décrire les documents n'est pas adaptée à l'analyse de documents contemporains.

#### 3.1.2 L'analyse diplomatique

L'analyse diplomatique d'un acte est traditionnellement divisée en plusieurs catégories. Les caractéristiques externes <sup>3</sup> d'un document concernent sa matérialité, notamment le support utilisé (papyrus, parchemin, papier, etc.). Les caractéristiques internes <sup>4</sup> concernent l'écrit lui-même et analyse son contenu.

---

1. Maria Milagros Cárcel Ortí (éd.), *Vocabulaire international de la diplomatie*, Commission internationale de diplomatie, 1997, URL : <http://www.cei.lmu.de/VID/> (visité le 19/08/2023), p. 21.

2. Luciana Duranti, « Pour une diplomatie des documents électroniques », *Bibliothèque de l'École des chartes*, 161-2 (2003), p. 603-623, DOI : [10.3406/bec.2003.463631](https://doi.org/10.3406/bec.2003.463631), p. 604.

3. *Vocabulaire international de la diplomatie...*, p. 45-50.

4. *Ibid.*, p. 51-68.

### 3.1.3 La diplomatie contemporaine

Durant la seconde moitié du XX<sup>e</sup> siècle, l'analyse diplomatique a été étendue aux documents administratifs<sup>5</sup> et les catégories de documents écrits ont été élargies. On parlera ainsi de « document dispositif » lorsque l'action juridique dépend de sa mise à l'écrit, de « document probatoire » lorsque sa mise à l'écrit apporte une preuve à l'action juridique<sup>6</sup>, de « document à l'appui » quand celui-ci contribue à l'action sans la créer ni la prouver, et enfin de « document narratif » pour tout document qui ne participe pas à l'action juridique<sup>7</sup>.

## 3.2 Typologie documentaire des éditions EHRI

Si l'on se rapporte à la typologie documentaire établie par Luciana Duranti, les éditions EHRI sont composées de documents probatoires, à l'appui et narratifs. Nous distinguons trois types de documents principaux : les témoignages, les rapports diplomatiques et les articles de presse.

Chacune des éditions de l'EHRI se concentre sur un aspect de l'histoire de la Shoah. De nombreux documents prennent la forme de témoignages. Au sens de la diplomatie contemporaine, ce sont des documents narratifs décrivant la persécution de la population juive en Europe à partir de 1933. Les témoignages sont de trois natures principales : les extraits de journaux intimes, la correspondance, et les témoignages oraux. Ces derniers ont été recueillis *a posteriori* et transcrits, notamment les témoignages judiciaires entendus lors des procès des dirigeants nazis après la guerre. Les témoignages ont été rédigés par des Juif·ve·s eux·elles-mêmes ou par des témoins de leur persécution.

Les rapports diplomatiques diffèrent des témoignages dans la mesure où ce sont des documents officiels. Ils prennent parfois la forme d'un document attaché à une lettre, parfois le contenu du rapport se trouve dans un télégramme. Ils témoignent du sort de la population juive, notamment au moment des déportations. Certains rapports sont rédigés sur du papier à lettre officiel, dont il convient d'encoder l'en-tête.

En plus des témoignages directs, les collections contiennent également de nombreux articles parus dans la presse. Seul le texte de l'article est encodé, même lorsque toute une page du journal est numérisée.

---

5. L. Duranti, « Pour une diplomatie des documents électroniques »..., p. 604-605.

6. *Vocabulaire international de la diplomatie*..., p. 22.

7. L. Duranti, « Pour une diplomatie des documents électroniques »..., p. 605.



## 3.3 Le standard TEI pour l'encodage des documents

### 3.3.1 Un standard ouvert et interopérable

La TEI est un standard qui a été développé « par la communauté scientifique, pour son propre usage<sup>8</sup> ». Elle s'appuie sur le métalangage XML et permet un encodage sémantique du texte. Contrairement au traitement de texte, qui s'attache plutôt à l'apparence du texte, la TEI permet par exemple de différencier sémantiquement des homonymes, de sorte que lors de la construction d'un index les occurrences ne seront pas confondues.

En outre, la TEI est un standard interopérable, et un texte encodé suivant ses recommandations sera toujours lisible. Elle est détachée de tout logiciel et des mises à jours dont celui-ci aurait besoin, ou de son obsolescence. L'encodage TEI est lisible par l'humain.

### 3.3.2 Diplomatique et TEI

Les documents encodés en TEI ont une structure initiale identique : le `<teiHeader>` et le `<text>`. Le `<teiHeader>` regroupe les métadonnées concernant le document. Cette section a une structure minimale obligatoire mais peut être étendue au besoin. Dans notre processus de création du nouveau modèle d'encodage pour les éditions en ligne de l'EHRI, nous avons considérablement précisé les champs contenus dans le `<teiHeader>` (Annexe D). Le `<text>` contient l'encodage de la partie textuelle du document. Il contient un élément `<body>` qui peut être divisé en sections avec des éléments `<div>` (« *ext division* ») selon la structure du texte.

La diplomatique, comme toutes les sciences, dispose d'un vocabulaire spécifique. La notion de « protocole », par exemple, n'est pas représentée en tant que telle au sein de la TEI<sup>9</sup>. Toutefois, un encodage rigoureux du texte et des métadonnées permet d'identifier toutes les informations qui seraient regroupées sous l'appellation « protocole » (initial ou final) dans une analyse diplomatique.

---

8. Lou Burnard, *Qu'est-ce que la Text Encoding Initiative ?*, trad. par Marjorie Burghart, 2015 (Encyclopédie numérique), DOI : [10.4000/books.oep.1237](https://doi.org/10.4000/books.oep.1237).

9. F. Clavaud, « L'édition numérique »..., p. 185.



Deuxième partie

Modèle d'encodage



# Chapitre 4

## Encodage des éditions EHRI

### 4.1 Analyse des pratiques d'encodage

Avec la publication de plusieurs éditions thématiques, l'équipe d'éditoriale de l'EHRI a pu expérimenter l'utilisation de la TEI et adapter sa pratique. Notre travail a donc débuté par l'observation des pratiques d'encodage des précédentes éditions.

#### 4.1.1 Extraction de l'information

Le corpus à analyser était composé des quatre éditions, soit un total de trois cent trente-cinq documents. À l'aide d'un *script* Python, nous avons pu extraire toutes les occurrences des balises et attributs utilisés par collection. Nous avons ensuite arrangé les résultats dans un tableau récapitulatif (Annexe C).

## Total number of documents in the collection

## Comments on the use of the tag and its attributes

Tag Name	header	body	Number of Occurrences				List of @Attributes used in EHRI editions	ODD EHRI	Comment(s)	Files Involved
			toOmnisight (104)	Early Treaties (119)	Diplomatic Reports (72)	Naka (40)				
ab	<input type="checkbox"/>	<input checked="" type="checkbox"/>				6	@type="diploma"	<input checked="" type="checkbox"/>	Use «metamark» instead (see Encoding Guidelines).	EHRI NISKO 19400102_DE.xml
abstract	<input checked="" type="checkbox"/>	<input type="checkbox"/>	97	113		3		<input checked="" type="checkbox"/>	If the title of the document is not explicit enough, a good practice is to add an «abstract» tag and a summary of the document's content.	
author	<input checked="" type="checkbox"/>	<input type="checkbox"/>	208	293		120		<input checked="" type="checkbox"/>		
availability	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		40		<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection.	
bibl	<input checked="" type="checkbox"/>	<input type="checkbox"/>	135	119	72	36	@type	<input checked="" type="checkbox"/>		
biblScope	<input checked="" type="checkbox"/>	<input type="checkbox"/>	10				@unit	<input type="checkbox"/>	Use «bibl» tag instead.	EHRI BF 19390202_DE.xml EHRI BF 19390316_DE.xml EHRI BF 19390409_DE.xml EHRI BF 19390418_DE.xml EHRI BF 19390421c_DE.xml EHRI BF 19390723_DE.xml EHRI BF 19390803_DE.xml EHRI BF 19390911_DE.xml
birth	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1				@when	<input type="checkbox"/>	Including the «birth» tag in the ODD is not relevant.	
body	<input type="checkbox"/>	<input checked="" type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
catRef	<input checked="" type="checkbox"/>	<input type="checkbox"/>	117			37	@target	<input checked="" type="checkbox"/>	Guidelines: «catRef» (category reference) specifies one or more defined categories within some taxonomy or text typology. Values for the @target attribute include: "Schmugge" (smuggling), "Schicksale" (fate/destiny), "Flüchtlingspolitik" (Refugee Policy), "WGO Wien" (Israelische Kultusgemeinde Wien), "Berichte" (reports), etc. Values must be normalized (capital letter at the beginning of German nouns).	

Tag's place of occurrence in the XML tree

List of the files containing (or missing) the tag

FIGURE 4.1 – Fonctionnement du tableau récapitulatif de l'encodage des éditions EHRI

Les balises sont classées par ordre alphabétique, et une case à cocher permet de voir si la balise se trouve dans le `<teiHeader>` et/ou dans le `<body>`. Chaque collection dispose ensuite de sa propre colonne, indiquant pour chaque balise son nombre d’occurrences. Nous avons ensuite établi, pour chaque balise, une liste de tous ses attributs utilisés dans l’encodage EHRI.

### 4.1.2 Évaluation du choix des balises et attributs

Une fois le tableau récapitulatif en partie rempli, nous avons pu procéder à une analyse plus approfondie de l’encodage. En fonction du contexte, nous avons ajouté des commentaires sur l’utilisation de certaines balises ; une précision sur l’usage recommandé par les *TEI Guidelines* ou une suggestion d’amélioration, par exemple.

Le tableau a été rendu le plus complet possible grâce à la rédaction d’un *script*, en Python, permettant de rechercher des informations précises dans le corpus (extrait du code en Annexe E). Le nombre d’occurrences de certaines balises a révélé des doublons, à l’instar de la balise `<creation>`, apparaissant parfois deux fois au sein d’un même fichier. Or, la création d’un document ne peut avoir qu’une seule occurrence. Ce *script* nous a permis de remplir la dernière colonne du tableau, « *Files Involved* », pour que les éditeur-ice-s puissent corriger ce type d’erreurs.

## 4.2 Identification des améliorations possibles

Le tableau complet donne une bonne vue d’ensemble de l’encodage des éditions EHRI, et notamment de leur évolution. Nous avons identifié deux pistes d’améliorations significatives : une meilleure conformité aux recommandations de la TEI et l’uniformisation des éditions déjà publiées.

### 4.2.1 Amélioration de la conformité

Pour être conforme au standard TEI, un fichier XML doit être bien formé, c’est-à-dire rédigé avec une syntaxe correcte. Le fichier XML doit également être valide. Pour cela, il faut que l’encodage respecte les règles du schéma qui lui a été associé. Enfin, le projet d’encodage doit être renseigné de façon exhaustive dans une ODD<sup>1</sup> (« *One Document Does-it all* »), elle aussi conforme.

Parmi les bonnes pratiques que nous recommandons dans le traitement d’un corpus comme celui des éditions EHRI, il y a tout d’abord la question de l’identifiant du fichier. Les fichiers EHRI portent tous un identifiant unique basé sur une même structure : « EHRI-`<collection>`-YYYYMMDD\_`<langue>` » (par exemple, « EHRI-BF-19380315\_DE »).

---

1. Nous aborderons la notion d’ODD dans le chapitre suivant.

Cet identifiant doit être la valeur de l'attribut `@xml:id` qui se trouve dans la balise ouvrante de l'élément racine `<TEI>`.

Le seconde recommandation qui nous paraît particulièrement importante est la structuration du contenu à l'intérieur de la balise `<body>`. Pour assurer un bon affichage du contenu et éviter la multiplication inutile de fichiers, nous proposons d'encoder la transcription du fac-similé et ses traductions au sein d'un même fichier. Il y aurait donc au moins une division `<div>` de premier niveau, complétée par un attribut `@type` dont la valeur serait "transcription" ou "translation". Dans le cas des traductions, l'attribut `@xml:lang` apporterait une précision sur la langue de la traduction contenue dans la division.

## 4.2.2 Uniformisation des édition

### 4.2.2.1 Utilisation de l'anglais

La nature européenne du projet EHRI implique que les différents acteurs ne parlent pas tous la même langue. La langue la plus communément parlée parmi les éditeur-ice-s et le public des éditions est l'anglais. Il nous paraît donc logique de normaliser l'utilisation de l'anglais comme langue d'encodage des métadonnées et de la valeur des attributs. Cela permet à quiconque accédant aux fichiers XML d'en comprendre l'encodage. Nous avons ainsi parfois été confronté-e à des valeurs d'attributs en allemand. Le problème posé par l'utilisation de l'allemand pour préciser la valeur des attributs est le risque d'une perte sémantique lors de sa traduction dans une autre langue. Or, les valeurs d'attributs proposées par les *TEI Guidelines* dans leur documentation d'éléments sont toujours en anglais, ce qui facilite le travail d'encodage et la réutilisation du schéma produit pour l'édition.

#### 4.2.2.2 Standardisation de la valeur de certains attributs

Il est essentiel que les valeurs des attributs soient uniformes. Pour certains attributs, il existe des normes que l'on peut choisir d'appliquer. Par exemple, il existe une norme d'encodage concernant le format des dates : l'utilisation de l'attribut `@when-iso` sera privilégié si la date est au format « YYYY-MM-DD » (norme ISO 8601 <sup>2</sup>), sinon on lui préférera l'attribut `@when`. Cela est d'autant plus important qu'une date incomplète avec un attribut `@when-iso` causera un bug d'affichage, et qu'une date complète avec un attribut `@when` ne s'affichera pas correctement. Il n'existe, en revanche, pas de normes concernant l'attribut `@type`, ses valeurs possibles étant aussi variées que la diversité des éléments auquel il peut être associé.

---

2. Récapitulatif de la norme ISO 8601 sur Wikipédia : [https://fr.wikipedia.org/wiki/ISO\\_8601](https://fr.wikipedia.org/wiki/ISO_8601) (visité le 03/09/2023).

L'importance de la valeur de l'attribut `@xml:lang` peut paraître abstraite. En général, et même s'il est possible d'écrire le nom de la langue en toutes lettres, nous utilisons un code pour donner la valeur de l'attribut de langue. Or, ce code doit s'inscrire dans un système établi : c'est le rôle des normes. Prenons l'exemple de la langue allemande : « Allemand » en français, « *German* » en anglais, et « *Deutsch* » en allemand. Sans système pré-établi, l'éditeur·ice sera naturellement tenté·e d'utiliser un code parlant pour sa propre langue. Nous obtiendrions donc les possibilités suivantes : « "all" », « "ger" » et « "deu" » respectivement. Le problème devient évident. Le choix ne peut pas revenir à l'éditeur·ice de déterminer lui·elle-même des codes à utiliser. Il existe en outre des préférences, qui ne regardent que les traducteur·ice·s, concernant le choix de traduire certains termes ou de les laisser apparaître dans leur langue originale. Ce problème s'applique également ici.

Plusieurs systèmes existent. La norme ISO 639<sup>3</sup> définit des codes pour les langues à deux ou trois caractères (ISO 639-1 ou ISO 639-2). Le problème posé par cette norme est qu'elle ne tranche pas entre l'utilisation de deux ou trois caractères, ce qui peut conduire à une certaine confusion. Nous avons donc opté pour le registre de codes établi par l'Iana<sup>4</sup>. Celui-ci a l'avantage de ne proposer qu'une norme, précise et très régulièrement mise à jour.

---

3. Récapitulatif de la norme ISO 639 sur Wikipédia : [https://fr.wikipedia.org/wiki/ISO\\_639](https://fr.wikipedia.org/wiki/ISO_639) (visité le 03/09/2023).

4. « *Iana Language Subtag Registry* » : <https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>.



# Chapitre 5

## Définition du nouveau modèle d'encodage

### 5.1 *One Document Does-it-all*

Une ODD (« *One Document Does-it-all* ») est un fichier XML encodé en TEI dont l'objectif est double. Elle permet de documenter le projet éditorial, en prose, mais également de créer des règles de validation. Ces dernières seront exportées dans un schéma RelaxNG, qui sera appliqué aux fichiers pour guider et valider leur encodage.

Il existe trois principales façons de générer une ODD : l'utilisation de l'interface ROMA<sup>1</sup> créée par le TEI Consortium, l'utilisation de la feuille de style XSL « *Oddbyexample* » à partir d'un encodage préexistant, ou la rédaction manuelle de l'ODD à partir de rien. Nous avons rapidement éliminé les deux dernières options et choisi de générer la base de notre ODD avec ROMA. En effet, la création de l'ODD manuellement aurait été trop chronophage, et la génération à partir de « *Oddbyexample* » aurait nécessité de corriger les fichiers EHRI en amont ; or la répartition des balises est parfois inégale au sein des trois cent trente-cinq fichiers.

La solution la plus optimale dans notre cas a donc été de travailler avec ROMA en partant du schéma le plus large : « *TEI All* ». L'application directe de ce schéma serait tout à fait inutile, puisqu'il contient tous les éléments de la TEI, alors que le projet n'en utilise qu'une partie. Il existe d'autres schémas spécifiques mais pour être pertinent, le schéma doit correspondre le plus possible aux caractéristiques du projet.

La customisation de l'ODD avec ROMA se présente sous la forme d'une liste des éléments de la TEI accompagnés d'une brève présentation et d'une case à cocher. Les éléments peuvent être classés par ordre alphabétique ou par module. Dans leur rapport,

---

1. ROMA : <https://roma.tei-c.org/>.

les éditeur·ice·s de l'EHRI ont souligné l'importance du module **namesdates** dans leur encodage<sup>2</sup>. Pour créer le modèle le plus pertinent possible, nous nous sommes attaché·e à consulter la documentation de tous les éléments qui nous paraissaient pertinents, en particulier concernant l'encodage des métadonnées dans le `<teiHeader>`. Une fois la base de l'ODD générée, il convient de rédiger consciencieusement la documentation et de créer des règles de validation fonctionnelles.

## 5.2 Rédaction de la documentation

Nous avons divisé la rédaction de la documentation de l'ODD en trois sections : les règles fondamentales, l'encodage du `<teiHeader>` et l'encodage du `<body>`. L'idée de la documentation de l'ODD est de s'affranchir de la seule consultation des *TEI Guidelines* et d'illustrer l'usage que nous en faisons dans le cadre du projet, notamment grâce à des exemples tirés de notre propre corpus.

La création du tableau récapitulatif (Annexe C), et les commentaires que nous y avons inclus, nous a permis de dégager certaines pratiques que les éditions en ligne de l'EHRI auraient tout intérêt à adopter pour assurer la conformité et l'uniformité de leur encodage. Ce sont les recommandations d'uniformisation que nous évoquons dans le chapitre précédent.

En analysant l'encodage des collections, nous avons pu adapter la structure du `<teiHeader>` (Annexe D). Cette partie de la documentation est particulièrement longue, mais il nous paraissait essentiel d'insister dessus dans la mesure où les membres de l'équipe éditoriale de l'EHRI sont nombreux et que la composition de l'équipe peut être amenée à changer au fil des années. C'est cette partie qui a nécessité le plus de clarifications, dans la mesure où nombre d'éléments figuraient déjà dans l'encodage des métadonnées, mais leur utilisation a été réadaptée dans notre nouveau modèle.

Enfin, nos recommandations pour l'encodage du `<body>` ont été guidées par la volonté de transcrire la structure du fac-similé le plus fidèlement possible. Nous avons donc insisté sur l'utilisation des balises `<pb/>` (« *page beginning* ») et `<lb/>` (« *line beginning* ») pour que l'affichage lors de la publication soit le plus proche possible du fac-similé. Nous avons également affiné l'encodage en introduisant les éléments `<opener>` et `<closer>`, contenant tous les deux les informations relatives aux en-têtes des documents et aux éléments de salutations.

---

2. M. Frankl, M. Bryant, J. Green, *et al.*, *Edition of documents...*

## 5.3 Création du schéma de validation

Les règles du schéma de validation sont contenues dans un élément `<schemaSpec>` (« *schema specification* »).

### 5.3.1 Déclaration des modules

Les éléments de la TEI sont classés en modules. Quatre d'entre eux sont obligatoires et communs à tous les encodages : `tei`, `header`, `core` et `textstructure`. Les modules sont déclarés à l'aide de l'élément `<moduleRef>` (« *module reference* »). Le module déclaré est précisé par la valeur de l'attribut `@key`.

Il existe ensuite deux façons de déclarer les éléments de chaque module acceptés par le schéma de validation : l'attribut `@include` déclare les éléments autorisés, tandis que l'attribut `@except` déclare les éléments non-autorisés. Dans un souci de légèreté et de lisibilité, il est généralement conseillé de choisir le mode de déclaration en fonction du nombre d'éléments à déclarer. Ainsi, si l'on souhaite utiliser tous les éléments du module `namesdates` sauf les éléments `<age>` et `<birth>`, par exemple, il est conseillé de les exclure avec l'attribut `@except`. Toutefois, nous avons choisi, dans notre modèle, de n'utiliser que l'attribut `@include`, que nous estimons plus parlant pour des personnes qui ne seraient pas familières de la TEI.

### 5.3.2 Règles de validation des éléments

La personnalisation des règles de validation des éléments se fait grâce à l'élément `<elementSpec>` (« *element specification* »), qui contient les informations sur la structure et le contenu possible de l'élément. Il n'est pas nécessaire de déclarer ici les éléments qui restent inchangés. Nous avons établi pour les éditions EHRI des règles de validation mettant en place les recommandations exprimées dans la partie documentation. Il s'agissait le plus souvent de déclarer le caractère recommandé ("`rec`" pour « *recommended* ») ou obligatoire ("`req`" pour « *required* ») d'un attribut, et éventuellement de définir une liste semi-ouverte de valeurs (comme pour l'attribut `@xml:lang`).

### 5.3.3 Déclaration des classes d'attributs

Les attributs sont regroupés en classes. Par exemple, la plupart des attributs précisant l'affichage souhaité de l'encodage appartiennent à la classe `att.global.rendition`. Certains attributs, comme `@type`, appartiennent à plusieurs classes d'attributs, et chaque élément prend en charge certaines classes définies. La déclaration des classes d'attributs permet donc de limiter l'usage des attributs à celui que l'on souhaite en faire dans le cadre du projet.



# Chapitre 6

## Semi-automatisation du modèle

### 6.1 Automatisation et semi-automatisation

La mission de l’Inria sur le projet EHRI est de proposer un soutien technique, que les éditeur·ice·s sont libres d’utiliser ou non. Si certaines tâches peuvent être automatisées, à l’aide de *scripts* Python par exemple, leur utilisation n’est pas forcément à la portée de tout le monde et les éditeur·ice·s pourraient préférer une solution plus accessible.

La créatrice de l’application DiScholEd, Floriane Chiffolleau, a rédigé une série de *scripts* d’automatisation<sup>1</sup> de plusieurs étapes de la chaîne éditoriale, notamment concernant l’encodage des métadonnées et la gestion des entités nommées. L’adaptation de ces *scripts* aux collections de l’EHRI fait partie des prochaines étapes du travail sur le projet.

Il n’est actuellement pas envisageable d’automatiser tout ou partie de la chaîne éditoriale des éditions EHRI. C’est pourquoi nous avons opté pour une semi-automatisation du processus d’encodage en proposant aux éditeur·ice·s des *templates* (« modèles<sup>2</sup> ») pour l’encodage de certaines parties des fichiers.

### 6.2 Création des *templates* d’encodage

Nous avons créé deux sortes de *templates* d’encodage. Le premier, et le plus conséquent, concerne l’encodage des métadonnées (Annexe D) ; le second concerne les entrées des index. Il existe trois index aux éditions EHRI : un index de lieux, un index de personnes et un index des organisations. Nous avons créé un quatrième *template* pour un « index de termes ». Actuellement, des renvois sont faits vers le vocabulaire contrôlé disponible sur le portail de l’EHRI<sup>3</sup>.

---

1. Voir le dépôt GitHub du projet DAHN : <https://github.com/FloChiff/DAHNProject>.

2. Nous utiliserons le terme anglais « *template* » afin d’établir une distinction avec la notion de « modèle d’encodage », qui fait référence à l’ODD et aux règles mises en place.

3. Vocabulaire sur le portail de l’EHRI : <https://portal.ehri-project.eu/vocabularies>.

L'idée derrière la création de ces *templates* est de permettre aux éditeur·ice·s de gagner du temps en leur proposant une structure à compléter. Cela permet de s'assurer que le plus de champs seront renseignés, et d'éviter des confusions entre différents éléments.

```
<place type="" xml:id="">
  <placeName xml:lang="en"/>
  <placeName xml:lang="{original language}"/>
  <location>
    <geo/>
    <country key="{country ID}"/>
  </location>
  <idno type="geonames"/>
  <idno type="wikidata"/>
  <note/>
</place>
```

FIGURE 6.1 – *Template* d'une entrée dans l'index de lieux

La Figure 6.1 représente le *template* d'entrée dans l'index de lieux. Il est bien sûr possible d'ajouter des éléments `<placeName>` pour chaque langue si le nom du lieu a été traduit, tout comme il est possible de supprimer l'élément `<placeName xml:lang="en">` si le nom du lieu est typique de sa langue originale et n'a jamais été traduit. L'objectif n'est pas de proposer une traduction, mais de la renseigner, si elle existe, pour faciliter la recherche d'information. Nous estimons qu'il est important d'établir un lien entre les index des éditions et les bases de données publiques largement utilisées comme Geonames<sup>4</sup> et Wikidata<sup>5</sup>, lorsque cela est possible.

---

4. Geonames : <https://www.geonames.org/>.

5. Wikidata : <https://www.wikidata.org/>.

## Troisième partie

### Publication et valorisation





# Chapitre 7

## Choix de l'outil de publication

Nous avons mentionné deux formes que peuvent prendre les éditions numériques : un *ebook* ou un site web. Selon la méthode de publication choisie, des connaissances en programmation et en développement web peuvent s'avérer nécessaires. Nous en verrons trois : la création d'un site à partir des fichiers XML grâce à une transformation XSLT, la création d'une application web, et l'utilisation d'une plateforme de publication.

### 7.1 Méthodes de publication

#### 7.1.1 Transformation XSLT vers HTML

XSLT est le langage de transformation de l'environnement XML. Un fichier XSL est une feuille de style qui contient des « *templates*<sup>1</sup> ». Cette feuille de style permet de transformer un fichier XML. Le fichier de sortie peut être un autre fichier XML ou dans un autre format ; dans notre cas, il s'agirait d'un fichier HTML.

Pour opérer une transformation, il faut indiquer au processeur XSL la position de l'élément à transformer grâce à une expression XPath. Pour chaque type de page web que nous souhaitons créer, il faudra créer un *template* contenant les règles de transformation. Par exemple, dans le cas des éditions EHRI, il faudrait créer des *templates* pour la page d'accueil générale des éditions, pour les collection, et pour l'affichage du document. L'élément `<xsl:template>` contient donc le code HTML que l'on souhaite voir apparaître dans le fichier de sortie, où chaque élément qui dépend du fichier d'entrée (son titre, par exemple) est identifié par son expression XPath. Cela permet de contrôler le niveau de l'arbre XML où sont appliquées les transformations<sup>2</sup>.

---

1. Elliotte Rusty Harold et W. Scott Means, *XML in a Nutshell*, 2<sup>e</sup> éd., 2002, URL : <https://docstore.mik.ua/oreilly/xml/xmlnut/index.htm> (visité le 28/08/2023), chapitre 8.

2. *Ibid.*

### 7.1.2 Création d'une application web

La création d'une application web requiert des connaissances en programmation. Plusieurs langages sont disponibles pour le développement d'une application web : Python, JavaScript, PHP, Ruby, etc. Nous nous concentrerons sur le langage Python.

Comme de nombreux langages de programmation, Python dispose de *frameworks*. Un *framework* permet de développer une application grâce à des outils intégrés. Par exemple, le *framework* Flask<sup>3</sup> permet de créer des « routes » (URL) qui afficheront la page web associée au *template* HTML. Cette méthode suppose la création d'une base de données dans laquelle Flask ira chercher les informations pour créer l'affichage du site, ce qui nécessite des compétences supplémentaires en gestion de bases de données.

### 7.1.3 Utilisation d'une plateforme de publication

Enfin, la troisième méthode, qui est sûrement la plus populaire, est le recours à un système de gestion de contenu (CMS). Cette méthode consiste en l'utilisation d'outils permettant aux chercheur·euse·s de publier leur édition numérique sans devenir développeur·euse·s. Nous n'évoquerons dans ce mémoire que deux solutions : Omeka et TEI Publisher.

## 7.2 Les éditions EHRI et le logiciel Omeka

Pour la publication de leurs éditions en ligne, l'EHRI a opté en 2018 pour l'utilisation du logiciel *open source* Omeka<sup>4</sup>. C'est un CMS, « dédié à l'organisation, à l'exposition, à la mise en ligne de données iconographiques, avec leur métadonnées, qui permet d'en faire très facilement une publication sur le Web<sup>5</sup> ». Il est également précisé qu'Omeka est « spécialisé dans l'édition de collections muséales, de bibliothèques numériques et d'éditions savantes en ligne<sup>6</sup>. » Il existe deux versions disponibles d'Omeka avec des fonctionnalités différentes : « Omeka Classic », que nous aborderons dans ce mémoire, et « Omeka S », davantage orienté vers le web sémantique<sup>7</sup>.

---

3. Documentation de Flask (version 2.3) : <https://flask.palletsprojects.com/en/2.3.x/>.

4. Omeka : <https://omeka.org/>.

5. Cécile Boulaire et Roméo Carabelli, « Du digital naïve au bricoleur numérique : les images et le logiciel Omeka », dans *Expérimenter les humanités numériques*, dir. Étienne Cavalié, Frédéric Clavert, Olivier Legendre et Dana Martin, Montréal, 2017, p. 81-103, DOI : [10.4000/books.pum.11115](https://doi.org/10.4000/books.pum.11115), p. 81.

6. Fatiha Idmhand, Claire Riffard et Richard Walter, « L'édition électronique de manuscrits modernes », dans *Expérimenter les humanités numériques*, dir. Étienne Cavalié, Frédéric Clavert, Olivier Legendre et Dana Martin, Montréal, 2017, p. 105-123, DOI : [10.4000/books.pum.11118](https://doi.org/10.4000/books.pum.11118), p. 115.

7. Pour en savoir plus sur Omeka : <https://omeka.org/about/project/>.

### 7.2.1 Fonctionnement d'Omeka Classic

Omeka permet la gestion, la valorisation et la diffusion d'importantes collections numérisées<sup>8</sup>. C'est un outil pensé pour la réalisation d'édition « numérisées<sup>9</sup> » (« *digitized edition* ») et non numériques (« *digital edition* »)<sup>10</sup>.

Deux utilisations se distinguent de l'utilisation d'Omeka. Le logiciel permet de « représenter de manière structurée et normalisée des connaissances sur des documents<sup>11</sup> », mais également de « construire des parcours raisonnés à l'aide de pages web dans une sélection ou une collection de documents<sup>12</sup> », et ce à des fins de partage ou de conservation.

Un « contenu Omeka<sup>13</sup> » est composé d'un document et de ses métadonnées ; le document lui-même pouvant être composé de plusieurs fichiers. Les métadonnées sont signalées à l'aide du standard Dublin Core et décrivent le contenu (titre, sujet, description, source, langue, relation et couverture), la propriété intellectuelle (créateur, contributeur, éditeur et gestion des droits) et l'instanciation (date, type, format, identifiant de la ressource)<sup>14</sup>.

### 7.2.2 Limites posées par le recours à Omeka

Pour les chercheur·euse·s souhaitant publier proprement leur édition sans avoir à apprendre la programmation, l'utilisation d'Omeka présente de nombreux avantages. Le logiciel permet notamment de gérer un volume conséquent de données et de valoriser son contenu par la création d'« expositions virtuelles, [de] cartes et [de] frises chronologiques<sup>15</sup> ».

Pourtant, et spécifiquement pour la création d'éditions numériques, Omeka présente un inconvénient non négligeable : le logiciel n'est pas conçu pour traiter des documents encodés en TEI. Ce sont les équipes des projets travaillant sur Omeka qui ont développé des modules (*plug-ins*) pour permettre le traitement et l'affichage de ces éditions<sup>16</sup>, alors que le standard TEI est le plus utilisé par l'édition scientifique numérique.

---

8. Elina Leblanc, « Omeka Classic, Un environnement de recherche pour les éditions scientifiques numériques », *RIDE*–11 (2020), p. 1-28, DOI : [10.18716/ride.a.11.3](https://doi.org/10.18716/ride.a.11.3), p.1.

9. P. Sahle, « What is a Scholarly Digital Edition ? »..., p. 27.

10. E. Leblanc, « Omeka Classic »..., p. 2.

11. Christine Daussin, *Présentation d'Omeka*, Association des usagers francophones d'Omeka, 11 avr. 2019, URL : <https://omeka.fr/document/38> (visité le 30/06/2023), diapo 7.

12. *Ibid.*

13. *Ibid.*, diapo 8.

14. Pour plus d'informations sur le standard Dublin Core : <https://www.bnf.fr/fr/dublin-core> (visité le 02/09/2023).

15. E. Leblanc, « Omeka Classic »..., p. 18-19.

16. *Ibid.*, p. 4.

À titre d'exemple, l'équipe éditoriale du WP12 de l'EHRI a dû mettre au point une correspondance entre certains éléments du `<teiHeader>` et les champs du Dublin Core<sup>17</sup> pour pouvoir travailler avec Omeka<sup>18</sup>. Cette opération s'apparente donc à une sorte de « bricolage numérique<sup>19</sup> » et nécessiterait systématiquement la présence d'un·e ingénieur·e dans l'équipe du projet, ce qui est loin d'être le cas en Sciences Humaines et Sociales<sup>20</sup>.

Enfin, les modules développés par les équipes de projets pour ajouter des fonctionnalités à Omeka en fonction de leurs besoins peuvent ne plus être compatibles avec le logiciel au fur et à mesure de ses mises à jour. En plus du lourd travail de développement nécessaire à la création d'un module, celui-ci doit être maintenu et la compatibilité avec la version courante d'Omeka doit être assurée<sup>21</sup>. Or, ce sont des missions que les chercheur·euse·s ne peuvent pas assurer eux·elles-mêmes, d'une part parce que c'est une activité très chronophage, et d'autre part parce que cela relève d'un domaine qui n'est pas le leur.

## 7.3 Une autre alternative : TEI Publisher

### 7.3.1 Fonctionnement de TEI Publisher

TEI Publisher se présente comme un outil permettant aux chercheur·euse·s et aux éditeur·ice·s de publier leurs éditions numériques sans apprendre la programmation, tout en laissant une place importante à la personnalisation. Son objectif est de fournir à ses utilisateur·ice·s une structure fonctionnelle, bien pensée et personnalisable<sup>22</sup>.

Une fois installée, la plateforme est dotée d'une « aire de jeu<sup>23</sup> » pour comprendre le fonctionnement de l'outil. L'affichage du document dépend des réglages de l'ODD<sup>24</sup> choisie. TEI Publisher prend en charge toutes sortes de documents en entrée, et pas seulement des documents encodés en TEI<sup>25</sup>, bien qu'il s'agisse du standard recommandé.

17. M. Frankl, M. Bryant, J. Green, *et al.*, *Edition of documents...*, p. 7-9.

18. E. Leblanc, « Omeka Classic »..., p. 9.

19. C. Boulaire et R. Carabelli, « Du digital naïve au bricoleur numérique : les images et le logiciel Omeka »..., p. 96.

20. *Ibid.*, p. 98.

21. E. Leblanc, « Omeka Classic »..., p. 16.

22. Documentation TEI Publisher : <https://teipublisher.com/exist/apps/tei-publisher/doc/documentation.xml?odd=docbook&view=div> (visité le 02/09/2023).

23. Disponible depuis la page d'accueil de la plateforme : <https://teipublisher.com/exist/apps/tei-publisher/index.html?query=&collection=playground&sort=title&field=text&start=1>.

24. À ne pas confondre avec la documentation de l'encodage du projet d'édition. L'ODD de TEI Publisher permet de gérer l'affichage du document.

25. Floriane Chiffolleau, *Publication of my Digital Edition – Working with TEI Publisher*, Digital Intellectuals, 4 déc. 2020, URL : <https://digitalintellectuals.hypotheses.org/3912> (visité le 17/05/2023).

La personnalisation de son application s'effectue à l'aide de modules appelés « *web components* »<sup>26</sup>.

### 7.3.2 Pourquoi choisir TEI Publisher

TEI Publisher est un outil *open source* et bien documenté. Sa communauté est très active et la plateforme est mise à jour régulièrement<sup>27</sup>. De nombreux projets d'éditions numériques utilisent désormais TEI Publisher, et les nouvelles fonctionnalités développées pour un projet en particulier sont mises à disposition des autres utilisateur·ice·s. C'est un avantage non négligeable par rapport au logiciel Omeka.

La capacité de l'outil à prendre en charge des formats comme le DOCX<sup>28</sup> est un avantage pour les utilisateur·ice·s qui n'auraient que peu, voire aucunes, compétences en encodage TEI. À titre d'exemple, les éditeurs du WP12 de l'EHRI encodent leurs documents à l'aide d'un traitement de texte simple avec des liens hypertextes. Le fichier est ensuite transformé en XML-TEI avec l'outil « Odette »<sup>29</sup>. En utilisant TEI Publisher, cette étape deviendrait superflue et le travail s'en trouverait accéléré.

En outre, il nous semble important de souligner que TEI Publisher propose un affichage horizontal des documents (Figure 7.1). Les éditions en ligne de l'EHRI disposent d'un affichage vertical (Figure 7.2) centré sur la transcription. Or, cet affichage ne nous semble pas des plus pratiques, notamment en raison du peu d'importance qui est accordé au fac-similé.

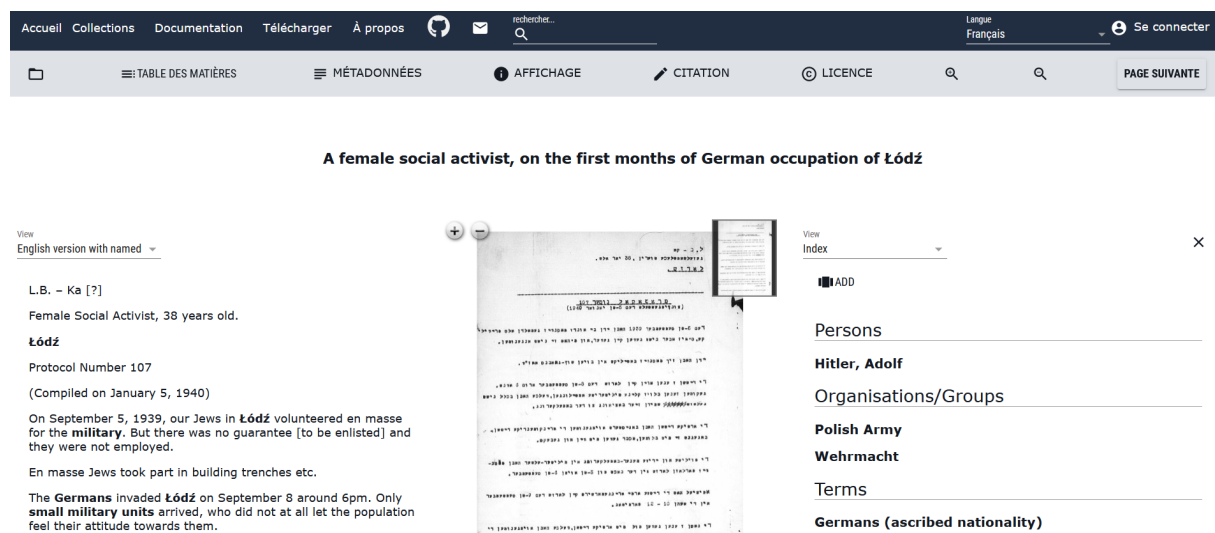


FIGURE 7.1 – Disposition horizontale avec TEI Publisher (Source : DiScholEd)

26. Documentation sur les modules : <https://cdn.tei-publisher.com/@2.12.6/dist/api.html>.

27. La dernière version de TEI Publisher (8.0) a été publiée le 28 mars 2023 : <https://www.e-editiones.org/posts/tei-publisher-8/>.

28. Format du traitement de texte Microsoft Word.

29. Odette : <https://obvil.huma-num.fr/odette/>.

< Ray Johnston on the reaction of the... Alexander Khan on the deposition of... >

## John F. Montgomery on the "Hungarian Jew Bill" in 1939

**METADATA** **DOCUMENT TEXT** **MAP** **REFERENCES**

**METADATA**

Q43-Q5-16940219-02

1939-Q5-19 | Subject

National Archives and Records Administration (, RG 22--Records of the Department of State relating to the internal affairs of Hungary 1919-1944, Wilmington, Del. 1937 (Montefiore), Doc. 864-0741-101. Original in English.


**DOCUMENT TEXT**

**ENGLISH**

Text from page2

The Department will note that the law was modified substantially. Primarily, the class of Jews to be considered officially as Christians was enlarged but the economic restrictions were increased, thus creating a sort of second-class citizenship. This represents the real compromise made between the two houses of Parliament. Also, the idea of placing the Jews in a distinct category for political purposes was discarded. Instead, the Jews retain free active and passive suffrage, but the conditions of their enfranchisement were made much more severe. Agricultural property may be taken from the Jews, their right to acquire real estate is limited, and the restriction of their employment is greatly increased. Jewish journalists complain bitterly against the special press restriction by which the law was amended. It appears that, in practice, Jewish youth will suffer the most.

**MAP** [Toggle fullscreen](#)



**REFERENCES**

**United States**  
Updated 4 years ago

The United States of America were neutral during the first two years of the Second World War. They were brought into the conflict by the Japanese attack on Pearl Harbor on 7 December 1941 and became one of the decisive belligerent nations, defeating Japan in the Pacific and heavily contributing to the war effort against the Axis powers, e.g. during the Allied landings in North Africa, Italy and France. American forces invaded German territory early in 1945. On the eve of the Second World War...

[All repository descriptions outside](#)

[View this item on the EHRP Portal](#)

**National Archives and Records Administration**  
NARA | National Archives

**United States**. 1937. Raleigh Road. College Park. Updated 8 months ago

[View this item on the EHRP Portal](#)

FIGURE 7.2 – Disposition verticale avec Omeka (Source : EHRI Online Editions)

# Chapitre 8

## EHRI sur l'application DiScholEd

### 8.1 Une application pour publier les ego-documents

#### 8.1.1 Qu'est-ce qu'un ego-document ?

La mention d'« ego-document » apparaît dans les années 1950, utilisée par l'historien néerlandais Jacques Presser pour décrire les textes sur lesquels se basent ses recherches : des journaux intimes, mémoires, autobiographies et de la correspondance personnelle (« *Ego-Dokument*<sup>1</sup> »). Le terme « ego-document » est plus largement employé dans le monde anglophone ; en France, on entendra volontiers parler d'« écrits du for privé<sup>2</sup> » ou d'« écriture de soi ». L'étude des ego-documents s'inscrit dans le courant historiographique de l'« Histoire par le bas » (« *History From Below* »).

#### 8.1.2 Les collections publiées sur DiScholEd

À ce jour, l'application DiScholEd<sup>3</sup> compte sept collections. Nous présenterons ici les plus documentées par l'application. L'une d'entre elles, la collection « *Early Holocaust Testimony* de l'EHRI », nous intéressera particulièrement.

### Correspondance de Paul d'Estournelles de Constant

Paul d'Estournelles de Constant (1852-1924) était un homme politique français. Il devient sénateur en 1905, fonction qu'il occupera jusqu'à sa mort. Prix Nobel de la paix, il a œuvré pour le maintien de la paix entre les pays, notamment avec la Grande-Bretagne

---

1. Alicia Altred Vigil, « Les ego-documents comme source pour l'histoire », trad. par Allison Taillot, *Matériaux pour l'histoire de notre temps*–143-144 (2022), p. 45-51, DOI : [10.3917/mate.143.0045](https://doi.org/10.3917/mate.143.0045), p. 46.

2. Juliette Deloye, « For privé (écrits du) » dans VOCES, *Vocabulaire pour l'Étude des Scripturalités*, Thomas Brunner (dir.), ARCHE UR3400 (Université de Strasbourg), édition électronique, 2019, DOI : [10.34931/vvxh-5046](https://doi.org/10.34931/vvxh-5046) (visité le 03/09/2023).

3. DiScholEd, *Éditions scientifiques numériques*, URL : <https://discholed.huma-num.fr/exist/apps/discholed/index.html>.

et l'Allemagne. Le corpus est composé de sa correspondance pendant la Première Guerre mondiale. Paul d'Estournelles de Constant y apporte son point de vue en tant qu'élus, mais aussi en tant que citoyen et père de famille<sup>4</sup>.

### Intellectuels Berlinois

Le collection « Lettres et textes : Le Berlin intellectuel des années 1800 » témoigne de la vie intellectuelle berlinoise de la fin du XVIII<sup>e</sup> et du début du XIX<sup>e</sup> siècle. Le corpus met en avant le développement du romantisme allemand et les relations que les cercles intellectuels berlinois entretenaient entre eux<sup>5</sup>.

### Catalogue des manuscrits d'Auguste Boeckh

Auguste Boeckh (1785-1867) était un philologue allemand. Il a enseigné à l'université de Berlin, avant de devenir le doyen de la faculté des Arts et Sciences Humaines. La collection est composée du catalogue de ses manuscrits, réunis dans le cadre du projet « *Boeckh-Nachlassprojekt*<sup>6</sup> ».

## 8.2 La collection « *Early Holocaust Testimony* » sur DiScholEd

Le contenu de la collection « *Early Holocaust Testimony* » sur DiScholEd est identique à celui de l'édition publiée avec Omeka<sup>7</sup>, seul l'affichage est différent. La simplicité de l'affichage proposé par TEI Publisher nous permet de présenter l'édition en donnant accès à l'utilisateur·ice aux informations sur l'édition et sur l'EHRI, aux textes édités et aux index directement depuis la page d'accueil (Figure 8.1). La présentation de la collection (section « Texts édités ») permet de naviguer directement dans toute la collection (Figure 8.2), alors que le site Omeka ne propose que d'explorer les exemples proposés par les éditeur·ice·s dans un premier temps.

La disposition horizontale (Figure 7.1) permet d'afficher plusieurs vues d'une importance équivalente, alors que la présentation de la transcription sur Omeka se fait au détriment du fac-similé (Figure 7.2), dont l'affichage nous semble tout particulièrement pertinent au sein d'une édition numérique de documents d'archives.

4. *Ibid.*, collection « Correspondance de Paul d'Estournelles de Constant », section « *History of the Corpus* » (visité le 03/09/2023).

5. *Ibid.*, collection « Lettres et textes : Le Berlin intellectuel des années 1800 », section « *History of the Corpus* » (visité le 03/09/2023).

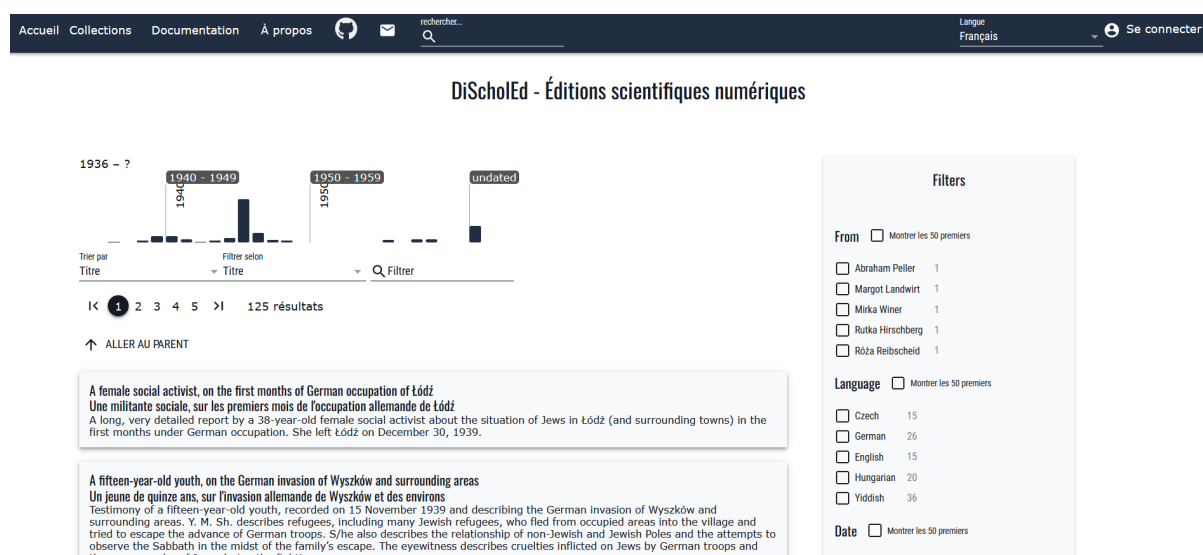
6. *Ibid.*, collection « Catalogue des manuscrits d'Auguste Boeckh », section « *History of the Corpus* » (visité le 03/09/2023).

7. Consultable à l'adresse : <https://early-testimony.ehri-project.eu/>.





FIGURE 8.1 – Page d'accueil de la collection sur DiScholEd

FIGURE 8.2 – Collection « *Early Holocaust Testimony* » sur DiScholEd

La publication de la collection « *Early Holocaust Testimony* » sur DiScholEd a permis à l'équipe éditoriale du WP12 d'avoir une idée du rendu des éditions qu'elle pourrait y publier. La compatibilité naturelle de TEI Publisher avec les fichiers encodés en TEI permet à la communauté de développer des modules plus simples à incorporer à son projet. En outre, TEI Publisher dispose d'un outil d'annotation intégré dont la maîtrise optimiserait davantage le travail éditorial.



# Chapitre 9

## Transition vers une application dédiée

### 9.1 Analyse des besoins

#### 9.1.1 Profil des utilisateur·ice·s

L'objectif principal de l'EHRI est de soutenir la recherche sur la Shoah, notamment en éditant des corpus thématiques. Les collections rassemblent des documents d'archives conservés dans les institutions partenaires du projet. De ce fait, les éditions de l'EHRI s'adressent plutôt à un public de chercheur·euse·s et d'étudiant·e·s, particulièrement en Histoire.

#### 9.1.2 Transposer les fonctionnalités d'Omeka

Le cahier des charges des éditions EHRI publiées avec Omeka a été rigoureusement étudié<sup>1</sup>. La transition vers TEI Publisher ne peut se faire sans s'assurer que les fonctionnalités dont dispose Omeka sont également disponibles dans TEI Publisher. Nous avons donc identifié quels étaient les principaux critères recherchés par l'équipe éditoriale, puis nous avons parcouru la liste des *web components* de TEI Publisher pour établir une correspondance :

- La possibilité d'ajouter une documentation à propos de l'édition (module `pb-mark-down`) ;
- Une barre de recherche (module `pb-search`) ;
- Une recherche facettée (modules `pb-browse-docs` et `pb-custom-form`) ;
- Une transcription lisible et sans distraction (modules `pb-view`, `pb-collapse`, `pb-grid` et `pb-grid-action`) ;

---

1. M. Frankl, M. Bryant, J. Green, *et al.*, *Edition of documents...*

- La possibilité de télécharger les fichiers dans plusieurs formats, principalement XML, PDF et ePub (module `pb-download`);
- La génération de cartes interactives (modules `pb-leaflet`, `pb-geolocation`, `pb-map-icon` et `pb-map-layer`);
- L’affichage d’informations contextuelles au survol de la transcription (modules `pb-highlight` et `pb-popover`).

## 9.2 Création du prototype d’application

### 9.2.1 Des fonctionnalités supplémentaires

Notre exploration des modules mis à disposition par TEI Publisher nous a permis d’établir une liste de fonctionnalités supplémentaires que l’application TEI Publisher dédiée à EHRI pourrait présenter :

- Une application disponible en plusieurs langues (module `pb-lang`);
- Une chronologie des documents (module `pb-timeline`);
- Une navigation claire à l’intérieur de la collection (module `pb-paginate`), ce que toutes les éditions EHRI avec Omeka ne proposent pas;
- Une navigation claire et simple à l’intérieur du document, notamment entre le fac-similé et la transcription (module `pb-navigation`);
- La possibilité de zoomer aisément sur le fac-similé sans perte de qualité (module `pb-facsimile`);

Certaines de ces fonctionnalités sont déjà disponibles sur l’application DiScholEd et pourront donc être facilement implémentées.

### 9.2.2 Maquettes visuelles

À partir des fonctionnalités décrites, nous avons proposé des maquettes visuelles pour le futur développement de l’application :

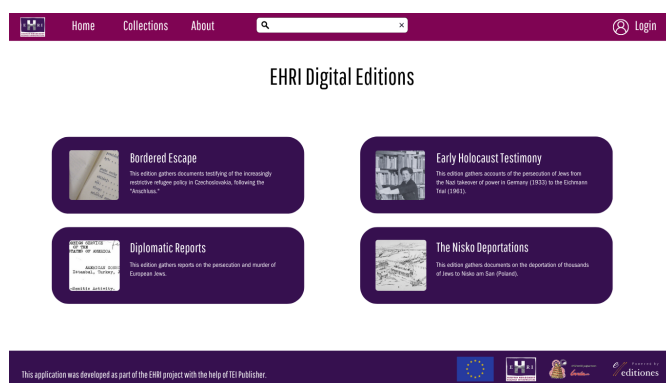


FIGURE 9.1 – Maquette de la page d’accueil EHRI avec TEI Publisher

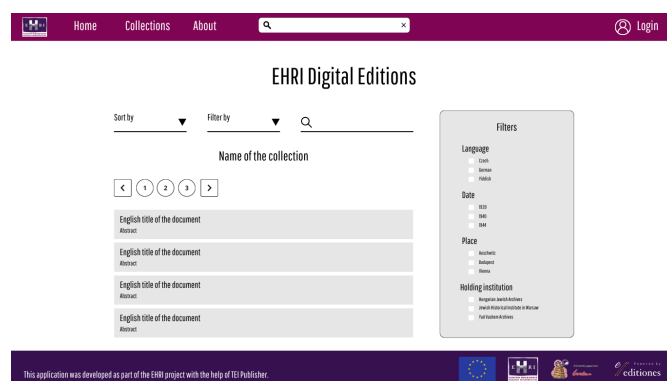


FIGURE 9.2 – Maquette de navigation dans la collection

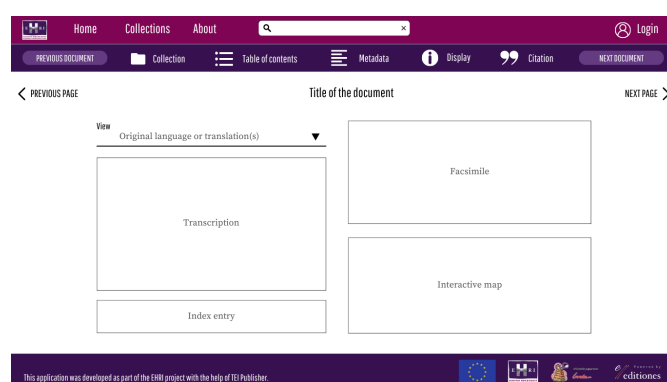


FIGURE 9.3 – Maquette de l'affichage du document avec TEI Publisher

Notre équipe a commencé le développement de l'application TEI Publisher EHRI, dont le prototype devrait être présenté aux éditeur·ice·s du WP12 entre octobre et novembre prochains.

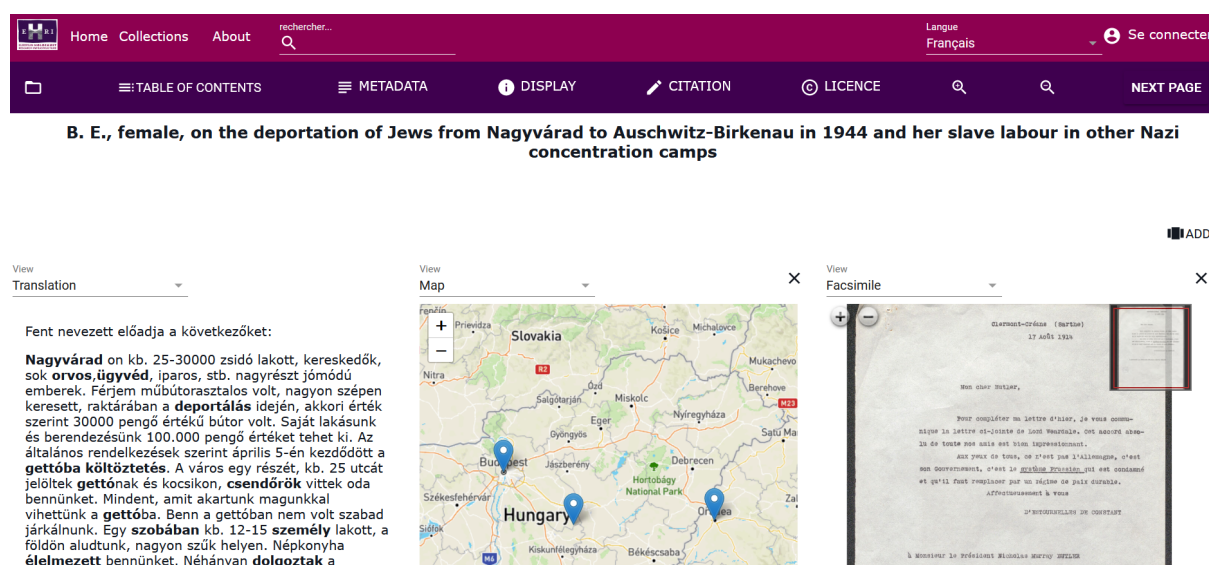


FIGURE 9.4 – Exemple de l'affichage d'un document sur l'application TEI Publisher EHRI

Grâce à son affichage horizontal, TEI Publisher permet de profiter de trois vues simultanément sans perdre en confort de lecture (Figure 9.4). Les différentes vues sont par ailleurs modulables, c'est-à-dire que l'utilisateur·ice peut choisir le nombre de vues souhaité et ce qu'elles affichent.

# Conclusion

La préparation d'une édition scientifique numérique est un processus long et passionnant. Leur publication permet de créer de formidables outils pour la recherche. Notre expérience nous a confirmé que la définition d'un modèle d'encodage est une étape primordiale afin d'assurer non seulement la validité et la conformité d'une édition, mais également l'uniformité à travers les collections. Le rôle de l'ingénieur·e à cet égard est extrêmement important, notamment pour conseiller les chercheur·euse·s et les aider à naviguer parmi les outils, formats et standards.

Les éditions de l'EHRI bénéficiant d'un aperçu de deux outils de publication, TEI Publisher et Omeka, les avantages de chacun d'eux ont pu être établis par rapport à un même corpus. Le développement de l'application TEI Publisher pour EHRI suit son cours, et les éditeur·ice·s pourront bientôt expérimenter l'outil d'annotation intégré à TEI Publisher, qui devrait encore davantage les aider dans leur travail éditorial.





# Annexe A

## Glossaire

Nous avons rédigé les définitions de ce glossaire à l'aide des références suivantes :

- BENSOUSSAN (Georges), *Histoire de la Shoah*, 7<sup>e</sup> éd., Paris, 2020 (Que sais-je ?), URL : <https://www.cairn.info/histoire-de-la-shoah--9782715404557.htm> (visité le 12/07/2023).
- BRUTTMANN (Tal), *Les 100 mots de la Shoah*, avec la coll. de Christophe Tarricone, 3<sup>e</sup> éd., Paris, 2020 (Que sais-je ?), URL : <https://www.cairn.info/les-100-mots-de-la-shoah--9782715404588.htm> (visité le 12/07/2023).

**Antisémitisme** Idéologie selon laquelle « le Juif » serait responsable de tous les maux. L'antisémitisme prend la forme de discrimination, voire de violences, à l'encontre des Juif·ve·s en raison de leur ethnie.

**Aryanisation** Processus visant à transférer la propriété d'une personne juive à un « aryen ». Le terme renvoie à une « opération de décontamination ». Il y a eu deux phases à l'aryanisation : l'« aryanisation volontaire » et l'« aryanisation forcée ». Cette deuxième phase commence le 14 juin 1938 lorsque le ministre de l'Intérieur, Wilhelm Frick, décide de transférer les biens et propriétés juifs à l'État pour les revendre à crédit aux classes moyennes allemandes.

**Camp de concentration** (*Konzentrationslager*). Symbole de la répression nazie, le camp de concentration est à l'origine un camp d'internement des opposants au régime. Le premier camp de concentration nazi est celui de Dachau en Allemagne.

**Camp d'extermination** (Centre de mise à mort). Après les expérimentations de la mort par le gaz menés sur les malades mentaux lors de l'« Aktion T4 », la méthode est utilisée dans les camps pour exterminer le plus de victimes en un minimum de temps. Le premier camp d'extermination est celui de Chelmno, ouvert le 8 décembre 1941.

**Déportation** Déplacement d'une personne ou d'un groupe de personne contre son gré. Durant la Shoah, les Juif·ve·s étaient déporté·e·s dans les ghettos, puis dans les camps de concentration et les camps d'extermination.

**Groupes d'action spéciale** (*Einsatzgruppen*). Unités mobiles du Reich allemand, chargées d'assassiner les Juif·ve·s et les commissaires politiques communistes. Au début, les groupes sont composés de presque 3 000 hommes, tous volontaires.

**Émigration** Action volontaire de quitter son lieu de résidence ou son pays. L'objectif premier du régime nazi était de forcer les populations juives à émigrer hors d'Allemagne par le biais d'intimidations. Cette politique s'est soldée par un échec certain avec l'expansion du Reich, menant ensuite à la politique d'extermination de la population juive d'Europe.

**Espace vital** (*Lebensraum*). Concept du pangermanisme selon lequel le peuple allemand (comprendre les « aryens ») doit disposer d'un espace suffisant pour vivre. L'extension du territoire est vue comme une nécessité, ce qui permet au régime nazi de justifier ses invasions successives.

**Génocide** Destruction partielle ou totale intentionnelle d'une population en raison de son ethnie ou de sa religion.

**Gestapo** Abréviation de « *Geheime Staatspolizei* ». Police politique de l'État nazi.

**Ghetto** Quartier imposé aux Juif·ve·s par l'État. La population juive est forcée de vivre à l'écart de la population non-juive. Dans l'Allemagne nazie, les ghettos sont clôturés et constamment surveillés. La densité de population y est très élevée et les conditions de vies difficiles. Les autorités y pratiquent l'extermination par la faim et l'épidémie. Avec la « Solution finale », les ghettos sont devenus une étape intermédiaire avant la déportation des Juif·ve·s vers les camps d'extermination.

**Lois de Nuremberg** Distinguent la citoyenneté de la nationalité. « Le Juif » est défini par ses ascendants (au moins trois grands-parents juifs) et ne peut pas être citoyen. Les mariages et relations sexuelles entre Juif·ve·s et non-Juif·ve·s sont interdits.

**Nazisme** Abréviation de national-socialisme. Idéologie antidémocratique et anti-marxiste, nationaliste et pangermaniste.

**Nuit de Cristal** (*Kristallnacht*). Suite à l'assassinat d'un conseiller d'ambassade allemand le 7 novembre 1938 à Paris, Joseph Goebbels prononce un discours qui incite la population à une grande violence antisémite. Des *pogroms* ont lieu partout : 267 synagogues sont pillées, saccagées et incendiées, 7 500 magasins sont dévalisés ou détruits, les habitations des Juif·ve·s sont saccagées. La « Nuit de Cristal » dure jusqu'au 10 novembre, dans l'après-midi. Une centaine de Juif·ve·s sont assassinés, les femmes sont violées malgré l'interdit racial et les enfants

Juif·ve·s sont chassé·e·s des orphelinats. Près de 11 000 hommes sont arrêtés et envoyés à Dachau, et 10 000 à Buchenwald.

**Pogrom** Terme russe signifiant « dévastation » ou « destruction ». Désigne les violences antisémites commises par la population, incitée par le pouvoir en place. Le régime nazi a provoqué des *pogroms* à de nombreuses reprises.

**Question juive** Expression faisant originellement référence à la capacité des Juif·ve·s à s'intégrer en Europe occidentale. Pour le régime nazi, la « question juive » est un problème auquel il faut apporter une solution, qui consiste en l'extermination de la population juive.

**Shoah** (Holocauste). Terme hébreu signifiant « catastrophe », synonyme du terme « Holocauste » qui, à l'origine, fait référence à un sacrifice dans un but religieux. Le terme fait aujourd'hui référence au génocide des Juif·ve·s d'Europe par le régime nazi. Les anglophones emploient le terme « *Holocaust* », tandis que les francophones parlent de « Shoah ».

**Solution finale** Projet d'extermination des Juif·ve·s d'Europe par le régime nazi. Ce « projet » est envisagé comme solution au « problème » posé par la « question juive ». La Solution finale accélère le processus entamé par l'enfermement des Juif·ve·s dans les ghettos et leur déportation dans les camps de concentration. Les premiers camps d'extermination sont créés dès 1941 et pratiquent la mort par le gaz et par balles.

**Spoliation** Désigne la confiscation des biens possédés par les Juif·ve·s. La politique de spoliation orchestrée par le régime nazi a commencé par l'identification de tous les Juif·ve·s, qui a été suivie par une série de lois visant à les déposséder et anéantir leurs droits.



## Annexe B

# Chronologie de la persécution des Juifs

Cette chronologie non exhaustive a été établie à partir des ressources suivantes :

- BENSOUSSAN (Georges), *Histoire de la Shoah*, 7<sup>e</sup> éd., Paris, 2020 (Que sais-je ?), URL : <https://www.cairn.info/histoire-de-la-shoah--9782715404557.htm> (visité le 12/07/2023).
- BRUTTMANN (Tal), *Les 100 mots de la Shoah*, avec la coll. de Christophe Tarricone, 3<sup>e</sup> éd., Paris, 2020 (Que sais-je ?), URL : <https://www.cairn.info/les-100-mots-de-la-shoah--9782715404588.htm> (visité le 12/07/2023).
- *Mémorial de la Shoah*, URL : <https://www.memorialdelashoah.org/>.

**XIII<sup>e</sup> siècle** Création des premiers ghettos et premières grandes migrations de l'Europe occidentale vers l'Europe de l'Est, en particulier vers la Pologne et la Lituanie.

**1215** Le IV<sup>e</sup> concile du Latran impose aux Juifs le port d'un signe distinctif : la « rouelle » (petit pièce de tissu jaune).

**15 novembre 1879** Heinrich Gothard von Treitschke publie un article antisémite dans les *Annales prussiennes*, dont les Nazis reprendront la formule : « Les Juifs sont notre malheur ! ».

**1881-1882** En Russie, les *pogroms* accélèrent l'émigration des Juifs de Russie.

**Années 1900** Montée du darwinisme social et de l'idéologie eugéniste en Allemagne.

**5 janvier 1919** Création du Parti ouvrier allemand, qui devient le Parti national-socialiste des travailleurs allemands (NSDAP) en 1921.

**Années 1920** Montée de la xénophobie et de l'antisémitisme. La défaite de l'Allemagne à l'issue de la Première Guerre mondiale pousse la classe moyenne allemande vers le nationalisme d'extrême-droite.

**1919-1921** Nombreux *pogroms* en Pologne.

**24 février 1920** Premier programme politique du NSDAP. Il entend retirer aux Juifs la citoyenneté allemande et réduire leurs droits à ceux des étrangers. Un point révoque l'accès aux emplois de la fonction publique pour les Juifs, et un autre souhaite expulser les Juifs entrés en Allemagne après le 2 août 1914.

**18 juillet 1925** Publication de *Mein Kampf* (Adolf Hitler).

**30 janvier 1933** Adolf Hitler est appelé à la Chancellerie.

**5 mars 1933** Élections législatives anticipées. Le NSDAP obtient 44 % des suffrages.

**27 février 1933** Incendie du Reichstag.

**22 mars 1933** Ouverture du premier camp de concentration à Dachau (Allemagne).

**23 mars 1933** Hitler obtient les pleins pouvoirs.

**Avril 1933** Les fonctionnaires juifs sont révoqués et les avocats juifs sont radiés du barreau.

**26 avril 1933** Création de la Gestapo.

**14 juillet 1933** Loi sur la stérilisation forcée. Les personnes atteintes de « maladies héréditaires » (physiques ou mentales) et les criminels « irrécupérables et dangereux » sont stérilisés avec l'appui des psychiatres.

**14 juillet 1933** Le NSDAP devient le seul parti autorisé en Allemagne.

**25 août 1933** L'Accord de « *haavara* » (de l'hébreu « transfert ») permet de négocier le départ de 60 000 Juifs vers la Palestine.

**29 juin 1934** Nuit des longs couteaux.

**Juillet 1934** Création de l'Inspection des camps de concentration.

**1<sup>er</sup> août 1934** Hitler se proclame « Führer » et Chancelier du Reich.

**15 septembre 1935** Lois de Nuremberg. Séparation physique des Juifs des autres Allemands.

**7 mars 1936** Entrée de la Wehrmacht en Rhénanie (zone démilitarisée par le Traité de Versailles en 1919).

**12 juillet 1936** Ouverture du camp de concentration d'Oranienburg-Sachsenhausen.

**15 juillet 1937** Ouverture du camp de concentration de Buchenwald.

**Mars 1938** Aux États-Unis, 82% de la population s'oppose à l'accueil de réfugiés juifs venus d'Allemagne et d'Autriche.

**12–13 mars 1938** Annexion de l’Autriche (*Anschluss*)

**Avril 1938** Création d’un centre d’émigration juive à Vienne, dirigé par Adolf Eichmann. En six mois, le quart de la communauté juive est expulsée.

**26 avril 1938** Les Juifs doivent déclarer tous leurs biens.

**Mai 1938** Première législation antisémite en Hongrie.

**14 juin 1938** Début de l’aryanisation forcée.

**Juillet 1938** Les médecins juifs doivent demander l’autorisation d’exercer et limiter leur pratique à une clientèle juive.

**6–15 juillet 1938** La conférence d’Évian réunit trente-deux états pour répondre à la question des réfugiés juifs. La Hongrie, la Roumanie et la Pologne envoient des « observateurs ». « Nul ne conteste à l’Allemagne sa souveraineté à l’égard de ses nationaux ».

**Août 1938** Les Juifs doivent ajouter un prénom de marquage à leur prénom courant : « Israël » pour les hommes, « Sara » pour les femmes.

**8 août 1938** Ouverture du camp de concentration de Mauthausen (Autriche).

**30 septembre 1938** Accords de Munich.

**9–10 novembre 1938** Nuit de Cristal (*Kristallnacht*).

**Décembre 1938** En Autriche, les Juifs perdent leur permis de conduire.

**3 décembre 1938** Les propriétaires juifs reçoivent l’ordre de vendre tous leurs biens restants.

**31 décembre 1938** Les travailleurs indépendants juifs doivent cesser toute activité.

**Janvier 1939** Création du « Bureau central pour l’émigration des Juifs » à Berlin, dirigé par Reinhard Heydrich.

**Mars 1939** La Suisse ferme ses frontières aux réfugiés juifs.

**15 mars 1939** Invasion de la Tchécoslovaquie et violation des Accords de Munich.

**Avril 1939** En Autriche, les locataires juifs perdent tous leurs droits face à leurs propriétaires.

**Mai 1939** Recensement général de la population allemande. Les fiches des Juifs sont marquées de la lettre « J ».

**Mai 1939** Ouverture du camp de concentration de Ravensbrück.

**Septembre 1939** Couvre-feu à 20h pour les Juifs uniquement.

**1<sup>er</sup> septembre 1939** Invasion de la Pologne par l’Allemagne.

**25 septembre 1939** La Pologne est divisée en quatre zones réparties entre l’Allemagne, l’URSS et la Lituanie.

- 27 septembre 1939** Début de l'« Aktion T4 » à l'asile de Kocborowo (Pologne).
- Octobre 1939** Début des déportations vers Nisko (Pologne).
- 8 octobre 1939** Création du ghetto de Piotrkow.
- 9 octobre 1939** Début du recensement des malades mentaux.
- 26 octobre 1939** Instauration du travail forcé pour les personnes âgées de 14 à 60 ans dans le « gouvernement général » (zone de Pologne occupée par l'Allemagne).
- 28 novembre 1939** Créations des « Conseils juifs » dans le « gouvernement général ». La communauté juive prend elle-même en charge les tâches administratives liées au recensement, à la spoliation et à la déportation de sa population.
- Décembre 1939** Exclusion des Juifs allemands des distributions spéciales de nourritures.
- 11 décembre 1939** Tout changement de domicile est interdit aux Juifs polonais.
- Mars 1940** Identification des Juifs par un « J » sur leur carte alimentaire.
- 20 avril 1940** Création du ghetto de Lodz.
- 9 avril 1940** Invasion du Danemark et du sud de la Norvège.
- 27 avril 1940** Ouverture du camp de concentration d'Auschwitz.
- 2 octobre 1940** Création du ghetto de Varsovie.
- Janvier 1941** Création des « *Einsatzgruppen* ».
- 3 mars 1941** Création du ghetto de Cracovie.
- 24 mars 1941** Création du ghetto de Lublin.
- Avril 1941** Début de l'« Aktion 14f13 ». Mise à mort des détenus incapables de travailler dans les camps de concentration.
- 29 juin 1941** Opération Barbarossa (invasion de la Russie).
- 15 août 1941** Heinrich Himmler se rend à Minsk et décide d'expérimenter le meurtre à l'aide des camions à gaz, déjà utilisés pour l'assassinat des malades mentaux.
- 24 août 1941** Hitler annonce la fin de l'« Aktion T4 ».
- Septembre 1941** Les Juifs allemands de plus de 6 ans doivent porter l'étoile de David sur le côté gauche de leur vêtement. Ils ne peuvent plus quitter leur commune de résidence.
- 5–6 septembre 1941** Les prisonniers de guerre soviétiques sont assassinés par le gaz Zyklon B à Auschwitz.
- 16 et 18 septembre 1941** Premiers essais des camions à gaz à Mohilev et à Minsk (Biélorussie).
- 29 septembre 1941** Massacre de Babi Yar (Ukraine).



- Octobre 1941** Le ghetto de Lodz devint un centre de transit pour les Juifs déportés du Reich.
- 1<sup>er</sup> octobre 1941** Début des travaux visant à transformer le Birkenau (Auschwitz) en camp d'extermination.
- 23 octobre 1941** L'Europe allemande est définitivement fermée à l'émigration juive.
- 1<sup>er</sup> novembre 1941** Construction du centre de mise à mort de Belzec (Pologne).
- 3 novembre 1941** Essais des camions à gaz à Sachsenhausen.
- 8 novembre 1941** Création du ghetto de Lwow (Ukraine).
- 24 novembre 1941** Ouverture du camp de concentration de Terezin (République Tchèque).
- 7 décembre 1941** Attaque de Pearl Harbor.
- 8 décembre 1941** Ouverture du camp d'extermination de Chelmno, près de Lodz. Début des exterminations par le gaz.
- 20 janvier 1942** Conférence de Wansee. Début de la « Solution finale ».
- 17 ars 1942** Ouverture du camp d'extermination de Belzec (Pologne).
- 7 mai 1942** Ouverture du camp d'extermination de Sobibor (Pologne).
- 22 juillet 1942** Ouverture du camp d'extermination de Treblinka (Pologne) et déportation des Juifs du ghetto de Varsovie.
- 13–14 mars 1943** Liquidation du ghetto de Cracovie.
- 19 avril 1943 – 16 mai 1943** Révolte du ghetto de Varsovie.
- 21 juin 1943** Décision de liquider tous les ghettos de Pologne.
- 3 novembre 1943** Début de l'Opération « *Erntefest* » (« Fête de la Moisson » : extermination Juifs de Lublin (environ 45000)).
- 19 mars 1944** Invasion de la Hongrie.
- Août 1944** Liquidation du ghetto de Lodz et déportation des survivants à Auschwitz.
- 30 octobre 1944** Déportation des Juifs de Theresienstadt vers Auschwitz.
- 27 janvier 1945** Libération d'Auschwitz.
- 30 avril 1945** Suicide d'Adolf Hitler.
- 7 mai 1945** Capitulation de l'Allemagne.
- 18 octobre 1945 – 1<sup>er</sup> octobre 1946** Procès de Nuremberg.



# Annexe C

## Tableau récapitulatif d'encodage

Tag Name	header	body	Number of Occurrences				List of @Attributes used in EHRI editions	ODD EHRI	Comment(s)	Files Involved
			BeGrenzte Flucht (104)	Early Testimonies (119)	Diplomatic Reports (72)	Nisko (40)				
ab	<input type="checkbox"/>	<input checked="" type="checkbox"/>				6	@type="dinkus"	<input checked="" type="checkbox"/>	Use <metamark> instead (see <a href="#">Encoding Guidelines</a> ).	EHRI-NISKO-19400102_DE.xml
abstract	<input checked="" type="checkbox"/>	<input type="checkbox"/>	97	113		3		<input checked="" type="checkbox"/>	If the title of the document is not explicit enough, a good practice is to add an <abstract> tag and a summary of the document's content.	
author	<input checked="" type="checkbox"/>	<input type="checkbox"/>	208	293		120		<input checked="" type="checkbox"/>		
availability	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		40		<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection.	
bibl	<input checked="" type="checkbox"/>	<input type="checkbox"/>	135	119	72	36	@type	<input checked="" type="checkbox"/>		
biblScope	<input checked="" type="checkbox"/>	<input type="checkbox"/>	10				@unit	<input type="checkbox"/>	Use <bibl> tag instead.	EHRI-BF-19390202_DE.xml EHRI-BF-19380316_DE.xml EHRI-BF-19380408_DE.xml EHRI-BF-19380418_DE.xml EHRI-BF-19380421c_DE.xml EHRI-BF-19380723_DE.xml EHRI-BF-19380830b_DE.xml
birth	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1				@when	<input type="checkbox"/>	Including the <birth> tag in the ODD is not relevant.	EHRI-BF-19380911_DE.xml
body	<input type="checkbox"/>	<input checked="" type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
catRef	<input checked="" type="checkbox"/>	<input type="checkbox"/>	117			37	@target	<input checked="" type="checkbox"/>	<a href="#">Guidelines:</a> <catRef> (category reference) specifies one or more defined categories within some taxonomy or text typology. Values for the @target attribute include: "Schmuggel" (smuggling), "Schicksale" (fate), "Flüchtlingspolitik" (Refugee Policy), "IKG Wien" (Israelitische Kultusgemeinde Wien), "Berichte" (reports), etc. Values must be normalized to English.	
cell	<input type="checkbox"/>	<input checked="" type="checkbox"/>	27	54			@role	<input checked="" type="checkbox"/>	See <table>.	EHRI-BF-19381231_DE.xml EHRI-ET-WL16560413_DE.xml EHRI-ET-WL16560413_EN_DE.xml EHRI-ET-VY3540315_YI.xml EHRI-ET-VY3540315_EN_YI.xml
collection	<input checked="" type="checkbox"/>	<input type="checkbox"/>	84	113	42	5	@ref	<input checked="" type="checkbox"/>	References the collection to which the document belongs to, and should systematically appear along with the <repository> or <institution> tag in the manuscript description <msDesc>.	
country	<input checked="" type="checkbox"/>	<input type="checkbox"/>	84	113	72	39	@ref	<input checked="" type="checkbox"/>	See <msDesc>.	
creation	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	73	47		<input checked="" type="checkbox"/>	The <creation> tag should only appear once in every file. Files with duplicates are indicated in the "Files Involved" column, and red files also have several occurrences of a <profileDesc> tag.	EHRI-DR-19400712-US_EN.xml EHRI-NISKO-19391015_DE.xml EHRI-NISKO-19391017_DE.xml EHRI-NISKO-19391018_DE.xml EHRI-NISKO-19391212_DE.xml EHRI-NISKO-19400102_DE.xml EHRI-NISKO-19400420_DE.xml EHRI-NISKO-1947111_DE.xml
date	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	365	984	229	133	@from @notBefore @to @type @notAfter @when	<input checked="" type="checkbox"/>	When the date is available in the YYYY-MM-DD format, please use the @when-iso attribute.	

FIGURE C.1 – Balises et attributs utilisés par les éditions EHRI (1/5)

Tag Name	header	body	Number of Occurrences				List of @Attributes used in EHRI editions	ODD EHRI	Comment(s)	Files Involved
			BeGrenzte Flucht (104)	Early Testimonies (119)	Diplomatic Reports (72)	Nisko (40)				
del	<input type="checkbox"/>	<input checked="" type="checkbox"/>	12			5	@rend	<input checked="" type="checkbox"/>	No values for @rend attribute in BeGrenzte Flucht and one occurrence of "overstrike" in Nisko. List of semi-closed values for the @rend attribute (based on the <a href="#">DAHJL Encoding Guidelines</a> ): - overwriten: a correction has been written over (must be used with the <add> tag and the @place attribute (values: "across", "above", "underneath") - (strikethrough): horizontal or vertical line(s) through a word/expression - erased: removed word/passages of the text (left blank) - none: the deletion is only marked by writing the correct word right next to it	EHRI-BF-19380425_DE.xml EHRI-BF-19381020_DE.xml EHRI-BF-19400503_DE.xml EHRI-NISKO-193910a_DE.xml EHRI-NISKO-19391013a_DE.xml
distinct	<input type="checkbox"/>	<input checked="" type="checkbox"/>	2	27			@xml:lang @type	<input checked="" type="checkbox"/>	The <distinct> tag must be used with a @type attribute for specific vocabulary. The value of said @type attribute should be either "camp_language" or "nazi_language". For passages in a foreign language, see <foreign>.	EHRI-BF-19380627_DE.xml EHRI-ET-JMP001_CS.xml EHRI-ET-JMP002_CS.xml EHRI-ET-JMP004_CS.xml EHRI-ET-JMP005_CS.xml EHRI-ET-JMP006_CS.xml EHRI-ET-JMP008_CS.xml EHRI-ET-JMP010_CS.xml EHRI-ET-JMP016_DE.xml EHRI-ET-JMP018_DE.xml
div	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1				@type	<input checked="" type="checkbox"/>	The <div> tag must appear at least once, just after the <body> tag. It should be added to all documents.	EHRI-BF-19380802_EN.xml
edition	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
editionStmt	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
emph	<input checked="" type="checkbox"/>	<input type="checkbox"/>	7				@rend	<input checked="" type="checkbox"/>	The <emph> tag is used to explicit the display of titles in the bibliography, but it should be used to encode a semantic emphasis. Use <chi> tag instead.	EHRI-BF-1939_DE.xml EHRI-BF-19420505_DE.xml EHRI-BF-1972_DE.xml EHRI-BF-1991_DE.xml EHRI-BF-1998_DE.xml EHRI-BF-2001b_DE.xml EHRI-BF-2001a_DE.xml
fileDesc	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
foreign	<input type="checkbox"/>	<input checked="" type="checkbox"/>		100			@xml:lang	<input checked="" type="checkbox"/>	The <foreign> tag is used to mark words or passages of the text in a different language than the main language of the document. It must be associated with a @xml:lang attribute, and its value must correspond to a code from the <a href="#">Iana Language Subtag Registry</a> .	
geo	<input checked="" type="checkbox"/>	<input type="checkbox"/>	710	1414	360	185		<input checked="" type="checkbox"/>		
head	<input type="checkbox"/>	<input checked="" type="checkbox"/>		3				<input checked="" type="checkbox"/>	Using the <head> tag is a good practice that should be implemented in all documents.	EHRI-ET-JMP008_CS.xml EHRI-ET-JMP008_EN.xml EHRI-ET-JMP008_EN_CS.xml EHRI-ET-JMP009_CS.xml EHRI-ET-JMP009_EN.xml EHRI-ET-JMP009_EN_CS.xml EHRI-ET-WL16560458_EN.xml EHRI-ET-WL16560458_EN_DE.xml
hi	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	65	84	12	6	@rend @style	<input checked="" type="checkbox"/>		
idno	<input checked="" type="checkbox"/>	<input type="checkbox"/>	188	155	117	79		<input checked="" type="checkbox"/>	To declare the file's identifier, use the @xml:id attribute in the <TEI> element.	
item	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	570	4300	1284	217	@xml:id	<input checked="" type="checkbox"/>		

FIGURE C.2 – Balises et attributs utilisés par les éditions EHRI (2/5)

Tag Name	header	body	Number of Occurrences				List of @Attributes used in EHRI editions	ODD EHRI	Comment(s)	Files Involved
			BeGrenzte Flucht (104)	Early Testimonies (119)	Diplomatic Reports (72)	Nisko (40)				
keywords	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		37		<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection. Files in which the <keywords> tag is missing are included in the "Files Involved" column.	EHRI-NISKO-19391213a_DE.xml EHRI-NISKO-19400413_DE.xml EHRI-NISKO-1940925b_DE.xml
lang	<input type="checkbox"/>	<input checked="" type="checkbox"/>		1			@xml:lang	<input type="checkbox"/>	Use the <foreign> tag.	EHRI-ET-ZIH3010481_PL.xml
langUsage	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		40		<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection.	
language	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	122		40	@usage @ident	<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection.	
lb	<input type="checkbox"/>	<input checked="" type="checkbox"/>	4	1				<input checked="" type="checkbox"/>	Using <lb> is a good practice. It must be used according to the <a href="#">TEI Guidelines</a> and indicate the beginning of a line in order to respect the physical structure of the facsimile.	EHRI-BF-19380409_DE.xml EHRI-BF-19380426_DE.xml EHRI-ET-WL05320070_YI.xml EHRI-ET-WL05320070_EN_YI.xml EHRI-ET-VV3540604_EN.xml EHRI-ET-VV3540604_EN_YI.xml
licence	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		40	@target	<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection (see <availability>).	
link	<input checked="" type="checkbox"/>	<input type="checkbox"/>	2090	7141	2095	657	@type @target	<input checked="" type="checkbox"/>	Way in which terms are encoded.	
linkGrp	<input checked="" type="checkbox"/>	<input type="checkbox"/>	1423	6097	1764	499		<input checked="" type="checkbox"/>	Way in which terms are encoded.	
list	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	107	119	87	40	@rend @type	<input checked="" type="checkbox"/>		
listOrg	<input checked="" type="checkbox"/>	<input type="checkbox"/>	78	96	50	34		<input checked="" type="checkbox"/>		
listPerson	<input checked="" type="checkbox"/>	<input type="checkbox"/>	74	113	73	35		<input checked="" type="checkbox"/>		
listPlace	<input checked="" type="checkbox"/>	<input type="checkbox"/>	102	119	72	40		<input checked="" type="checkbox"/>		
location	<input checked="" type="checkbox"/>	<input type="checkbox"/>	710	1414	360	185		<input checked="" type="checkbox"/>		
msDesc	<input checked="" type="checkbox"/>	<input type="checkbox"/>	84	119	72	39		<input checked="" type="checkbox"/>	Files with no <msDesc> file are all newspaper articles. They are indicated in the "Files Involved" column, and red files indicate that the facsimile is available on the EHRI edition website. The bold orange file seems to be missing from the EHRI BeGrenzte Flucht edition website.	EHRI-BF-19380316_DE.xml EHRI-BF-19380401b_DE.xml EHRI-BF-19380408_DE.xml EHRI-BF-19380418_DE.xml EHRI-BF-19380421a_EN.xml EHRI-BF-19380421c_DE.xml EHRI-BF-19380421d_DE.xml EHRI-BF-19380427_DE.xml EHRI-BF-19380723_DE.xml EHRI-BF-19380812b_DE.xml EHRI-BF-19380830b_DE.xml EHRI-BF-19380923_DE.xml EHRI-BF-1939_DE.xml EHRI-BF-19390202_DE.xml EHRI-BF-19420505_DE.xml EHRI-BF-1972_DE.xml EHRI-BF-1991_DE.xml EHRI-BF-1998_DE.xml EHRI-BF-2001a_DE.xml EHRI-BF-2001b_DE.xml EHRI-NISKO-194711_DE.xml
msIdentifier	<input checked="" type="checkbox"/>	<input type="checkbox"/>	84	119	72	39		<input checked="" type="checkbox"/>	See <msDesc>.	
name	<input checked="" type="checkbox"/>	<input type="checkbox"/>	548	4300	1250	217		<input checked="" type="checkbox"/>	Used to list <items> in a <list> in the <teiHeader>.	

FIGURE C.3 – Balises et attributs utilisés par les éditions EHRI (3/5)

Tag Name	header	body	Number of Occurrences				List of @Attributes used in EHRI editions	ODD EHRI	Comment(s)	Files Involved
			BeGrenzte Flucht (104)	Early Testimonies (119)	Diplomatic Reports (72)	Nisko (40)				
note	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	217	815	244	297	@place @xml:lang @type	<input checked="" type="checkbox"/>		
org	<input checked="" type="checkbox"/>	<input type="checkbox"/>	154	311	105	66	@xml:id	<input checked="" type="checkbox"/>		
orgName	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	466	1190	221	218	@rend @type @ref	<input checked="" type="checkbox"/>		
p	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	1499	4750	971	915	@rend @rendition @xml:lang	<input checked="" type="checkbox"/>	The only value for the @rendition attribute is "book" and only occurs in the <teiHeader>. The use of the @rendition attribute is therefore irrelevant.	
pb	<input type="checkbox"/>	<input checked="" type="checkbox"/>	190	617	164	31	@n @facs	<input checked="" type="checkbox"/>	Using <pb> is a good practice.	
persName	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	1024	3961	352	655	@ref @type	<input checked="" type="checkbox"/>		
person	<input checked="" type="checkbox"/>	<input type="checkbox"/>	309	1146	140	228	@xml:id	<input checked="" type="checkbox"/>		
physDesc	<input checked="" type="checkbox"/>	<input type="checkbox"/>	84	119	28	35		<input checked="" type="checkbox"/>	The <physDesc/> element is empty in all collections. It should be filled in according to the <a href="#">TEI Guidelines</a> .	
place	<input checked="" type="checkbox"/>	<input type="checkbox"/>	730	1436	378	189	@xml:id	<input checked="" type="checkbox"/>		
placeName	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	2639	6407	1061	602	@ref @type @rend	<input checked="" type="checkbox"/>		
profileDesc	<input type="checkbox"/>	<input type="checkbox"/>	104	119	72	46		<input checked="" type="checkbox"/>	Files that contain more than one occurrence of the <profileDesc> tag are indicated in the "Files Involved" column, and red files are common with the duplicates of the <creation> tag.	EHRI-NISKO-19391015_DE.xml EHRI-NISKO-19391017_DE.xml EHRI-NISKO-19391018_DE.xml EHRI-NISKO-19391212_DE.xml EHRI-NISKO-19400420_DE.xml EHRI-NISKO-194711_DE.xml
publicationStm	<input type="checkbox"/>	<input type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
publisher	<input type="checkbox"/>	<input type="checkbox"/>	104	119	72	80		<input checked="" type="checkbox"/>		
q	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	101	771	29			<input checked="" type="checkbox"/>		
ref	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	106	119	97	80	@target	<input checked="" type="checkbox"/>		
relatedItem	<input checked="" type="checkbox"/>	<input type="checkbox"/>			2		@target	<input type="checkbox"/>	Links in the @target attribute are broken. Including the <relatedItem> tag in the ODD is not relevant.	EHRI-DR-19421120-US.xml EHRI-DR-19430325-US_EN.xml
repository	<input checked="" type="checkbox"/>	<input type="checkbox"/>	84	113	72	39	@ref	<input checked="" type="checkbox"/>	The <repository> tag is missing when the <msDesc> tag is missing.	
row	<input type="checkbox"/>	<input checked="" type="checkbox"/>	9	24			@role	<input checked="" type="checkbox"/>	See <table>.	EHRI-BF-19381231_DE.xml EHRI-ET-WL16560413_DE.xml EHRI-ET-WL16560413_EN_DE.xml EHRI-ET-YV3540315_YI.xml EHRI-ET-YV3540315_EN.xml EHRI-ET-YV3540315_EN_YI.xml
seg	<input type="checkbox"/>	<input checked="" type="checkbox"/>				8	@rend	<input type="checkbox"/>	The arbitrary segment <seg> tag is irrelevant in the context of EHRI editions. (see "Files Involved") This file should be modified to delete the <seg> tag.	EHRI-NISKO-19391212_DE.xml
sic	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1					<input type="checkbox"/>	Irrelevant. Will not be included in the ODD.	EHRI-BF-19380509c_DE.xml

FIGURE C.4 – Balises et attributs utilisés par les éditions EHRI (4/5)

Tag Name	header	body	Number of Occurrences				List of @Attributes used in EHRI editions	ODD EHRI	Comment(s)	Files Involved
			BeGrenzte Flucht (104)	Early Testimonies (119)	Diplomatic Reports (72)	Nisko (40)				
signatures	<input type="checkbox"/>	<input checked="" type="checkbox"/>	2					<input type="checkbox"/>	Use <signed> tag instead.	EHRI-BF-19380509b_EN.xml EHRI-BF-19380509c_DE.xml
signed	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1					<input checked="" type="checkbox"/>	Should be embedded in a <closer> tag.	EHRI-BF-19380802_EN.xml
sourceDesc	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		
space	<input type="checkbox"/>	<input checked="" type="checkbox"/>		1	4		@rend	<input checked="" type="checkbox"/>	The <space> tag is used to describe primary source by indicating a significant space in the text. It should be used to describe an unusual space and have @quantity and @unit attributes whenever possible to best describe the space. In the "Files Involved", the <space> tag is either empty or has an @rend="tab" attribute and does not justify the use of such a tag. The <space> tag will be referenced in the ODD but should be used correctly, according to the TEI Guidelines.	EHRI-ET-JMP004_EN.xml EHRI-ET-JMP004_EN_CS.xml EHRI-ET-YV5245989_DE.xml EHRI-ET-YV5245989_EN_DE.xml EHRI-DR-19411007-IT_EN.xml
sponsor	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104			40		<input checked="" type="checkbox"/>		
stamp	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1					<input checked="" type="checkbox"/>	Should be used for every stamp in all collection.	EHRI-BF-19380330a_DE.xml
table	<input type="checkbox"/>	<input checked="" type="checkbox"/>	1	3			@cols @rows	<input checked="" type="checkbox"/>	Give an example of the complete table structure in the ODD.	EHRI-BF-19381231_DE.xml EHRI-ET-YV3540315_YI.xml EHRI-ET-YV3540315_EN.xml EHRI-ET-YV3540315_EN_YI.xml EHRI-ET-WL16560413_DE.xml EHRI-ET-WL16560413_EN_DE.xml
term	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	1333	16508	2186	656	@ref @rend @type	<input checked="" type="checkbox"/>		
textClass	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		37		<input checked="" type="checkbox"/>	Must be included in the Diplomatic Reports collection. See "Files Involved" for files without <textClass> tag.	EHRI-NISKO-19391213a_DE.xml EHRI-NISKO-19400413_DE.xml EHRI-NISKO-19470925b_DE.xml
textLang	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	69	35	@mainLang	<input checked="" type="checkbox"/>	Used in the <bibli> element. The @mainLang attribute is required and the @otherLangs attribute is recommended. All attributes' values must correspond to a code from the <a href="#">iana Language Subtag Registry</a> . See "Files Involved" for files in which the <textLang> tag must be added.	EHRI-DR-19430327_SK.xml EHRI-DR-19430601-US_EN.xml EHRI-DR-19430745_US_EN.xml EHRI-NISKO-193910a_DE.xml EHRI-NISKO-193910a_EN_DE.xml EHRI-NISKO-1946_DE.xml EHRI-NISKO-1948_DE.xml EHRI-NISKO-1970_DE.xml
title	<input checked="" type="checkbox"/>	<input type="checkbox"/>	142	119	72	42	@type @xml:lang	<input checked="" type="checkbox"/>		
titleStm	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119	72	40		<input checked="" type="checkbox"/>		

The values for all @attributes MUST be written in ENGLISH.

Values for the @xml:lang attribute should be normalized with [iana Language Subtag Registry](#) codes.

FIGURE C.5 – Balises et attributs utilisés par les éditions EHRI (5/5)



# Annexe D

## Encodage des métadonnées

```
1 <teiHeader>
2   <fileDesc>
3     <titleStmt>
4       <title xml:lang="en"/>
5       <title xml:lang=""/>
6     <principal>
7       <affiliation>
8         <orgName ref="https://www.ehri-project.eu">European
          ↪ Holocaust Research Infrastructure</orgName>
9       </affiliation>
10    </principal>
11    <respStmt>
12      <resp/>
13      <persName/>
14    </respStmt>
15  </titleStmt>
16  <publicationStmt>
17    <authority>
18      <ref target="https://www.ehri-project.eu">European
        ↪ Holocaust Research Infrastructure</ref>
19    </authority>
20    <availability>
21      <licence
        ↪ target="http://creativecommons.org/licenses/by-sa/4.0">
22        Attribution-ShareAlike 4.0 International
23      </licence>
24    </availability>
```

```
25     </publicationStmt>
26     <seriesStmt>
27         <title ref="{link to the online edition}"/>
28     </seriesStmt>
29     <sourceDesc>
30         <msDesc>
31             <msIdentifier>
32                 <institution>
33                     <orgName/>
34                     <address>
35                         <street>
36                             <num/>
37                         </street>
38                         <postCode/>
39                         <settlement/>
40                         <country/>
41                     </address>
42                 </institution>
43                 <collection/>
44                 <idno/>
45             </msIdentifier>
46             <physDesc>
47                 <p/>
48             </physDesc>
49         </msDesc>
50         <bibl>
51             <textLang/>
52         </bibl>
53     </sourceDesc>
54 </fileDesc>
55 <encodingDesc>
56     <projectDesc>
57         <p xml:lang="en"/>
58     </projectDesc>
59 </encodingDesc>
60 <profileDesc>
61     <creation>
62         <origDate when=""/>
63         <origPlace ref="{GeoNames link}"/>
```



```
64         <persName ref="{EHRI entity}"/>
65     </creation>
66     <textClass>
67         <catRef target="{link to EHRI portal}"/>
68         <keywords>
69             <term/>
70         </keywords>
71     </textClass>
72     <langUsage>
73         <language ident=""/>
74     </langUsage>
75     <abstract>
76         <p xml:lang="en"/>
77     </abstract>
78 </profileDesc>
79 <revisionDesc>
80     <change when="" who="{ }"/>
81 </revisionDesc>
82 </teiHeader>
```



# Annexe E

## Extrait du script de recherche

```
1      import os # Allows interaction with the computer, regardless of its
      ↪ operating system.
2      import sys # Allows for the argument to be called directly from the
      ↪ command line.
3      from bs4 import BeautifulSoup
4
5      def parse(corpus):
6          """ Parses the given corpus.
7              Parameter
8              -----
9              corpus
10                 Absolute path for the corpus.
11                 Corresponds to "arg1" in the command line.
12             """
13         for root, dirs, files in os.walk(corpus):
14             for filename in files:
15                 with open(os.path.join(root, filename), 'r',
16                     ↪ encoding='UTF-8') as xml_file:
17                     # Opens the files in 'read' mode and stores the opened
18                     ↪ file in a variable called 'xml_file'.
19                     xml_content = xml_file.read()
20                     # The 'xml_content' variable stores the content of
21                     ↪ the file opened in 'xml_file' with the '.read()'
22                     ↪ method.
23                     soup = BeautifulSoup(xml_content, 'xml')
```

```

20         # Creates a 'soup' object, which parses 'xml_content'
        ↪ with the XML parser included in the BeautifulSoup
        ↪ library.
21     yield filename, soup
22     # The 'yield' statement suspends the execution of the
        ↪ function to return a result, and then resumes
        ↪ execution instead of starting over every time.
23
24
25     def missing_tag(tag):
26         """ Finds the files from which the element is missing.
27         Parameter
28         -----
29         tag : str
30             Name of the researched element, between quotation marks.
31         Returns
32         -----
33         set()
34             Set containing the names of the files from which the given
        ↪ element is missing.
35         """
36         files_missing_tag = set()
37         for filename, soup in parse(sys.argv[1]):
38             # Uses the 'parse' function to parse the corpus provided by
            ↪ 'arg1' in the command line.
39             if not soup.find_all(tag):
40                 files_missing_tag.add(filename)
41             # Adds the name of the file to the set if the
            ↪ '.find_all()' method does not find any occurrence.
42             print(filename)
43     # Example --> missing_tag("catRef")

```

# Table des figures

1.1	Chaîne éditoriale de l'application DiScholEd (source : F. Chiffolleau) . . . .	7
2.1	Frontières du Reich en mars 1938 (Source : Wikipédia) . . . . .	11
4.1	Fonctionnement du tableau récapitulatif de l'encodage des éditions EHRI .	21
6.1	<i>Template</i> d'une entrée dans l'index de lieux . . . . .	30
7.1	Disposition horizontale avec TEI Publisher (Source : DiScholEd) . . . . .	37
7.2	Diposition verticale avec Omeka (Source : EHRI Online Editions) . . . . .	38
8.1	Page d'accueil de la collection sur DiScholEd . . . . .	41
8.2	Collection « <i>Early Holocaust Testimony</i> » sur DiScholEd . . . . .	41
9.1	Maquette de la page d'accueil EHRI avec TEI Publisher . . . . .	44
9.2	Maquette de navigation dans la collection . . . . .	45
9.3	Maquette de l'affichage du document avec TEI Publisher . . . . .	45
9.4	Exemple de l'affichage d'un document sur l'application TEI Publisher EHRI	45
C.1	Balises et attributs utilisés par les éditions EHRI (1/5) . . . . .	59
C.2	Balises et attributs utilisés par les éditions EHRI (2/5) . . . . .	60
C.3	Balises et attributs utilisés par les éditions EHRI (3/5) . . . . .	60
C.4	Balises et attributs utilisés par les éditions EHRI (4/5) . . . . .	61
C.5	Balises et attributs utilisés par les éditions EHRI (5/5) . . . . .	61



# Table des matières

Résumé	i
Remerciements	iii
Liste des sigles et abréviations	v
Sommaire	v
Bibliographie	ix
Introduction	xiii
<b>I Édition scientifique et numérique</b>	<b>1</b>
<b>1 Le processus d'édition</b>	<b>3</b>
1.1 Édition et édition scientifique . . . . .	3
1.1.1 Définition de la notion d'édition . . . . .	3
1.1.2 Caractéristiques de l'édition scientifique . . . . .	3
1.2 Édition et édition numérique . . . . .	4
1.2.1 Caractéristiques d'une édition numérique . . . . .	4
1.2.2 Limites de l'édition imprimée . . . . .	5
1.3 La chaîne éditoriale . . . . .	6
1.3.1 Définition de la chaîne éditoriale . . . . .	6
1.3.2 Un exemple : la <i>pipeline</i> DiScholEd . . . . .	7
<b>2 Les éditions du projet EHRI</b>	<b>9</b>
2.1 Une nouvelle approche des archives numériques . . . . .	9
2.1.1 Une initiative européenne . . . . .	9
2.1.2 Développer les éditions numériques sur la Shoah . . . . .	9
2.2 Les éditions en ligne EHRI . . . . .	10
2.2.1 BeGrenzte Flucht (2018) . . . . .	10

2.2.2	Early Holocaust Testimony (2020) . . . . .	11
2.2.3	Diplomatic Reports (2021) . . . . .	12
2.2.4	Nisko (2023) . . . . .	12
<b>3</b>	<b>La diplomatique à l'épreuve du numérique</b>	<b>15</b>
3.1	La diplomatique . . . . .	15
3.1.1	En tant que science . . . . .	15
3.1.2	L'analyse diplomatique . . . . .	15
3.1.3	La diplomatique contemporaine . . . . .	16
3.2	Typologie documentaire des éditions EHRI . . . . .	16
3.3	Le standard TEI pour l'encodage des documents . . . . .	17
3.3.1	Un standard ouvert et interopérable . . . . .	17
3.3.2	Diplomatique et TEI . . . . .	17
<b>II</b>	<b>Modèle d'encodage</b>	<b>19</b>
<b>4</b>	<b>Encodage des éditions EHRI</b>	<b>21</b>
4.1	Analyse des pratiques d'encodage . . . . .	21
4.1.1	Extraction de l'information . . . . .	21
4.1.2	Évaluation du choix des balises et attributs . . . . .	22
4.2	Identification des améliorations possibles . . . . .	22
4.2.1	Amélioration de la conformité . . . . .	22
4.2.2	Uniformisation des édition . . . . .	23
4.2.2.1	Utilisation de l'anglais . . . . .	23
4.2.2.2	Standardisation de la valeur de certains attributs . . . . .	23
<b>5</b>	<b>Définition du nouveau modèle d'encodage</b>	<b>25</b>
5.1	<i>One Document Does-it-all</i> . . . . .	25
5.2	Rédaction de la documentation . . . . .	26
5.3	Création du schéma de validation . . . . .	27
5.3.1	Déclaration des modules . . . . .	27
5.3.2	Règles de validation des éléments . . . . .	27
5.3.3	Déclaration des classes d'attributs . . . . .	27
<b>6</b>	<b>Semi-automatisation du modèle</b>	<b>29</b>
6.1	Automatisation et semi-automatisation . . . . .	29
6.2	Création des <i>templates</i> d'encodage . . . . .	29



<b>III</b>	<b>Publication et valorisation</b>	<b>31</b>
<b>7</b>	<b>Choix de l'outil de publication</b>	<b>33</b>
7.1	Méthodes de publication . . . . .	33
7.1.1	Transformation XSLT vers HTML . . . . .	33
7.1.2	Création d'une application web . . . . .	34
7.1.3	Utilisation d'une plateforme de publication . . . . .	34
7.2	Les éditions EHRI et le logiciel Omeka . . . . .	34
7.2.1	Fonctionnement d'Omeka Classic . . . . .	35
7.2.2	Limites posées par le recours à Omeka . . . . .	35
7.3	Une autre alternative : TEI Publisher . . . . .	36
7.3.1	Fonctionnement de TEI Publisher . . . . .	36
7.3.2	Pourquoi choisir TEI Publisher . . . . .	37
<b>8</b>	<b>EHRI sur l'application DiScholEd</b>	<b>39</b>
8.1	Une application pour publier les ego-documents . . . . .	39
8.1.1	Qu'est-ce qu'un ego-document ? . . . . .	39
8.1.2	Les collections publiées sur DiScholEd . . . . .	39
8.2	La collection « <i>Early Holocaust Testimony</i> » sur DiScholEd . . . . .	40
<b>9</b>	<b>Transition vers une application dédiée</b>	<b>43</b>
9.1	Analyse des besoins . . . . .	43
9.1.1	Profil des utilisateur·ice·s . . . . .	43
9.1.2	Transposer les fonctionnalités d'Omeka . . . . .	43
9.2	Création du prototype d'application . . . . .	44
9.2.1	Des fonctionnalités supplémentaires . . . . .	44
9.2.2	Maquettes visuelles . . . . .	44
	<b>Conclusion</b>	<b>47</b>
<b>A</b>	<b>Glossaire</b>	<b>49</b>
<b>B</b>	<b>Chronologie de la persécution des Juifs</b>	<b>53</b>
<b>C</b>	<b>Tableau récapitulatif d'encodage</b>	<b>59</b>
<b>D</b>	<b>Encodage des métadonnées</b>	<b>63</b>
<b>E</b>	<b>Extrait du script de recherche</b>	<b>67</b>