

EHRI Online Editions Meeting

July 4th, 2023

Sarah Beniere

Inria Paris • ALMAAnaCH

Objectives

- For the internship: Setting up a semi-automated process to help with the encoding and publishing of the future EHRI Online editions
- Methods:
 - Creating user-friendly tools for the editions
 - Providing advice for homogenization technique
 - Suggesting elements for enhancement and TEI conformance
 - Documenting the whole process

What has been done so far

Presenting the EHRI ODD

Quick reminder: what is an ODD?

- *Stands for* ▶ One Document Does it all
- *Contains* ▶ Specifications (TEI-XML) & documentation (prose)
- *Generates* ▶ RelaxNG schema applicable to a corpus to ensure TEI conformance

Observations on the current encoding

- The [overview table](#) synthesizes the TEI encoding of the EHRI files for all online editions currently available.
- The encoding is not homogeneous throughout the editions, especially in the `<teiHeader>`.
- By running a [Python script](#), it is possible to identify the files that include or are missing a certain element.

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
			deGrenzte 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 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-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"/>	Guidelines : <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/destiny), "Flüchtlingspolitik" (Refugee Policy), "IKG Wien" (Israelitische 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

Observations on the current encoding – EXAMPLE

Tag Name	header	body	Number of Occurrences			
			BeGrenzte Flucht (104)	Early Testimonies (119)	Diplomatic Reports (72)	Nisko (40)
availability	<input checked="" type="checkbox"/>	<input type="checkbox"/>	104	119		40

- All files should have an `<availability>` element to explicitly state the sharing and reusing conditions, so this should be added to the entire “Diplomatic Reports” collection.

Improving TEI conformance with the ODD

- General recommendations:
 - Managing the file's identifier with `@xml:id` attributes.
 - Making sure the files have a standard structure.
 - Managing date attributes (`@when` and `@when-iso`) and their format.
 - Establishing English as the encoding language for attributes' values and metadata.
- Specific adjustments for EHRI:
 - Making some attributes mandatory.
 - Creating lists of semi-closed values for several attributes.

General recommendations

- Observation: The identifier appears several times throughout the `<teiHeader>`.
- Solution: The identifier only appears once, as an `@xml:id` attribute in the `<TEI>` element.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="EHRI-NISKO-19391112-DE">
```

General recommendations

- Observation: There is no `<div>` element in the `<body>`.
- Solution: The `<div>` element becomes mandatory, and requires a `@type` attribute with a value of either `"transcription"` or `"translation"` (list can be expanded).

```
<text>
  <body>
    <div type="transcription">
      <pb n="1" facs="EHRI-BF-1946_01.jpg"/>
      <p>In Sommer <lb/>
        [...]</p>
      <p>[...]</p>
    </div>
  </body>
</text>
```

- Best practice: marking the page beginnings (`<pb />`) and line breaks (`<lb />`).

General recommendations

- Observation: Only the @when attribute is used with the <date> tag.
- Solution: Using @when-iso whenever the date is of YYYY-MM-DD format, and @when in every other case.
- Exception: Some tags constrain the use of the @when attribute, which will be specified in the ODD.

```
<date when-iso="1939-01-12">January 12 1939</date>  
<date when="1942-09">September 1942</date>
```

General recommendations

- Observation: English is predominant in the encoding.
- Solution: English should be the main language of EHRI files, and the metadata should appear **at least** in English. The elements appearing in original language must be translated into English and signaled with an `@xml:lang="en"` attribute.

```
<keywords>  
  <term>staatliche Dokumente</term>  
</keywords>
```

Current encoding

```
<keywords>  
  <term>government document</term>  
</keywords>
```

Best practice

General recommendations

- Observation: The values for the attributes are very often in German.
- Solution: For a better understanding of the encoding by all users, the values for all attributes must be in English only.

```
<catRef target="Vertreibungspolitik"/>  
<catRef target="expulsion_policy"/>
```

Current encoding

Best practice

General recommendations

- The values for the `@xml:lang` attribute should match the values recommended by the [iana Language Subtag Registry](#):

- Czech = "cs"
- Dutch = "nl"
- English = "en"
- French = "fr"
- German = "de"
- Hebrew = "he"
- Hungarian = "hu"

- Italian = "it"
- Polish = "pl"
- Russian = "ru"
- Slovak = "sk"
- Ukrainian = "uk"
- Yiddish = "yi"

Adjustments for EHRI

- In some cases, specific attributes were made mandatory:
 - `@type` for the `<distinct>` element
 - `@xml:lang` for the `<foreign>` element
- A list of suggested values was established for some attributes:
 - `<distinct>` `@type` → “camp_language” or “nazi_language”
 - `` `@rend` → “**strikethrough**” or “overwritten” etc.

Future perspectives

Homogenization, easier encoding & TEI Publisher

Homogenizing the editions

- The new ODD for EHRI electronic editions comes with a [validation schema](#).
- The schema should be applied to new files to ensure improved TEI conformance.
- It should also be applied to the previous editions to facilitate their homogenization.

Making the encoding easier

- Observation: The encoding of texts can be a long and tedious task.
- Solution: Creating a semi-automated process of encoding makes this step easier for the editors, and the encoding can be done faster.
- Methods:
 - Creating ready-to-use templates for an encoding by hand.
 - Creating Python scripts for a semi-automated encoding.
- Both methods will be thoroughly documented for them to be within anyone's reach.

TEI Publisher application for EHRI

- Observation: The “Early Holocaust Testimony” collection has been published on the [DiSchoEd](#) application. TEI Publisher is fully adapted to display documents encoded in the TEI standard.
- Suggestion: Creating a dedicated TEI Publisher application for the EHRI collections.
- Methods: The basis for the application architecture would be the same as DiSchoEd, which means it could be ready faster.
- Advantage of TEI Publisher: The platform is [well-documented](#), regularly maintained, and open source. The TEI Publisher community is large, very active and easily reachable.

EHRI Online Editions

The **EHRI Online Editions** consist of annotated digitized documents from various sources gathered around a theme, and are a new way of presenting digital archival content. A new edition is published at irregular intervals.

Available editions:

- [Von Wien ins Nirgendwo: Die Nisko-Deportationen 1939](#) (in German) (2023)
- [Diplomatic Reports](#) (2021)
- [Early Holocaust Testimony](#) (2020)
- [BeGrenzte Flucht](#) (in German) (2018)

[Von Wien ins Nirgendwo: Die Nisko-Deportationen 1939](#) (in German)



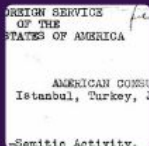
General homepage for all the editions

EHRI Digital Editions



Bordered Escape

This edition gathers documents testifying of the increasingly restrictive refugee policy in Czechoslovakia, following the "Anschluss."



Diplomatic Reports

This edition gathers reports on the persecution and murder of European Jews.



Early Holocaust Testimony

This edition gathers accounts of the persecution of Jews from the Nazi takeover of power in Germany (1933) to the Eichmann Trial (1961).



The Nisko Deportations

This edition gathers documents on the deportation of thousands of Jews to Nisko am San (Poland).

VON WIEN INS NIRGENDWO: DIE NISKO-DEPORTATIONEN 1939

Search



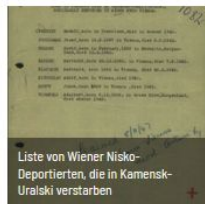
Nach dem Überfall auf Polen im September 1939 eröffnete sich für die NS-Führung erstmals die Möglichkeit, Jüdinnen und Juden in eroberte Gebiete abzuschieben. In der zweiten Oktoberhälfte 1939 verschickte die Zentralstelle für jüdische Auswanderung 1.500 jüdische Männer aus Wien nach Nisko am San. Weitere Transporte gingen aus Mährisch-Ostrau (Ostrava) und Kattowitz (Katowice) nach Nisko. Nur ein kleiner Teil der 4.750 bis 4.900 Deportierten kam jedoch in dem Lager unter, die Übrigen wurden von der SS verjagt. Hunderte irrten im Distrikt Lublin umher. Die Mehrheit gelangte in die Sowjetunion, wo sich ihre Spuren verlieren. 198 Männer konnten nach Abbruch der Aktion im März 1940 nach Wien zurückkehren.

Diese Quellenedition führt kommentierte Dokumente aus verschiedenen Ländern zusammen. Sie soll auch in den kommenden Jahren kontinuierlich erweitert werden und so die wissenschaftliche Aufmerksamkeit auf die unzureichend genutzten Zeugnisse der Opfer und auf das Schicksal der Deportierten in diesem Grenzgebiet lenken.

ZULETZT HINZUGEFÜGTE DOKUMENTE



Zimmerkommandantinnen an Josef Löwenherz über die Zustände im Sammellager Gänsbachergasse



Liste von Wiener Nisko-Deportierten, die in Kamensk-Uralski verstarben



IKG Wien Liste an den World Jewish Congress betreffend Wiener Nisko-Deportierte in Kamensk-Uralski



Nisko-Deportierte aus Zamosc an die IKG Wien mit der Bitte um Unterstützung

Homepage for the Nisko collection

EHRI Digital Editions

Sort by



Filter by



Name of the collection



English title of the document
Abstract

English title of the document
Abstract

English title of the document
Abstract

English title of the document
Abstract

Filters

Language

- ☐ Czech
- ☐ German
- ☐ Yiddish

Date

- ☐ 1939
- ☐ 1940
- ☐ 1944

Place

- ☐ Auschwitz
- ☐ Budapest
- ☐ Vienna

Holding institution

- ☐ Hungarian Jewish Archives
- ☐ Jewish Historical Institute in Warsaw
- ☐ Yad Vashem Archives

Current display of the document

[← Ray Aderton on the reaction of the...](#)[Alexander Kolb on the deportation of...](#)

John F. Montgomery on the “Hungarian Jew Bill” in 1939

[METADATA](#)[DOCUMENT TEXT](#)[MAP](#)[REFERENCES](#)

METADATA

649-09-102070PLUS

1939-09-19 | Budapest

National Archives and Records Administration 6, RG 59 - Records of the Department of State relating to the internal affairs of Hungary 1919-1964, Wilmington, Del. 1967 (Montreal), Doc. 884.0218-161. 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 its 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 3 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 available](#)

[View this item on the D4U Portal](#)

National Archives and Records Administration

NAUSA | National Archives

United States 649-09-102070PLUS | College Park | Updated 7 months ago

[View this item on the D4U Portal](#)

View

Original language or translation(s) ▼

Transcription

Index entry

Facsimile

Interactive map

Thank you for your attention!

Any questions?