

Historical Animal Observation Records by Bavarian Forestry Offices (1845): Documentation of the XML Structure

Version 1.0 as of 2024-10-12

Malte Rehbein, Bettina Haas, Giada Matheisen, Tobias Perschl

Files:

- AOD_transcriptions.xml
- AOD_transcriptions.xsd

This documentation provides a detailed overview of the XML structure used in the file AOD_transcriptions.xml to represent transcriptions from the archival records "Historical animal observation records by Bavarian forestry offices (1845)". The XML captures metadata and content related to the fauna survey conducted in all forestry districts of the Kingdom of Bavaria in 1845. The original (archival) provenance of the data is documented in the XML-file.

For validation purposes, an XML Schema Definition is available in the file AOD_transcriptions.xsd.

Overview

The XML file is structured hierarchically to reflect the organization of the archival documents. It consists of a root `<collection>` element that contains metadata about the collection and multiple `<file>` elements representing the individual archival files.

```
<collection>
  <fileDesc>
    <!-- Metadata about the collection -->
  </fileDesc>
  <file ID="210" UUID="...">
    <!-- First archival file -->
  </file>
  <file ID="211" UUID="...">
    <!-- Second archival file -->
  </file>
  <!-- Additional <file> elements -->
</collection>
```

Elements and Attributes

1. `<collection>`

- Description: The root element encapsulating the entire collection.
- Children:
 - `<fileDesc>`
 - One or more `<file>` elements.

2. `<fileDesc>`

- Description: Contains metadata about the entire collection. These follow the recommendation by the Text Encoding Initiative (TEI).
- Children:
 - `titleStmt`
 - `editionStmt`
 - `publicationStmt`

3. `<titleStmt>`

- Description: Holds the title of the collection.
- Child:
 - `<title>`: The title as text

4. `<editionStmt>`

- Description: Provides details about the genesis of the online data set.
- Children:
 - `<edition>`: Indicates the type of publication, data set in this case
 - `<editor>`: Name of the editor responsible for the online data set
 - `<date>`: Year of publication
 - `<respStmt>`: Statements of responsibility and contribution.
 - `<resp>`: Role (e.g., transcribed by)
 - `<name>`: Names of the people holding this role.

5. `<publicationStmt>`

- Description: Information about the publishers, availability, and usage licences of the online data set.

6. `<file>`

- Description: Represents an individual (original) archival file within the collection.
- Children:
 - `archivaldescription`: Contains descriptive metadata about the archival file.
 - `content`: Content included in the file; mostly the responses from the governmental district in question
- Attributes:
 - `@ID`: Unique identifier for the file that corresponds with the signature of the Bavarian State Archive (e.g., "210").
 - `@UUID`: Universal Unique Identifier linking to the representation of the source file in the online archive.

7. `<archivaldescription>`

- Description: Contains descriptive metadata about the archival file.
- Children:
 - `<title>`: Title of the archival file according to the archival catalog.
 - `<contains>`: Description of the contents as provided by the State Archive.
 - `<signature>`: Archival signature or reference as provided by the State Archive.
 - `<provenance>`: Origin of the file before transfer to the State Archive.
 - `<dateOfOrigin>`: Date (year) when the original document was created.

7. <content>

- **Description:** Contains the main transcribed content.
- **Children:**
 - `<opener>`: Contains any opening notes.
 - `<body>`: Contains the main content of the report.
 - `<closer>`: Contains any closing notes.
- **Attribute:**
 - `@districtName`: Name of the governmental district in German language (e.g., "Niederbayern").

9. <opener>

- **Description:** When given, represents any notes on the front cover of the file before the reports, likely by the collecting institution.

10. <body>

- **Description:** Main body containing the reports by the forestry offices, in most cases four pages. See `./entry/text@page` for page numbers.
- **Children:**
 - `<report>`: Main (tabular) content.

11. <report>

- **Description:** This is the main content; represents one report from a specific office.
- **Children:**
 - `<opener>`: When given, represents any opening notes on the reports by the responding forester, usually at least the name of the office. These may also contain handwritten notes by the collecting institution or Andreas Wagner. Usually give at the top of the first page of the survey form.
 - `<entry>`: Content of one table row in the questionnaire. Usually each report contains 44 entries; the 45th row (about fish) is omitted here.
- **Attributes:**
 - `@officeID`: Unique identifier for the office (e.g., "OF_0079"); this refers to the identifier in the main data table.
 - `@officeName`: Name of the office (e.g., "Altdorf").

Example:

```
<report officeID="OF_0079" officeName="Altdorf">
  <!-- Reported content -->
</report>
```

12. <entry>

- **Description:** Describes the content of one table row in the questionnaire, representing the report about one animal species.
- **Child:**
 - `<text>`
- **Attributes:**
 - `@id`: Unique identifier for the entry. This corresponds with the main data table.
 - `@speciesID`: Identifier for the species (e.g., "SP_0001"). This corresponds with the main data table.
 - `@page`: Page number in the original file.

Example:

```
<entry xml:id="E_03355" speciesID="SP_0001" page="3">
  <text>Kömmt in dem hiesigen Amtsbezirk nicht vor.</text>
</entry>
```

13. <text>

- **Description:** Represents the entry's content as raw text. May contain unclear transcriptions, marked with '#'. The text is described 'as is', i.e. spelling errors in the original are preserved and abbreviations are not expanded.

14. <closer> (within <report>)

- **Description:** Closing remarks, signatures, or dates when given by the forester. These are usually placed at the bottom of the last page of the survey form.

15. <closer> (within <file>)

- **Description:** When given, represents any notes on the back cover of the file after the reports, likely by the collecting institution

16. Other Elements

- `<additionaltext>`: Used for additional notes outside the main structure but related to a specific entry.

Detailed Explanation with Examples

Metadata Section

The metadata provides essential information about the collection and its publication.

```
<fileDesc>
  <titleStmt>
    <title>Historical animal observation records by Bavarian forestry offices (1845)</title>
  </titleStmt>
  <editionStmt>
    <edition>Data set publication</edition>
    <editor>Malte Rehbein</editor>
    <date>2024</date>
    <respStmt>
      <resp>transcribed by</resp>
      <name>Bettina Haas, Felix Lambsdorff, Giada Matheisen, Tobias Perschl, Malte Rehbein</name>
    </respStmt>
  </editionStmt>
  <publicationStmt>
    <publisher>Chair of Computational Humanities, University of Passau, Germany</publisher>
    <publisher>Generaldirektion der Staatlichen Archive Bayerns, Munich, Germany</publisher>
    <publisher>German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany</publisher>
    <availability status="free">
      <licence target="https://creativecommons.org/licenses/by/4.0/">
        <p>This work is licensed under a Creative Commons Attribution 4.0 International License.</p>
      </licence>
    </availability>
  </publicationStmt>
</fileDesc>
```

```
</publicationStmt>
</fileDesc>
```

Purpose: Provides context and credit to those involved in the creation and transcription of the dataset.

Archival File Example

Each `<file>` element represents an archival file, including its metadata and content.

```
<file ID="210" UUID="https://www.gda.bayern.de/show/09d0febb-5bcd-4e6e-baf1-8718ce13a725/1">
  <archivaldescription>
    <title>Erfassung der Fauna im Kreis Niederbayern</title>
    <contains>Ausgefüllte Erfassungsbögen der Forstämter Deggendorf, Freyung, Kelheim, Landau a.d. Isar,
      Neustadt a.d. Donau, Passau, Schönberg, Thurn-und-Taxis, Wolfstein und Zwiesel; Amtsbericht des
      Forstamts Landau a.d. Isar (1846)
    </contains>
    <signature>BayHStA, Zoologische Staatssammlung 210</signature>
    <provenance>Zoologische Staatssammlung</provenance>
    <dateOfOrigin>1845 (1846)</dateOfOrigin>
  </archivaldescription>
  <content districtName="Niederbayern">
    <!-- Content specific to Lower Bavaria district -->
  </content>
</file>
```

Report and Entries

Within the `<content>` element, the reports from all offices based on the questionnaires are included.

Report Example

```
<report officeID="OF_0002" officeName="Freysing (niederbayerischer Anteil)">
  <opener>
    Forstamt Freysing (:niederbayerischer Anteil)
  </opener>
  <body>
    <!-- Entries -->
  </body>
  <closer>
    koenigl. Forstamt Freysing // [signature]
  </closer>
</report>
```

Entry Example

Each `<entry>` corresponds to observations about a species.

```
<entry xml:id="E_03355" speciesID="SP_0001" page="3">
  <text>Kömmt in dem hiesigen Amtsbezirk nicht vor.</text>
</entry>
```

This documentation may assist users in understanding the structure and content of the XML file, enabling effective utilization of the dataset in scientific studies.

References

The XML data is part of a project by the Chair of Computational Humanities, University of Passau, in cooperation with the Generaldirektion der Staatlichen Archive Bayerns, Munich, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, and other partners from the German consortium NFDI4Biodiversity.

The original archival materials are held by the Bavarian State Archive, Bayerisches Hauptstaatsarchiv (BayHStA).