

## Dokumentation des TEI Schemas:

### Einleitung

Die Bildannotationen im Datathon wurden nach den [P5 Guidelines der Text Encoding Initiative](#) erstellt und basieren im Wesentlichen auf dem [Modul Representations of Primary Sources](#).

Die im Rahmen des Datathons entstandenen TEI-Dokumente seien im Folgenden hinsichtlich ihrer Struktur und Elemente beschrieben. Es wird unterschieden zwischen Elementen und Attributen, die beim RAW-XML-Export aus der [Annotationsumgebung](#) mithilfe eines XSL-Transformationskript **automatisch** erstellt wurden – diese sind **grau unterlegt** – und solchen Elementen und Attributen, die im Annotationstool unter Nutzung eines vereinfachten Tagsets **manuell** angelegt wurden und mithilfe des XSLT in TEI transformiert wurden – diese sind **blau unterlegt**.

### TEI-Header

Der TEI-Header **<teiHeader>** enthält das Element **<fileDesc>**. Dieses enthält die Elemente **<titleStmt>**, **<publicationStmt>** und **<sourceDesc>**.

**<titleStmt>** enthält das Element **<title>**, darin werden die „Bildannotationen in mittelalterlichen Handschriften der SBB-PK“ genannt.

**<publicationStmt>** enthält das Element **<authority>** mit Angabe der Veranstaltung: „Erstellt im Rahmen des Datathons der SBB-PK und der FU Berlin“, das Element **<date>** mit Angabe des Entstehungsjahres sowie das Element **<availability>** mit dem Kind-Element **<licence>** und der Angabe der Rechteinformation: „veröffentlicht unter“  
`<ref target="https://creativecommons.org/publicdomain/zero/1.0/" />` mit der Angabe der Lizenz „CCO 1.0 Universal Licence“.

Das Element **<sourceDesc>** enthält das Element **<msDesc>** mit dem Kind-Element **<msIdentifier>**. Dieses wiederum hat die Kind-Elemente **<settlement>** mit Angabe des Aufbewahrungsortes „Berlin“, **<repository>** mit Angabe der bestandshaltenden Institution „Staatsbibliothek zu Berlin – Preußischer Kulturbesitz“ und **<idno>** mit Angabe der Signatur der Handschrift.

### TEI-SourceDoc

Das Element **<sourceDoc>** enthält das Attribut `@source` mit dem IIF-Manifest der Handschrift. Innerhalb des `<sourceDoc>` gibt es ein Kind-Element **<surface>** für jede annotierte Seite der betreffenden Handschriften. Dieses Kind-Element **<surface>** hat mehrere Attribute: `@facs`: Image-URL der Seite des digitalen Objekts in voller Auflösung; `@source`: Canvas-Id aus dem IIF-Manifest; `@ulx/uly/lrx/lry`: Bildkoordinaten der gesamten Seite des digitalen Objektes. **<surface>** enthält weiterhin ein Kind-Element **<zone>**.

Das Element **<zone type="page">** wurde im Datathon manuell vergeben. Es umfasst die Umriss der betreffenden Handschriftenseite, ohne den schwarzen Hintergrund des Digitalisats. Das Element hat neben dem Attribut `@type="page"` noch weitere Attribute: `@xml_id`="Signatur\_Blattzahl aufgenullt" – z.B. "Mgo9\_01v" oder "Mgf244\_011v" – und `@points`="Koordinaten der Handschriftenseite". Die `xml:id` wurde manuell vergeben. Die Koordinaten ergeben sich aus dem gezeichneten Polygon/Rechteck. **<zone type="page">** kann folgende Kind-Elemente enthalten: **<zone type="figure">** für Bildregionen und **<line>** für Textzeilen.

Das Element `<zone type="figure">` wurde im Datathon manuell vergeben. Es umfasst eine Bildregion auf einer Handschriftenseite. Diese kann als Rechteck oder Polygon gezeichnet sein. Das Element hat neben `@type="figure"` folgende Attribute: `@points="Koordinaten der Bildregion"`; `@xml:id="Signatur_Blattzahl aufgenullt_laufende Nummer"`. Die Attribute wurden automatisch vergeben. `<zone type="figure">` enthält stets das Element `<note>` und kann weitere Elemente `<zone type="figure">` und `<line>` umfassen, wenn diese innerhalb der Bildregion liegen.

Das Element `<note>` enthält stets die Elemente `<desc>` und `<ref>`. `<desc>` hat das Attribut `@type="description"`: Textkorrelat der ICONCLASS-Notation. `<ref>` hat die Attribute: `@type="iconclass"` und `@target="URI der ICONCLASS-Notation"`.

Das Element `<line>` hat die Attribute: `@n="Zeilennummer in Leserichtung"`; `@points="Koordinaten der Zeile"`; `@xml:id="Signatur_Blattzahl aufgenullt_laufende Nummer"`. `<line>` enthält die Transkriptionen der Textzeile. Jedes transkribierte Wort steht dabei in einem `<w>`-Tag. Eine zeitgenössische handschriftliche Tilgung wurde mit dem Element `<del>` codiert und mit dem Attribut `@rend="striketrough"` spezifiziert.

## Transkriptionsregeln

Grundsätzlich wurde buchstabengetreu transkribiert, d.h. Groß- und Kleinschreibung sowie i/j-, u/v- und s/f-Schreibung folgen der Handschrift. Sonderzeichen und Abkürzungszeichen wurden, sofern nicht über die Tastatur vorhanden oder durch Kopieren eingefügt, in Unicode (hex) codiert. Folgende Sonderzeichen sind vorhanden: Lang-S: ꝛ Unicode (hex) 017F; Geschwänztes Z: ꝛ Unicode (hex) 0292; Tilde: ~ Unicode (hex) 0303 und Er-Kürzung: ʿ Unicode (hex) 035B.