

## **Aufgabe 04a: Kodierung des `teiHeader` und der allgemeinen Dokumentstruktur (Birds of Asia)**

Christof Schöch

DARIAH-DE

Lizenz: Creative Commons Attribution 4.0 (CC-BY).

Das vorliegende Dokument wurde für DARIAH-DE neu erstellt. Das vorliegende Dokument beruht teilweise auf den folgenden Materialien: James Cummings, Renée Baalen, Ylva Berglund-Prytz: "An Introduction to XML and the Text Encoding Initiative", Digital.Humanities@Oxford Summer School 2012, University of Oxford, 2012, <http://digital.humanities.ox.ac.uk/dhoxss/2012/>, CC-BY-SA.

### **Kodierung des `teiHeader` und der allgemeinen Dokumentstruktur (Birds of Asia)**

#### **1. Lernziele**

- \* Kodierung grundlegender Metadaten mit Elementen aus dem Modul `teiHeader`;
- \* Einbinden einer Schema-Datei;
- \* Kodierung von übergreifenden Textstrukturen.

#### **2. Materialien**

- \* Ein partielles digitales Faksimile von John Goulds Birds of Asia (`Gould-Birds.pdf`)
- \* Eine (fast) leere XML-Datei (`Birds-Start.xml`)
- \* Eine Schema-Datei `tei_lite.rnc`

#### **3. Grundidee**

Stellen Sie sich vor, Sie wollten eine digitale Edition von John Goulds Birds of Asia erarbeiten und sich dabei auf die digitalen Faksimiles stützen, welche die BPU de Neuchâtel in der Sammlung e-rara zur Verfügung stellt. Der entsprechende Katalogeintrag findet sich unter folgender Adresse: <http://www.e-rara.ch/doi/10.3931/e-rara-8153>. In dieser Übung legen wir eine der Grundlagen für die Edition des Textes, indem wir die wesentlichen Metadaten und die Textstruktur mit TEI festhalten. Welche weiteren grundlegenden Überlegungen würden am Anfang des Editionsprojekts stehen müssen?

#### **Vorbereitung**

##### **Eine Schema-Datei mit dem XML-Dokument assoziieren**

- \* Öffnen Sie das XML-Dokument `Birds-Start.xml` mit Oxygen.

\* Assoziieren Sie die Schema-Datei mit dem Dokument; wählen Sie im Menü: Dokument > Schema > Schema assoziieren; wählen Sie dann die Schema-Datei `tei_all.rnc` aus und bestätigen Sie dies im Dateidialog.

## A. Metadaten im `teiHeader`

\* 1. Die wichtigsten Metadaten für die neue digitale Edition versammeln:

- \* Titel
- \* Autor
- \* Herausgeber
- \* ...
- \* ...
- \* ...
- \* ...
- \* ...

\* 2. Ordnen Sie nun jeder dieser Informationen in der Liste das TEI-Element zu, mit dem sie sich kodieren lässt;

\* 3. Finden Sie außerdem heraus und notieren Sie, in welchem Teil des `teiHeader`s dieses Element kodiert wird.

\* 4. Kodieren Sie nun die Metadaten im `teiHeader`; achten Sie darauf, dass das Dokument immer valide bleibt!

\* 5. Die wichtigsten Metadaten für das Quelldokument der Edition versammeln:

- \* Autor
- \* Titel
- \* Publikationsort
- \* ...
- \* ...
- \* ...
- \* ...

\* 6. Ordnen Sie nun auch diesen Informationen das TEI-Element zu, mit dem sie sich kodieren lässt.

\* 7. Finden Sie außerdem heraus und notieren Sie, in welchem Teil des `teiHeader`s dieses Element kodiert wird.

\* 8. Kodieren Sie nun die Metadaten im `teiHeader`; achten Sie darauf, dass das Dokument immer valide bleibt!

## B. Die Struktur des Dokuments analysieren und kodieren

\* 1. Das digitale Faksimile `Gould-Birds.pdf` öffnen; es handelt sich um eine verkürzte Version, die nur wenige Seiten enthält.

\* 2. Legen Sie hier eine Liste der strukturellen Einheiten an, die Sie hier unterscheiden können; gehen Sie dabei nicht weiter ins Detail, als notwendig ist, um zur Ebene der Absätze oder Listen zu kommen.

\* Titelseite

\* ...

\* ...

\* ...

\* ...

\* ...

\* ...

\* ...

\* ...

\* 3. Finden Sie heraus und notieren Sie, welche TEI-Elemente für die verschiedenen strukturellen Einheiten verwendet werden.

\* 4. Darüber, wie man eine Titelseite kodiert, haben wir bisher wenig gehört. Nutzen Sie also das entsprechende Kapitel der Guidelines der Text Encoding Initiative, um herauszufinden, wie das geht. Die Guidelines finden Sie unter folgendem Link: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

\* 5. Kodieren Sie auf Grundlage des digitalen Faksimiles möglichst exakt die Titelseite.

\* 6. Kodieren Sie die anderen übergeordneten Strukturen oder Teile des Dokuments. Sie sollten hier maximal bis zur Ebene der Absätze oder Listen gehen.

### **C. Zusatzaufgabe**

\* 1. John Goulds Birds of Asia hat nicht nur einen Band, sondern insgesamt sieben. Recherchieren Sie in den Guidelines, wie man eine solche Struktur in einem einzigen TEI-Dokument abbilden kann.

\* 2. Setzen Sie das Ergebnis ihrer Recherche einmal in Ihrem TEI-Dokument um, indem Sie einen weiteren Band (ohne detailliertere Angaben) kodieren und in dem Dokument verorten.

\* 3. Bereiten Sie sich darauf vor, das Ergebnis Ihrer Recherchen und ggfs. der Kodierung den anderen Teilnehmern kurz zu erläutern.

Ende der Übung!