Stainforth Library of Women’s Writing Data Model

Kirstyn Leuner
http://stainforth.scu.edu/about
kleuner@scu.edu
Documentation: go to stainforth.scu.edu/about
• What is your data model / data structure and how did you arrive at it?

• What if any standards did you consult in arriving at your data model?

• How do you check for accuracy / completeness with your data?

• Do you have any regrets about how you have organized your data?

• What are the biggest challenges you have faced in wrangling your data?
What is our data model?
(great question!)
Most-used tables in Stainforth relational MySQL database
Output in TEI & 18Connect/NINES standards

- Countries
- Entries (transcribed catalog entries with some TEI encoding)
- Entries backup
- Formats
- Genres
- Line counts
- Names
- Persons
- Shelfmarks
- TEI_changes
- Titles
### Snapshot of our Database

**PHPMyAdmin**

| EntryID | EntrySection | EntryPageNumber | EntryLineNumber | EntryTransText_ old | EntryTransText | EntryType | EntryType
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>frontmatter</td>
<td>1</td>
<td>1</td>
<td>15-1-7</td>
<td>15-1-7</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>2</td>
<td>frontmatter</td>
<td>2</td>
<td>2</td>
<td>Catalogue</td>
<td>Catalogue</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>3</td>
<td>frontmatter</td>
<td>3</td>
<td>3</td>
<td>of the Library of Female Authors</td>
<td>of the Library of Female Authors</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>4</td>
<td>frontmatter</td>
<td>4</td>
<td>4</td>
<td>of the Rev J. F. Stainforth</td>
<td>of the Rev J. F. Stainforth</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>5</td>
<td>frontmatter</td>
<td>5</td>
<td>5</td>
<td>The end of the book contains a list of the works o...</td>
<td>The end of the book contains a list of the works o...</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>6</td>
<td>frontmatter</td>
<td>6</td>
<td>6</td>
<td>not obtained by Mr. S before his death.</td>
<td>not obtained by Mr. S before his death.</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>7</td>
<td>acquisitions</td>
<td>1</td>
<td>1</td>
<td>&lt;del&gt;&lt;del rend=&quot;overstrike&quot;&gt;C (A) See&lt;/del&gt;&lt;/del&gt;</td>
<td>&lt;del&gt;&lt;del rend=&quot;overstrike&quot;&gt;C (A) See&lt;/del&gt;&lt;/del&gt;</td>
<td>AuthorOnly</td>
<td>AuthorOnly</td>
</tr>
<tr>
<td>8</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td>A (E) Rhymes. Newry 1831</td>
<td>A (E) Rhymes. Newry 1831</td>
<td>AuthorTitle</td>
<td>AuthorTitle</td>
</tr>
<tr>
<td>9</td>
<td>acquisitions</td>
<td>3</td>
<td>3</td>
<td>A (M.E) See Arnold</td>
<td>A (M.E) See Arnold</td>
<td>AuthorOnly</td>
<td>AuthorOnly</td>
</tr>
<tr>
<td>10</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>Blank</td>
<td>Blank</td>
</tr>
<tr>
<td>11</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td>F9 Abbott (Mrs Charles) Ms. Verses. ND</td>
<td>F9 Abbott (Mrs Charles) Ms. Verses. ND</td>
<td>AuthorTitle</td>
<td>F9</td>
</tr>
<tr>
<td>12</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td>F7 Abdy (Mrs) Arthur's 9th Birthday ND</td>
<td>F7 Abdy (Mrs) Arthur's 9th Birthday ND</td>
<td>AuthorTitle</td>
<td>F7</td>
</tr>
<tr>
<td>13</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>Blank</td>
<td>Blank</td>
</tr>
<tr>
<td>14</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td>? Achilles to Chiron by Lady * * 1738</td>
<td>? Achilles to Chiron by Lady * * 1738</td>
<td>TitleOnly</td>
<td>Blank</td>
</tr>
<tr>
<td>15</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>Blank</td>
<td>Blank</td>
</tr>
<tr>
<td>16</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td>Acton (Mary Harriet) in Friendship's Offering 1844</td>
<td>Acton (Mary Harriet) in Friendship's Offering 1844</td>
<td>AuthorTitle</td>
<td>AuthorTitle</td>
</tr>
<tr>
<td>17</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td>F9-gap reason=&quot;Impliedperson&quot;&gt;Ms. Places</td>
<td>F9-gap reason=&quot;Impliedperson&quot;&gt;Ms. Places</td>
<td>AuthorTitle</td>
<td>F9</td>
</tr>
<tr>
<td>18</td>
<td>acquisitions</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td>Blank</td>
<td>Blank</td>
</tr>
</tbody>
</table>
TEI output from database for a catalog entry

<item xml:id="p6.01">
  <bibl type="transcription">
    <label place="margin" type="shelfmark">M4</label> Alexander (C. F) Hymns for Little Children 25th Ed 4to ND</bibl>
  <bibl type="pagerecord">
    <persName type="stainforth">Alexander (C. F)</persName>
    <idno type="stainforth">M4</idno>
    <title type="stainforth">Hymns for Little Children. 25th Ed</title>
    <title type="authority">Hymns for Little Children. 25th Ed</title>
    <date type="stainforth" when="----">ND</date>
    <note type="details" n="format" resp="#f.quarto">Quarto</note>
  </bibl>
</item>
TEI output from database
For Person Record (ex.)

<person xml:id="pa_AlexanderCecilFrances" sex="F" ref="http://viaf.org/viaf/35879345"/>

<persName>

<surname>Alexander</surname>, <forename>Cecil</forename>

<addName type="other">Fanny</addName>

</persName>

<birth when="1818" />

<death when="1895" />

<sex value="F">female</sex>

</person>
How did we get there?

(see the whole story in our documentation stainforth.scu.edu/about)
General Steps to Digitization: Ms Catalog to Database 1/2

- Scan each page of MS, assign line #s and page #s to PDFs

- Create LUNA edition of the “flippable” digital manuscript, but it has flaws bc it does not replicate the tete-beche experience of the manuscript.

- Transcription in Google Sheets, 1 file per page; We made transcription editorial guidelines in Google Docs and were zealots about updating + following it; In Google Drive Folder: Sheets for tracking editorial progress at each stage; multiple sub-folders for each editorial stage (Transcribed and Ready to Edit, In Process_anypass, Edit Complete)

- Edit transcriptions (in Google Sheets); editors “signed out” each other’s pages on Editorial Progress sheet. Slack for discussing issues/questions.

- Light encoding of edited transcriptions in TEI; Assign Entry Type; Another editorial pass on these components on each page.

- Parse Master Files (by script) into database by Entry Type, Edit parsed data
General Steps to Digitization: Names to Person Records 2/2

• From the database, we generated a list of names that appear in the catalog. They are not regularized, many are hard to recognize/find. Who is in this catalog? How well-known are they?

• Make “person record” for each name containing: Display name, TEI-XML ID, person role (author, editor, translator, composer, publisher, printer, unknown), name(s), gender, birth year, death year, nationality, VIAF permalink, image link, Wikipedia link, ODNB permalink, notes

• Output via script with TEI XML encoding. As-yet unpublished. (Coming Spring/Summer, 1/3 of names?)
It's hard to regret decisions that resulted in flaws bc, at the time, we made the best decision we could with what we knew. But I do regret:

- Letting the former editor of the mapping project tie the Carto map to his personal gmail account. (Bad, Leuner!) He left the project and I had to make a whole new map.
- Not pushing harder for the manuscript scans to represent its tete-beche format and all blank pages.
- Our search interface doesn’t search the whole catalog manuscript. E.g., cross references. But it is all searchable in XML output. Will remedy!
- Not being more fluent in SQL or database construction or javascript. I can write SQL queries to extract data, but can’t on my own change the database itself. I hate that.
- Not being able to delete erroneous person records (programmer; don’t ask … )
- The “Gender” field options in Person Records isn’t quite right. We have Male, Female, Other, N/A, Unknown. This doesn’t account for diff btw sex and gender. Maybe change field name to “Sex”?
- Not pushing harder as a PhD student to make this project my dissertation. My dissertation committee did not let me write on DH.
- Where a title has multiple editions, the editions are not broken out into their own title records. I had originally transcribed and encoded this way, but in parsing something went wrong and it didn’t make it into the database. (Not really a regret, just a bummer … )
Some of our challenges

• Dependency on programmer and scarcity of $$$: I am dependent on our programmer to get content in our database to show up on our website. And to make structural changes to our database. And we have so much data that making fixes across the data set cannot be done by hand. They have to be scripted. Size of data set is HUGE. Everyone says this is awesome. I feel they are wrong.

• Team size has shrunk considerably. And the RA we sometimes have at CU-Boulder is often pulled off of our work to do other library work. (What the heck?)

• Matching names to person records project taking forever (years), isn’t close to done, and is dependent upon funding for RA to help with it. RA funding/allocation is hard to get, and tied to making the experience somewhat like an independent study. Read: labor.

• I want to tag for genre; I badly want to add Worldcat records to make title searches more robust; and doing this by hand would take forever.

• We have a few TEI validation issues, and one expert suggested I write up our own ODD. Need to learn how to do this.

• We have 2 NACO-certified librarians who will submit name authority records for us, but library budget cuts added to their workloads. So: the pace is more snail than hare.