Function Follows Form:

A Practical Guide to Research Data Curation

Julia Collins
Function Follows Form:

A Practical Guide to Research Data Curation

Julia Collins
About the title...

[...] form ever follows function, and this is the law. Where function does not change, form does not change.

*Louis Sullivan, American architect*

https://en.wikipedia.org/wiki/Form_follows_function

Photo: http://calumet412.com/page/159
Data lifecycle elements

- Acquire data
  - Reformat and clean data
- Explore alternatives
- Debug
- Inspect outputs
- Execute scripts
- Edit analysis scripts
- Analysis
- Make comparisons
- Write reports
- Dissemination
- Take notes
- Employ others
- Hold meetings
- Share experiment
- Archive experiment
- Reflection
Data lifecycle elements

Act 1
Data lifecycle elements

**Act 1**

- Acquire data
- Reformat and clean data

**Act 2**

- [Discover and] re-use data

[Diagram showing data lifecycle elements with key steps such as acquire data, reformat and clean data, analysis, execute scripts, inspect outputs, debug, explore alternatives, make comparisons, take notes, hold meetings, write reports, and share experiment.]
Act 1: Field Work
## Act 1: Data compilation

<table>
<thead>
<tr>
<th>Date</th>
<th>File</th>
<th>Site</th>
<th>Community</th>
<th>Treatment</th>
<th>Year</th>
<th>Temperature</th>
<th>Relative Humidity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE14</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 5.78</td>
</tr>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE15</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 6.05</td>
</tr>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE16</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 6.13</td>
</tr>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE17</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 6.05</td>
</tr>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE18</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 2.69</td>
</tr>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE20</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 -999.10</td>
</tr>
<tr>
<td>1995081102</td>
<td>1995081018</td>
<td>BDE21</td>
<td>BD</td>
<td>E</td>
<td>1995</td>
<td>222 8</td>
<td>18 5.96</td>
</tr>
</tbody>
</table>

- **One year per file**
- **Temperature**
- **Site, community, treatment**
- **Relative humidity**

**Note:** Each row represents a data entry with the date, file number, site, community, treatment, year, temperature, and relative humidity.
Act 1: Archive

PI transfers data to archive as per funding agency requirements
Act 2: New application
Act 2: New application
Act 2: New application

Click!
That’s a nice story, but...

How is this relevant to me?

• When you find yourself in the data set creation pipeline, do the right thing.

• Educate domain scientists regarding data structure pros and cons

• Keep in mind the need for care and feeding of previously-created (and archived) data. Multi-season data sets may have components that change over the data collection timespan.
References and Reading


