Data in Dspace: Linking Archival Primary Documents and Quantitative Datasets

Ann Marshall, Rush Rhees Library
IASSIST 2008 Conference
May 30th, 2008

University of Rochester River Campus Libraries
An idea

• The library (partially) funds graduate student research on the condition that the unique dataset collected for the project is deposited into our digital repository. The library archives originating documents with the downloadable dataset created from these documents.

• What if the library helps fund data at the point of data creation rather (or in addition to) the point of data access and/or data archiving?

• Caveats
Overview

• Ongoing frustrations
• A fortuitous request
• The project
• Permission issues & plan B
• Cost/benefit analysis
• Potential challenges
Ongoing frustrations

• Challenges related to proprietary data

• Underutilization of digital repositories
A fortuitous request

• Faculty member request

• Immediate need for support of a graduate student’s research
The project
About Arnd Plagge’s dissertation

The Economic Foundations of World War II

My dissertation stands at the intersection of security studies and international political economy and focuses on two main questions, namely:

- How do trade patterns and internal economic structures change before and during a major war?
- How and why do countries (especially dictatorships) conceal their economic and military capabilities in the shadow of war both in theory and practice?

I try to answer the above questions with the help of an in-depth study of the German, French, British, American, and Soviet experience from 1933 to 1945, and in order to do so, I have compiled several data sets based upon primary sources that will be made available more widely at some point in the future.

For some first results, see the [working paper](#) section.
“Materials acquired under this program will, upon completion of use by the graduate student, become part of the permanent collection of the libraries.”
Requesting Permission from Bundesarchiv

- Role of the graduate student
- Translated correspondence
- Possible ambiguities
- Plan B as a back-up
- Seems like a win-win
- Follow-up with the Bundesarchiv
Conversion of Files to Bundesarchiv Required Format

• Maximum resolution – 75 dpi
• Maximum size – 800 pixels
• Watermark that reads “Bundesarchiv”
• Inclusion of the document’s signature at the bottom of the page (e.g. Bundesarchiv Berlin, R3102/6040)
### Statistisches Reichsamt
Abteilung VIII
Industrielle Produktionsstatistik

Einfuhr kriegswichtiger Rohlinge in Tonnen

Die Zahlen beruhen teilweise auf der Außenhandelsstatistik, teilweise auf Beute- und Frischgut. In allgemeinen ist die Außenhandelsstatistik zugrunde gelegt, worauf der zugehörige Inlandseinsatzen, vgl. die jeweiligen Einzelübersichten.

<table>
<thead>
<tr>
<th>Produktgruppe</th>
<th>Durchschnitt 1938</th>
<th>1940</th>
<th>1941</th>
<th>Januar 1941</th>
<th>Februar 1941</th>
<th>März 1941</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eisenerze, effektiv</td>
<td>1.627.508</td>
<td>331.414</td>
<td>1.446.414</td>
<td>463.046</td>
<td>629.811</td>
<td>740.697</td>
</tr>
<tr>
<td>Eisenerze, Fe-Inhalt</td>
<td>372.165</td>
<td>444.357</td>
<td>676.463</td>
<td>327.457</td>
<td>300.603</td>
<td>326.471</td>
</tr>
<tr>
<td>Stahl</td>
<td>90.199</td>
<td>34.677</td>
<td>35.215</td>
<td>40.234</td>
<td>66.798</td>
<td>123.002</td>
</tr>
<tr>
<td>Eisenkriecherzeugnisse</td>
<td>38.953</td>
<td>35.756</td>
<td>34.732</td>
<td>23.137</td>
<td>46.324</td>
<td>44.632</td>
</tr>
<tr>
<td>Eisenkriecherzeugnisse (Gew.-, Temper- u. Stahlformguss)</td>
<td>67</td>
<td>1.002</td>
<td>1.701</td>
<td>2.482</td>
<td>1.721</td>
<td>2.846</td>
</tr>
<tr>
<td>Stahl- und Gießereizeugnisse</td>
<td>35.553</td>
<td>35.846</td>
<td>35.436</td>
<td>25.805</td>
<td>34.345</td>
<td>38.860</td>
</tr>
<tr>
<td>Eisenerze (über 30 mm De), effektiv</td>
<td>38.700</td>
<td>6.517</td>
<td>13.741</td>
<td>2.412</td>
<td>4.674</td>
<td>11.501</td>
</tr>
<tr>
<td>Eisenerze (über 30 mm De), Fe-Inhalt</td>
<td>18.683</td>
<td>4.227</td>
<td>6.120</td>
<td>1.132</td>
<td>2.161</td>
<td>5.322</td>
</tr>
<tr>
<td>Stahl-Ferroelemente</td>
<td>171</td>
<td>1.423</td>
<td>774</td>
<td>259</td>
<td>342</td>
<td>104</td>
</tr>
</tbody>
</table>
Guide to archival documents
(authored by Arnd Plagge)

A Guidebook to Primary Materials Acquired in the
German Federal Archives and the Political
Archive of the German Foreign Ministry
(as well as several other sources)

⇒ 14. R/3102/5705 — Jahresübersicht über die Rohstoffversorgung 1941
These are a few excerpts from a collection of documents called “Full-year overview of [Germany’s]
raw material supply 1941” that contain data on monthly volumes of Germany’s war-relevant raw
material imports in 1941 [4 pp.].
Title: Jahresübersicht über die Rohstoffversorgung 1941

Keywords: Bundesarchiv
Archival military and trade documents
World War II
Germany
Numerical tables

Issue Date: 1941

Abstract: These are a few excerpts from a collection of documents called "Full-year overview of [Germany's] raw material supply 1941" that contain data on monthly volumes of Germany's war-relevant raw material imports in 1941 [4 pp.].

Description: Source: German Federal Archives of the Bundesarchiv. Gathered by Arnd Plagge in support of his doctoral dissertation research.

URI: http://hdl.handle.net/1802/5568

Appears in Collections: Research Projects in Political Science

Show full item record

All items in UR Research are protected by copyright, with all rights reserved.
Cost/Benefit Analysis: Costs

• **UR Libraries/Friends:** dissertation grant, processing, cataloging, management, and server costs

• **Bundesarchiv:** physical archival costs, staff, management, potential loss of reproduction or other income/benefits.

• **Graduate Student:** time, travel, living and research costs.
Cost/Benefit Analysis: Benefits
(implicit collaboration)

• **UR Libraries/Friends:** exposure, cost-effective sharing of data, serve future researchers

• **Bundesarchiv:** exposure, cost-effective sharing of data, new digital archival copy

• **Graduate Student:** exposure, cost-effective sharing of data, future publications, academic career, etc.

• **Wider academic community:** access to primary documents and data
Potential Challenges

• Permission/legal Issues
• Metadata/Codebook
• Data in Dspace
• Semantics
• Funding
Questions & Suggestions