

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

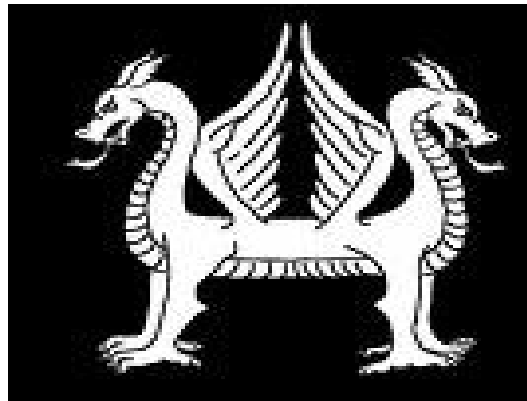
- The library (partially) funds graduate student research on the condition that the unique dataset collected for the project is deposited into our digital repository. The library archives originating documents with the downloadable dataset created from these documents.
- What if the library helps fund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

Friends of the University of Rochester Libraries



[About the Friends](#)

[Benefits of Membership](#)

[Executive Committee & Past Presidents](#)

[Membership Application](#)

[Book Sale](#)

[Calendar 2007 - 2008](#)

[Events](#)

[Metzdorf Award](#)

[Metzdorf Recipients](#)

[Metzdorf Lectures](#)

Graduate Dissertation Fund

“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

Willkommen beim 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)

Sample Archival Document

Statistisches Reichsamt
Abteilung VIII
Industrielle Produktionsstatistik

(Einfuhr kriegswichtiger Roh-
waaren in Tonnen)

Die Zahlen über die Einfuhr beruhen teils auf der Außenhandelsstatistik, teils auf
Seute- und Preisopfer. Im allgemeinen ist die Außenhandelsstatistik zugrunde ge-
nommen. Für das zugrundeliegende Inlandsgebiet vgl. die jeweiligen Einzelübersichten.

	Jahresdurchschnitt			Januar 1941	Februar 1941	März 1941
	1938	1940	1941			
Gußeisen, effektiv	1 027 508	631 454	1 440 434	663 048	620 811	740 697
Gußeisen, Fe-Inhalt	872 195	444 357	676 463	327 467	300 909	321 471
Stahl	90 109	34 677	81 223	40 230	60 798	123 062
Stahl (einschl. Hochofen-Ferroschrott)	30 797	10 142	13 641	13 332	9 586	19 170
Stahl (Rohblöcke und Schweißstahl)	.	4 604	1 461	219	5 972	1 204
Stahlerzeugnisse	35 563	35 766	54 735	23 137	49 634	48 632
Stahlerzeugnisse (Grau-, Temper- u. Stahlforguß)	67	1 062	1 731	2 402	1 714	2 346
Stahler- und Gießereierzeugnisse	35 563	35 148	56 458	25 635	51 340	48 660
Gußeisen (über 30 vH Si), effektiv ¹⁾	35 504	6 917	13 741	2 412	4 574	11 501
Gußeisen (über 30 vH Si), Si-Inhalt ¹⁾	14 653	4 227	6 123	1 132	2 161	5 322
Hochofen-Ferroschrott	171	1 423	774	269	349	104

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.].

Dspace Record

Please use this identifier to cite or link to this item: <http://hdl.handle.net/1802/5568>

Files in This Item:

File	Description	Size	Hits
R3102_5705.pdf	PDF of document	1886Kb	3

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](#)

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