

## Humanities Publishing & Data Curation: Eternal Life & Eternal Damnation



Patrick Alexander, Penn State University Press

Michael Furlough, Penn State University Libraries and Scholarly Communication

June 25, 2010

Fourth Annual Bloomsbury Conference on E-publishing and E-Scholarship



Eternal life = what we are trying to achieve in curation/publishing

Eternal damnation – what trying to do so feels like

The image is of a broadside titled: Das ewige Leben und die ewige Verdammnis, held by Penn State Libraries special collections.

Serves as a source for some other broadsides

## **Pennsylvania State University's Office of Digital Scholarly Publishing (ODSP)**

- conceptualized in the 1990s, collaborations in early 2000s, formally began in March of 2005
- Jointly run by Press and University Libraries
- Several types of cooperative projects underway:
  - The Publishing/Archiving/Hosting of Journal Content
  - The Digital Monograph
  - The Conference Proceedings
  - Digitization of Primary Sources (newspaper project)
  - Paperback POD Editions of Formerly Out-of-Print Books



## **Publishing, Curation, Stewardship**

- Innovation versus preservation
- Publishing in the lifecycle
- Organizing for the lifecycle
- “Adding value”



The tension between research and innovation versus standardization for economies of scale and preservation

Humanists have less interest in replication than scientists. None, probably.

Possible press/library collaboration: Library posts non-publishable appendices (bibliography, images, map) but interest is in pushing functionality.

Libraries: decentralized model of digital content production/mangement/services. Not well organized for hosting/publishing data if we don't control creation. (or at least inexperienced)

Curation is a broad activity, encompasses nearly all aspects of digital life cycle. (assuming joyce ray will introduce this)

Hiring Digital Collection Curator: a starter effort to bring a wider perspective: as manager of all digital collections, she is supposed to coordinate across all aspects.

Problem: too much for one person. Hiring for cultural shift as much as functional activities.

## **German Language Broadside in North American, 1730-1830: An annotated bibliography**

- Extends earlier bibliographies of North American German language printing
- Over 1,800 known single sheet works
- Editors: Hermann Wellenreuther, Reimer Eck, Anne Von Kampe, and Carola Wessel
- Funding from Deutsche Forschungsgesellschaft (DFG)



Extends earlier bibliographic work: *The first century of German language printing in the United States of America*, edited by Karl John Richard Arndt and Reimer C. Eck; compiled by Gerd-J. Bötte and Werner Tannhof using a preliminary compilation by Annelies Müller



## **Publishing the Bibliography**

- Print edition by Penn State Press in the Max Kade Series
  - About 20 illustrations
- Open access electronic edition by Penn State Libraries
  - Include as many facsimiles as possible (all?)



For facsimiles, the researchers acquired a total of 2,851 images of the 1800 broadsides:

A host of issues here, including image quality, provenance, assessment, permissions.

An assumption/hope that OA electronic will allow us to do more with images than the print edition will.

## Database, Manuscript, Curation Object

- Bibliographic data is a primary output of humanities scholarship **and** its production is a form of scholarship itself.
- Digital humanities research is primarily textual/descriptive in nature.
- The raw Broadsides database is an object for curation and a source for two modes of access.
- Replication requires cooperative production work.

## **Penn State Personnel**

- *Project managers*: Furlough & Alexander
- *Technologists*: digitization specialist, metadata specialist, digital collections curator
- *Content managers*: manuscript editor, U.S. history librarian, German studies librarian, rare books curator
- *Design*: interface designer, book designer, production editors

## **Curation actions**

- Selection process
- Assessment and appraisal of data
- Description and documentation
- Determining access restrictions



### 2. Some basic curation actions and how they apply to this project:

#### Selection

"Fit" of the project with Press's list and Libraries' collections  
Do-ability, e.g., capacity and expertise there if not experience

Test premise: Bibliographic data publishing should be simple and repeatable  
How much variation can we tolerate?

#### Assessment & Appraisal of the Data

Making sense of the database, identifying gaps, areas for clarification in content itself  
Images taken as reference photos, not documentary photos  
[[This would be a good place to show some bad examples]]

#### Description and documentation

Actual metadata and documentation somewhat limited  
Two day workshop planning kickoff to clarify many of these matters

#### Determining access restrictions

No complete access to the entire dataset, e.g., no data dump.  
Retaining internal notes as part of the project's record.  
PLUS: where the hell did all these images come from?

## **Appraisal concerns**

- Incomplete documentation
- Image provenance and quality
- Translation accuracy
- Physical communication gap

Identifier	514	Size	19 x 21 cm
Title	Ein schönes Passion Lied	Print Space	
English Title	[A beautiful passion hymn.]	Columns	2
Alternate Title		Lines	28
Start Line	Ihr Sunder kommt gegangen, Seht euren Jesum an...	Typeface	Fraktur
Uniform Title		Watermark	
Author		Paper Quality	
Publisher	Henkel	Meter	6 numbered stanzas, eight lines each.
Place of Publication	New Market, Va.	Tune	
Date of Publication	1816	Main Copy 40	VIU
Language	German	Shelf Number	Broadside 1816 .S35*
Keywords		Additional Copy	VI, Broadside 1816 .S37 Box
Subject - Corporate		Additional Physical Form	Foto gut, VIU. Image, von VI übernommen.
Subject - Personal		URI	<a href="http://vimage.lib.va.us/bsides/images/060883_01.jpg">http://vimage.lib.va.us/bsides/images/060883_01.jpg</a>
Publication		Assignment 50	
German Summary	Beschreibung von Jesu Leiden am Kreuz.	Citation 51	Not in Al. Dolmetsch 128.
English Summary	Description of Jesus' suffering on the cross.	Publications 58	
Project Index	2.3.2. Hymns.	Created 52	20020227
Project Index 28	2.3.2.	Reviewed 53	20070426
Index Term 29	Hymns.	Internal Notes 69	Vgl. B 210; B 438, B 531; U 767, U 767 A. - MWA gives: Pennsylvania, 179-? - MWA hat Original hat, daher von CW als Main Copy geändert. Im Mennonitischen Gesangbuch 2. Aufl., 1808, S. 415-416 enthalten. Nicht in Lutherischem Gesangbuch 3. Auflage (1803); nicht im Reformierten Gesangbuch, 2. Aufl., 1799; nicht in Gemeinschaftlichen Gesangbuch Lutherische und Reformierte Gemeinden, 1834; nicht im Marburger Gesangbuch, 1762; nicht in Reformiertes Gesangbuch, 1829; nicht in Lieder der Hutterischen Brüder, 1962, nicht in Berthelsdorfer Gesangbuch, 1725, nicht im Gesangbuch der Evangelischen Brüdergemeinen, 1741.
Commentary 60d and e		Data Source 95	GyGoN
Page Count	1	TannhofNumber65	B0619C-VIU
Illustration	Border ornament, fillet beneath title and wavy line between columns.	Tannhof.jpg	B0619C-VIU.jpg

This is a straight dump from the original database as handed off to us.

Internal notes here are innocuous...show which checklists that they have checked for this example.

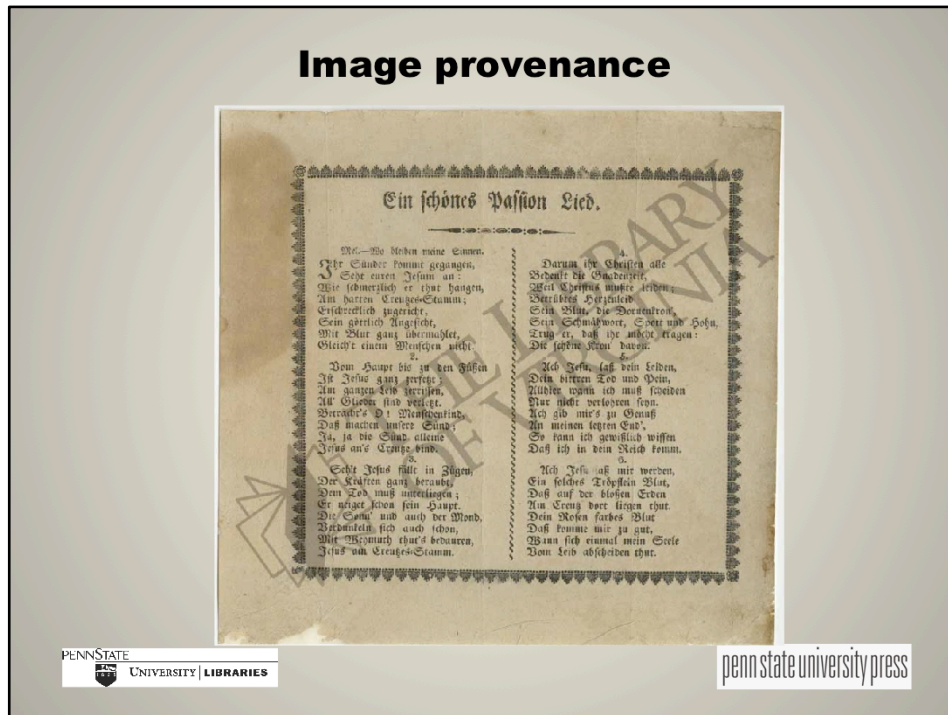
Other comments:

“Original finden, Angaben ergänzen!: Original find! Add details”

“Scholar X says printed in Lexington”

“What is meant by bars? In the left text, the seven words of Jesus are mentioned on the cross. In the right text Hermit and Pope mentioned, Catholic background? Texts of this kind probably very common, even under the title: The seven sacred sky-bars, which received a pious hermit from his protective angels. The seven cross, Jesus' words were in superstition as a powerful spell, they decide the so-called sky bar (n. The sixth and seventh books of Moses, 201 and 206).”

## Image provenance



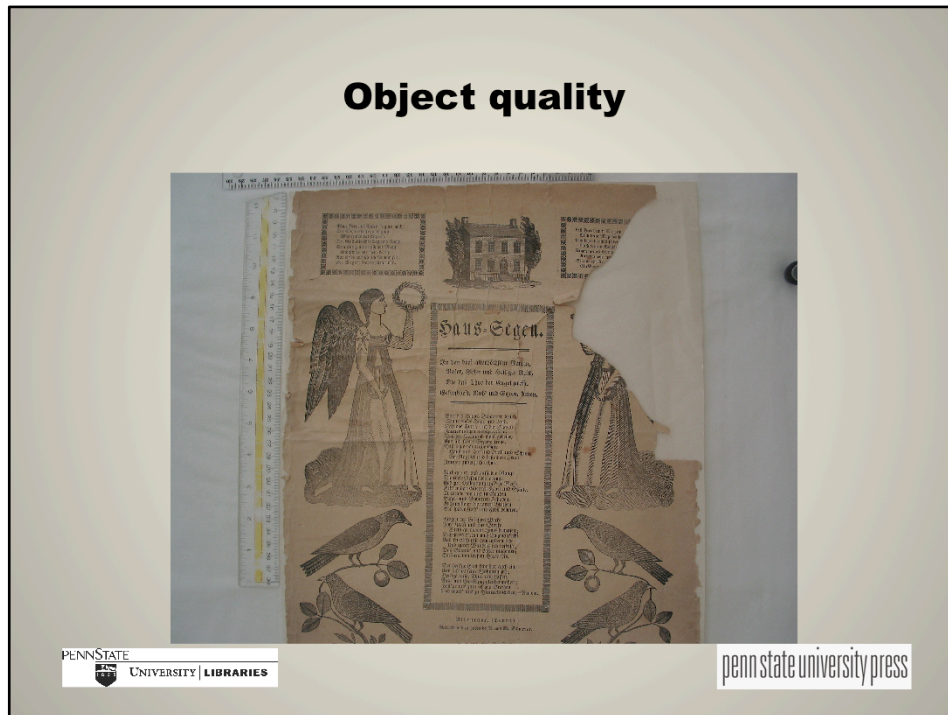
This is the broadside described in previous table.

Broadsides held by 73 different institutions.

165 broadsides held by private owners...who are identified only by last name.



## Object quality



Haus-Segen. In den drei allerhöchsten Namen, Vater, Sohn und Heil'ger Geist [House blessing. In the three highest names, Father, Son and Holy Spirit...]

Researchers consulted four other examples of this same broadside, photographed them all. Others have some lighting problems.

## Image quality 1/2

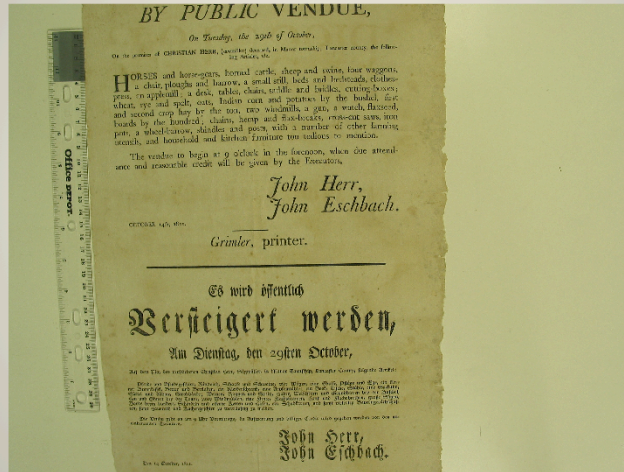


PENNSTATE  
UNIVERSITY LIBRARIES



penn state university press

## Image quality 2/2



PENNSTATE  
UNIVERSITY LIBRARIES

penn state university press

Es wird öffentlich Versteigert werden, Am Dienstag, den 29sten October... 1811 // To be Sold, By Public Vendue, On Tuesday, the 29th of October... 1811.

## Advertisement of animals for sale

## **Publication: beyond curation**

- Data clean-up: the challenge of versions
- Interface development: meeting user needs
- Securing rights and permissions
- Performing work on behalf of the authors/editors



3. Preparation for publication is one step beyond curation (or hyper-curation):  
Data massaging & cleanup

Our hands are deep in the database performing cleanup, etc:

"1 master copy" rule & Keeper of the database  
Taking "executive liberties" for public access

Interface development

This will provide a set of customized search/access points to a dataset

Involvement of subject librarians is in part to consult on user needs

Seeking/soliciting permissions and/or new images

Key area of difference in philosophy between Press v Library  
Press would ordinarily require authors to seek permissions and high-resolution  
images for publication  
Libraries doing this because

We know many of the archives where the originals are held -- personal  
approach  
We don't entirely trust the researcher to get what we need  
And for the entire database it will be way more than the primary editor can  
do

Performing work on behalf of the author/editors

## **The print bibliography process**

- fields selected for filtering, limited to twenty most critical for print
- exported as csv files to library
- converted to Word via mail merge
- editorial preview for missing information, editorial queries, potential stylistic problems
- creation of list of abbreviations and sigla

## The print bibliography: sample output

UNKNOWN. Adam und Eva im Paradies. [Adam and Eve in Paradise]  
“Als Gott die Welt erschaffen, Und alle Creatur...” [“As God created  
the world and every creature”]. *German Summary*: Erschaffung von  
Adam und Eva, Sündenfall. *English Summary*: Creation of Adam  
and Eve, fall. Reading(?): C. F. Egelmann, n.d. 1 page. *Illustr.*: Cut  
of Adam and Eve standing under the tree, Eve to the left, the  
serpent giving Eve an apple, Eve handing it to Adam. Border  
ornaments (two to three lines, various). Border around caption  
(sunbursts). Line around cut. Illustrations at bottom: seal in the  
middle, birds left and right. Colored. *Size*: 40 x 32,7 cm. *Type area*:  
37 x 26 cm. 15 unnumbered stanzas, 8 lines each, additional  
stanza, 8 lines, as caption to illustration. IV,65. f. PLF. OSDR 1,  
Adam N#\*. Not in Müller.



## Antemortem Postmortem

- Imperfect convergences
  - organizational
  - skills
  - missions
- Curation perspective: re-tooling process and outlook.



•In other words: this project is still ongoing (antemortem) but we can still do some assesment (postmortem) [NOTE: too clever]

•Imperfect convergences include

- Organizational expertise and skill, but those can be seen as complementary.
- The mission perspectives of the two project managers/organizations still requires negotiation
- irregular contact can weaken opportunity to leverage expertise
- Faulty assumption: that we had a perfect convergence of project, organizational fit, and funder interest.

Also that of scholar and librarian: approaches to documentation in conflict

- Lib: "where's the admin metadata? Who took the pix?" Hermann: "This is a group effort, we are not taking individual credit."

different perspectives can be a learning experience for all

Anecdote:

Metadata specialist making unilateral decisions (he was authorized to do so) but some were questioned on basis of usability.

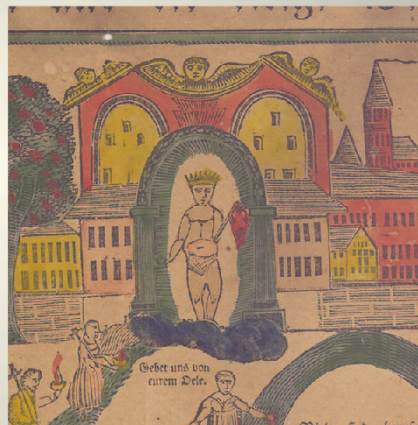
after discussion of potential for confusion in field labels, Metadata Specialist said "I guess I never thought much about the user perspective on the metadata."



## The way of all data?



PENNSTATE  
UNIVERSITY LIBRARIES



penn state university press

This is the slide formerly known as “pulling the rabbit out of the hat once more” e.g., repeatability/sustainability

- Challenge of achieving scale
- Striving for sustainability
- Decision-tree critical choices for the future
- Convergence of interests

We know that the effort we are putting into this will not scale to many other examples of similar nature.

But We think that once created the project is fairly sustainable. Using existing tools in the digital library toolset.

Interface/platform: need to avoid making a decision that will hamstring us or tie our hands for a few years.

Can we afford to include an example of every broadside? Cost of permissions, new copies, etc.

Warning against false dichotomies (Mike’s commentary, maybe not Patricks)

- Raw results // non-peer review // no value vs. Cleaned // peer reviewed // high valuable
- What the collaboration is teaching us is that there is a continuum of value for research outputs
- And there is a continuum of modes of access and preservation durability (faithfulness & authenticity is chimerical)



Variations on a theme: [Die Wege zum ewigen Leben oder dem Ewigen Verderben. Das Neue Jerusalem.] [The paths to eternal life or eternal damnation. The new Jerusalem.]