

# Crowdsourcing Cultural Heritage

Mia Ridge and Ben Brumfield

HILT 2014

August 4-8, Maryland Institute for  
Technology in the Humanities

# Welcome!

- Session 10: Wednesday, 10:45 am – 12:00 pm

# OCR: Print vs. Script

W. W. CALKINS' NORTH AMERICAN LICHENS.	
No	226, <i>Gypsethelium</i>
Name	<i>ochroleucum</i> , var <i>hallescens</i> , Nyl.
Locality	Ft. George, Fla
On	Shrubs

W. W. CALKINS' NORTH  
AMERICAN LICHENS.

yields

**W. W. CALKINS' NORTH  
AMERICAN LICHENS '**

224, *Trypethelinum*  
*ochroleucum*, var  
*hallescens*, Nyl. Isl. Fla  
lity, Ft. George Isl. Fla

yields

— / 2 L -  
` / 17 flfl / v ; / r , 1 `  
\$ / /

# Implications

Methods: Humans vs. Software

Goals: Accuracy vs. Findability

Workflow: Context vs. Granularity

# Findability

WEDNESDAY, OCTOBER 23, 1918

296 days past

69 to come

A good day There was a killing  
first this morning Ben worked  
on Nat woodings car  
Nat come tonight and he got  
his car Ben went to  
Staerla this evening to see  
a boat getting some Lum lumber  
a boat Mr Owens House

# Workflow

◀ BACK TO FRONT PAGE

POST SCORE TO MY WALL

Block...

arfipäivä:  
artipäivä: ✓

81  
81 ✓

>> ✓

127.  
127. ✓

Level: 1

Score: 0

Show help

= Aa Uu Ää Öö f f f f f f

DIGITALKOOT BACKGROUND MEDIA GIVE FEEDBACK TERMS & PRIVACY CONTACT

Swedish

English

### Current project

The current project consists of the National Library's archive of the issues of the newspaper [Aamulehti](#) from the end of the 19th century.

So far 93,460 people have visited the Digitalkoot web site. Volunteers have contributed a total of 266,288 minutes (5,481,490 microtasks) of their time.

DIGI TALKOOT

Digitalkoot on Facebook

Like 4,642

DIGI TALKOOT

**Digitalkoot**  
Digitalkoot wins the Press Jury Prize for most innovative product at MindTrek LaunchPad 2011!

**Lupaavia it-yrityksiä palkittiin Tampereella - Tietokone**  
[www.tietokone.fi](http://www.tietokone.fi)  
Tampereella keskiviikosta perjantaille järjestetty MindTrek-tapahtuma ei houkutellut aivan yhtä paljon osallistujia kuin viime vuonna, mutta järjestäjä...

October 1 at 1:56pm



# Workflow

page

transcribe

versions

[Previous Page](#)

[Next Page](#)

 [zoom in](#)

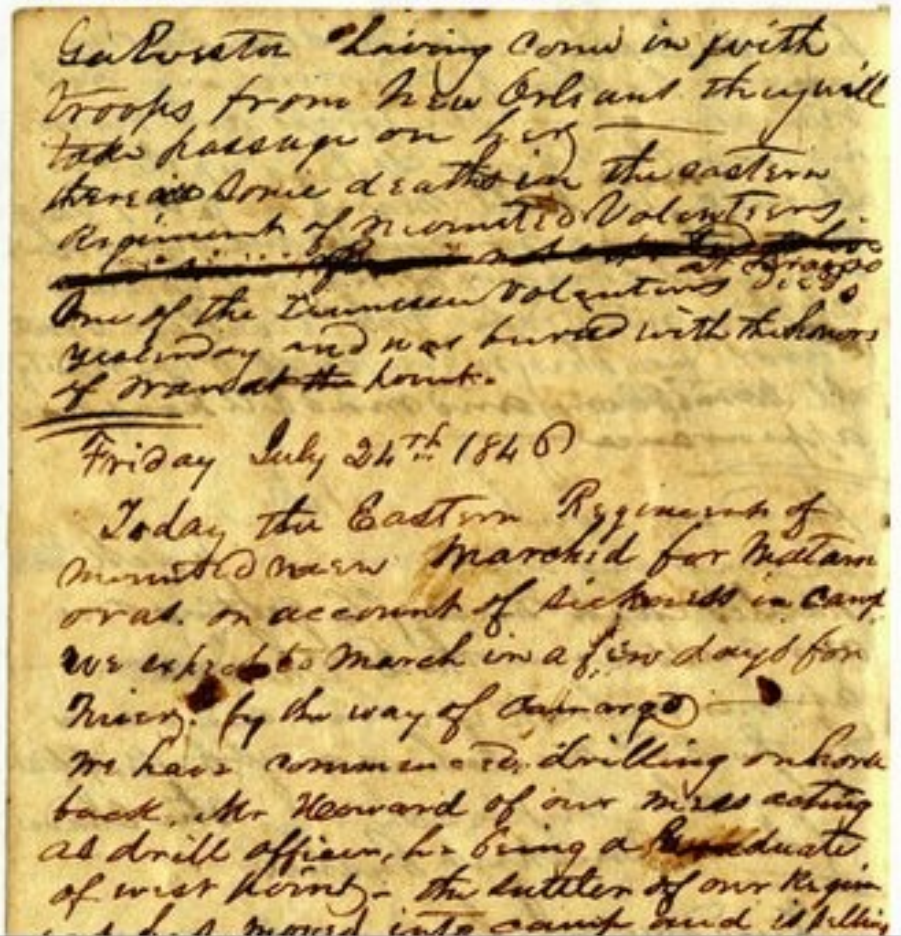
[too small?](#)

19

Galveston having come in with troops from [New Orleans](#) they will take passage on her. There are some deaths in the eastern Regiment of Mounted Volunteers One of the [Tennessee](#) Volunteers at [Brazos](#) died yesterday and was buried with honors of war at the [point](#).

Friday July 24th 1846

Today the Eastern Regiment of Mounted Men Marched for [Matamoros](#) on account of sickness in camp we expect to March in a few days for Mier by way of [Camargo](#). We have commenced drilling on horse back. [Mr Howard](#) of our mess acting as drill officer he being a Graduate



Galveston having come in with troops from New Orleans they will take passage on her. There are some deaths in the eastern Regiment of Mounted Volunteers. One of the Tennessee Volunteers at Brazos yesterday and was buried with honors of war at the point.

Friday July 24th 1846

Today the Eastern Regiment of Mounted Men Marched for Matamoros on account of sickness in camp we expect to March in a few days for Mier by the way of Camargo. We have commenced drilling on horse back. Mr Howard of our mess acting as drill officer, he being a Graduate of West Point. The sutler of our Regiment has moved into camp and is filling

# Workflow

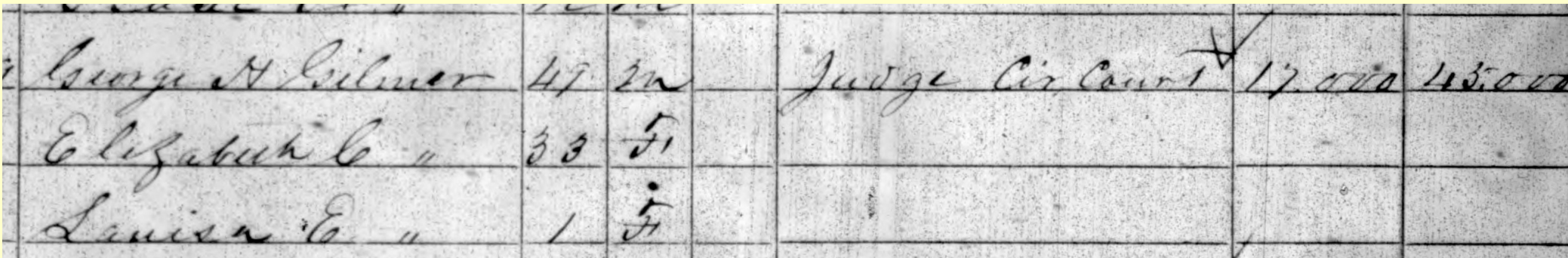
9/29/2011	Available
11/23/2011 (AM)	Transcription
11/23/2011 (PM)	Revision
12/8/2011	6 Revisions
12/9/2011	2 Revisions
8/5/2012	Revision

All by the same volunteer!

# Methodological Origins

What is transcription?

# Indexing

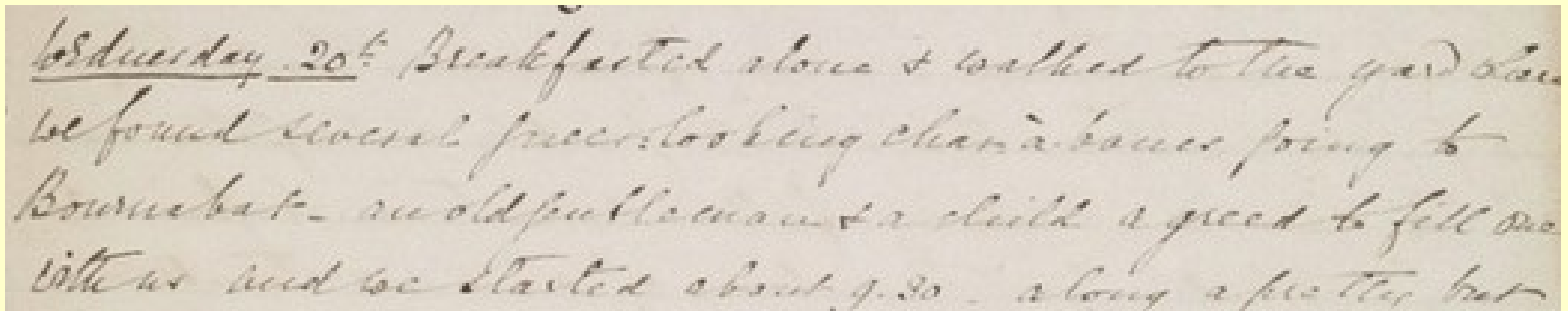


A photograph of a handwritten ledger with multiple columns. The entries are written in cursive. The first column contains names, the second and third columns contain numbers, and the fourth column contains descriptions. There are checkmarks in the fourth column for the first two entries.

George H. Silmer	47	2m	Judge Cir Court ✓	17.000	43.000
Elizabeth C. "	33	51			
Lewis E. "	1	51			

- Structured Data
- Extracting from Text
- Databases for Search and Analysis
- Granular Quality Control
- Gamification

# Editing



Wednesday 20<sup>th</sup> Breakfasted alone & walked to the yard. There we found several free-looking char-a-bancs going to Bournebat - an old publican & a child agreed to fill me with us and we started about 9.30 - along a pretty but

- Books, Diaries, Letters, Articles
- Representing Text
- Traditional Editorial Workflow
- Digital or Print Editions

# Community Origins

- Libraries and Archives
- Documentary & Scholarly Editing
- Genealogy
- Bioinformatics & Astronomy
- Investigative Journalism
- Free Culture



4

We all fasted until after meeting I attended  
conjoint meeting. studied after all retired.

Feb 7. Monday Mapusaga.

I taught school for Willard as he was  
painting. I enjoyed it very well  
but De Soy had to help. I got dinner for the  
Elders, Api, Iaaga and Tumu I visited Va  
who was sick. studied while they all went  
to Priesthood meeting. then retired.

Feb 8. I taught school again had very good

Feb 9. De Soy called on us again.

Now What?

# Libraries and Archives

- Material:
  - Hand-written letters
  - OCRed newspaper articles
- Goal: Findability
- Format: Plaintext transcripts
- Destination: Search engines, finding aids





## Navigation

- [Transcription Desk](#)
- [Transcription Guidelines](#)
- [Select a Manuscript](#)
- [Blog](#)
- [Discussion Forum](#)
- [Recent changes](#)
- [Random page](#)
- [Contact Us](#)
- [Help](#)

## Search

[Page](#) [Edit](#) [Discussion](#) [Watch](#) [History](#) [Move](#)

★ [Transcribe Bentham wins Award of Distinction in Prix Ars Electronica 2011](#); [View manuscripts in full screen](#); [Are you having trouble transcribing after registering your user account? see here](#); [New material available to transcribe: see here](#)

## JB/070/219/002

[Click Here To Edit](#)

### Forgery -- Indirect Legislation

10.

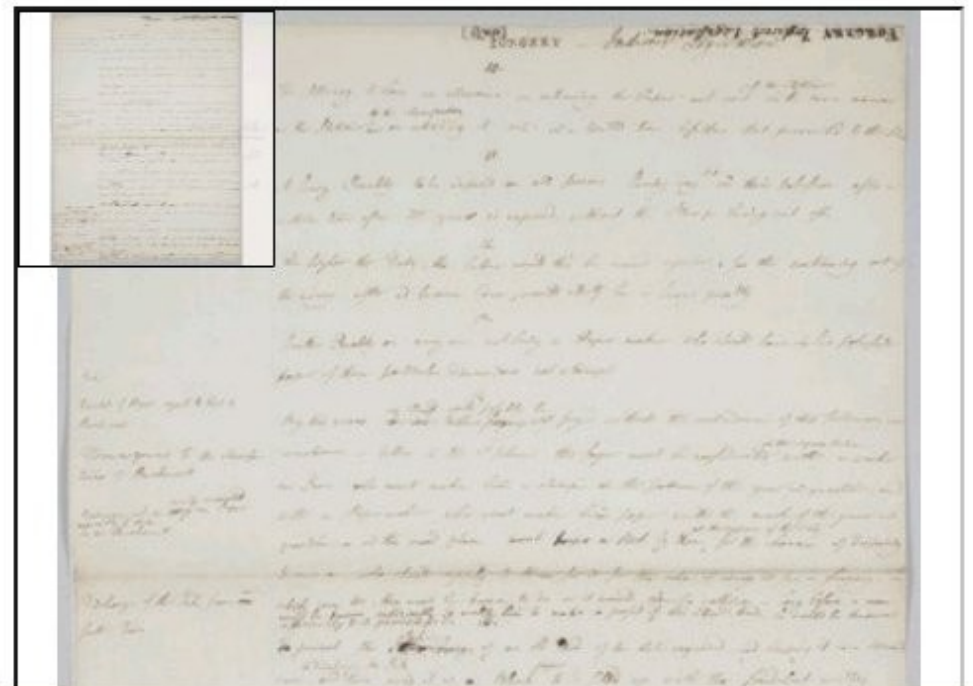
The attorney to have an allowance on returning the Paper not used to the Stationer in the same manner

To prevent more being returned  
to the Stationer than he [ . . . ]

he may mark the sheets with  
his own mark. or the Stationer to the Manufactory on  
returning it viz: in a limited time less than that  
presented to the Stat

11.

A heavy Penalty to be imposed on all persons having



# Documentary & Scholarly Editing

- Material:
  - Literary drafts
  - Historic correspondence
- Goal: High-quality editions
- Format: TEI or other XML
- Destination: Human-readable print or digital editions

# Burials

9

1684	October 30	Janet Wives
	November 6.	The wife of Robert Arthur Burrows
	November 6	The daughter of John Sam Burrows
	November 13	John son of Thomas Burrows Burrows
	November 17	Thomas son of Thomas Burrows Burrows
	November 23	Thomas son of Richard Burrows Burrows
	November 24	Francis son of David Burrows Burrows
	November 2.	Widow Burrows Burrows
	December 31	Robert son of John Cobb Burrows
	January 19	Ann wife of Francis Burrows Burrows
	January 21	Widow Burrows Burrows
	January 27	The wife of Thomas Burrows Burrows

# Genealogy

- Material: Handwritten records
- Goal: Findability
- Format: Structured data
  - Spreadsheets
  - Proprietary databases
- Destination: Searchable databases





U.C. Berkeley  
EMEC 609.711

Vallecillo  
N.L., Mex.  
VI-2-5-1951

P. D. Hurd  
Collector

Cerceris	♂
chilopsidis	♀
Det. by V. & C.	
H.A. Scullen '52	

Cerceris convergens

# Bioinformatics

- Material: Specimens
- Goal: Analysis
- Format: Custom Databases
- Destination:
  - Analytic Databases
  - Scientific Journals
  - Museum Collection Databases

# Tell Us About This Document

DOCUMENTPAGESTEXT

Zoom

Search

p. 1

Contract Agreement Between:

**WNYW**  
205 E. 67th Street  
New York, NY 10065-6050  
(212) 452-5555

**CONTRACT**

Contract / Revision	Alt Order #
4071707 / 4	LINDA MCMAHON

Product	Contract Dates	Estimate #
LINDA MCMAHON/SENATE	10/16/12 - 10/22/12	1342

Advertiser	Original Date / Revision
McMahon / Senate / CT-R	09/12/12 / 10/22/12

Billing Cycle	Billing Calendar	Cash/Trade
EOM/EOC	Broadcast	Cash

Station	Account Executive	Sales Office
WNYW	Anthony Castaneda	FSS Philadelphi

Special Handling	Demographic
	Adults 35+

IDB#	Advertiser Code	Product Code
TV12736		

Agency Ref	Advertiser Ref

And:

**Mentzer Media Service**  
Attention: Anne Braunschweig  
600 Fairmount Avenue  
Suite 306  
Towson, MD 21206

*Line	Ch	Start Date	End Date	Description	Start/End Time	Days	Length	Spots/Week	Rate	Type	Spots	Amount
1	WNYW	10/16/12	10/19/12	Good Day 7a	7a-9a		:30			NM	4	\$5,400.00
Candidate Discount												
		<u>Start Date</u>	<u>End Date</u>	<u>Weekdays</u>				<u>Spots/Week</u>	<u>Rate</u>			
Week:		10/15/12	10/21/12	-TWTF--				4	\$1,350.00			
2	WNYW	10/16/12	10/19/12	Fox News 5p	5p-6p		:30			NM	4	\$3,600.00
Candidate Discount												
		<u>Start Date</u>	<u>End Date</u>	<u>Weekdays</u>				<u>Spots/Week</u>	<u>Rate</u>			
Week:		10/15/12	10/21/12	-TWTF--				4	\$900.00			
E 3	WNYW	10/16/12	10/21/12	FOX 5 News	Tu-Su 6-630p		:30			NM	6	\$6,000.00
Candidate Discount												
		<u>Start Date</u>	<u>End Date</u>	<u>Weekdays</u>				<u>Spots/Week</u>	<u>Rate</u>			
Week:		10/15/12	10/21/12	-T--WTF--				6	\$1,000.00			
	<u>Spot</u>	<u>Ch</u>	<u>Date Range</u>	<u>Description</u>	<u>Start/End Time</u>	<u>Weekdays</u>	<u>Length</u>		<u>Rate</u>	<u>Type</u>		
	2	WNYW	10/15/12-10/21/12	FOX 5 News	Tu-Su 6-630p	-Tu-ThFSaSu	:30		<del>\$4,000.00</del>	NM		
	See MG 3.7											
7	WNYW	10/22/12-10/22/12	10/22/12	Fox 5 News 6p	6p-630p	M-----	:30		\$1,000.00	NM		
	MG for 3.2 10/20											
E 4	WNYW	10/17/12	10/17/12	NLCS Game Adj	4-7p		:30			NM	1	\$12,000.00
Immediately Preemptible												
		<u>Start Date</u>	<u>End Date</u>	<u>Weekdays</u>				<u>Spots/Week</u>	<u>Rate</u>			
Week:		10/15/12	10/21/12	--W----				1	\$6,000.00			
	<u>Spot</u>	<u>Ch</u>	<u>Date Range</u>	<u>Description</u>	<u>Start/End Time</u>	<u>Weekdays</u>	<u>Length</u>		<u>Rate</u>	<u>Type</u>		
	2	WNYW	10/15/12-10/21/12	NLCS Game Adj	4-7p	---W-----	:30		<del>\$6,000.00</del>	NM		
	See MG 4.3											
3	WNYW	10/15/12-10/21/12	10/21/12	NLCS NonHome PrePich/Ter/Various	M-FuWThFSaSu		:30		\$12,000.00	NM		
	MG for 4.2 10/17											
5	WNYW	10/16/12	10/21/12	Late News	M-Sun 10p-11p		:30			NM	6	\$21,000.00
Candidate Discount												
		<u>Start Date</u>	<u>End Date</u>	<u>Weekdays</u>				<u>Spots/Week</u>	<u>Rate</u>			
Week:		10/15/12	10/21/12	-TWTF--				6	\$3,500.00			
6	WNYW	10/20/12	10/20/12	College Football	7p-1000p		:30			NM	2	\$4,000.00
Immediately Preemptible												
		<u>Start Date</u>	<u>End Date</u>	<u>Weekdays</u>				<u>Spots/Week</u>	<u>Rate</u>			
Week:		10/15/12	10/21/12	-----S-				2	\$2,000.00			

(\* Line Transactions: N = New, E = Edited, D = Deleted)

Notwithstanding to whom bills are rendered, advertiser, agency and service, jointly and severally, shall remain obligated to pay to station the amount of any bills rendered by station within the time specified and until payment in full is received by station. Payment by advertiser to agency or to service or payment by agency to service, shall not constitute payment to station. Station will not be bound by conditions, printed or otherwise contracts, insertion orders, copy instructions or any correspondence when such conflict with the above terms and conditions. Two week advance cancellation notice is required unless otherwise specified.

## 1. Who Bought It?

HELP

Which candidate or committee bought these ads?

## 2. What Agency?

HELP

What is the name of the advertising agency that bought these ads?

## 3. Contract Number?

HELP

What is the contract number on this ad buy? (Sometimes labeled "Rep. Order")

## 4. How Much?

HELP

What was the gross total cost? (Sometimes labeled "Report Totals")

Is there something else notable about this file?

Submit

— or skip because —

This is an invoice

This is unreadable or not an ad order

There are multiple contracts in this filing

# Investigative Journalism

- Material: Receipts, FOIA Responses
- Goal: Findability
- Format: Custom Databases
- Destination: News Articles





WIKISOURCE

Hauptseite

Systematischer  
Einstieg

Themenübersicht

Autorenindex

Zufällige Seite

▼ Mitmachen

Letzte Änderungen

Neuer Artikel

Korrekturen des  
Monats

Gemeinschafts-Portal

Skriptorium

Auskunft

Hilfe

Spenden

► Drucken/exportieren

► Werkzeuge

Anmelden / Benutzerkonto erstellen



Seite

Diskussion

Scan



Lesen

Quelltext anzeigen

Versionsgeschichte

Suche

## Seite:Tagebuch Russlandfeldzug 0018.jpg



**F. W. Winkler: Bemerkungen über den Feldzug gegen Rußland in den Jahren 1812 und 1813**

**Fertig.** Dieser Text wurde zweimal anhand der Quelle korrekturgelesen. Die Schreibweise folgt dem Originaltext.

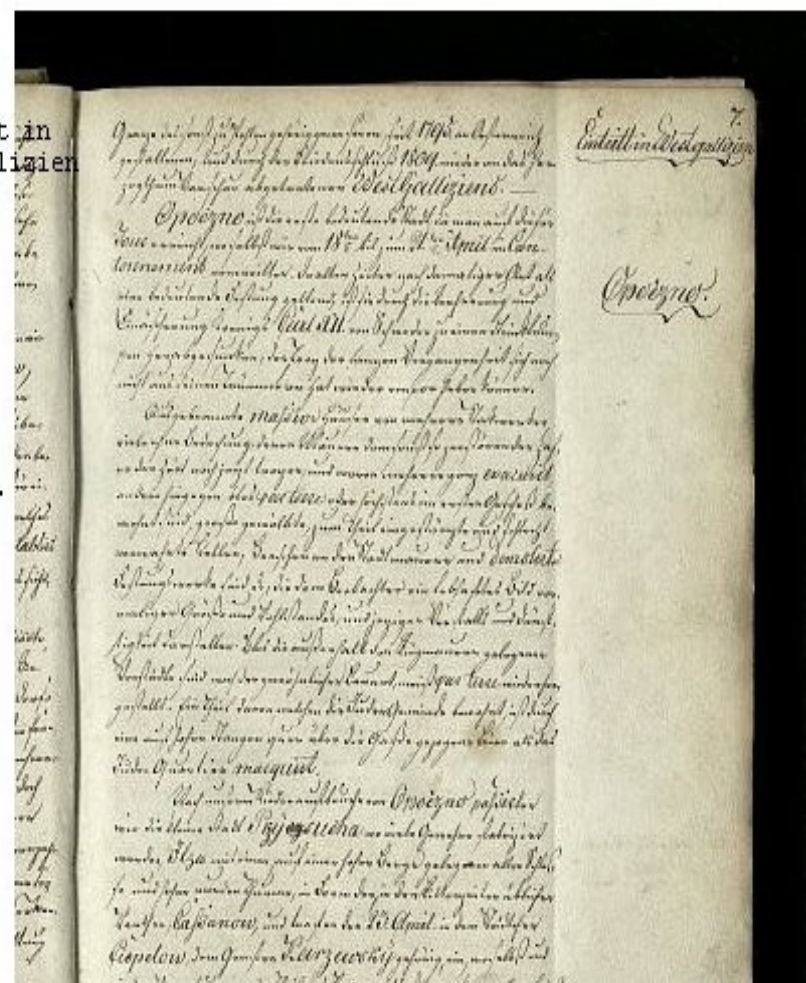
Grenze des sonst zu Pohlen gehörig gewesenen, seit 1795. an Österreich gefallenen, und durch den Friedensschluß 1809. wieder an das Herzogthum Warschau abgetretenen WestGalliziens. -

Opoczno<sup>[1]</sup> ist die erste bedeutende Stadt, die man auf dieser Tour erreicht, woselbst wir vom 18<sup>ten</sup> bis zum 21<sup>ten</sup> April im Cantonnement verweilen. In alten Zeiten nach damaliger Art als eine bedeutende Festung geltend, ist sie durch die Verheerung und Einäscherung Königs Carl XII. von Schweden<sup>[2]</sup> zu einen Steinklumpen herabgesunken, der, Trotz der langen Vergangenheit sich noch nicht aus seinen Trümmern hat wieder empor heben können.

Ausgebrannte massive Häuser von mehrern Stockwerken, viele ohne Bedachung, deren Mauern dem sonst so zerstörenden Zahne der Zeit noch jezt trotzen und wovon mehrere ganz evacuir. andere

Eintritt in  
Westgallizien

Opoczno.



# Free Culture

- Material: OCR and e-Texts
- Goal: Readability
- Format: Plaintext, wiki mark-up
- Destination: Digital editions

# Lab Session

Transcribe exactly what you see!  
Click image to rotate if necessary

### Species/Name of Bird

Common or Scientific Name

AOU Number

### Observer

Observer Name

### Location

Country

State/Province

City/Town

County/Parish

Name of bird <u>Rosebreasted Grosbeak</u>						Year <u>1938</u>
Locality <u>Racine, Wisconsin</u>						
Observer <u>Lulu M. Lunn</u>						
First seen	Number seen	Next seen	Became common	Last seen	Breeds	Winters
				<u>9/13</u>	<u>tol.com.</u>	<u>no</u>

Form BI-801

GPO

8-5280

### Event Observation Data

Reference Year


	Event Month	Day	No. of Birds
First Arrival	<input type="text"/>	<input type="text"/>	<input type="text"/>
Next Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Most Common	<input type="text"/>	<input type="text"/>	<input type="text"/>
Last Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Breeds?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Overwinters?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Commonness?	<input type="text" value="Not Given"/>		

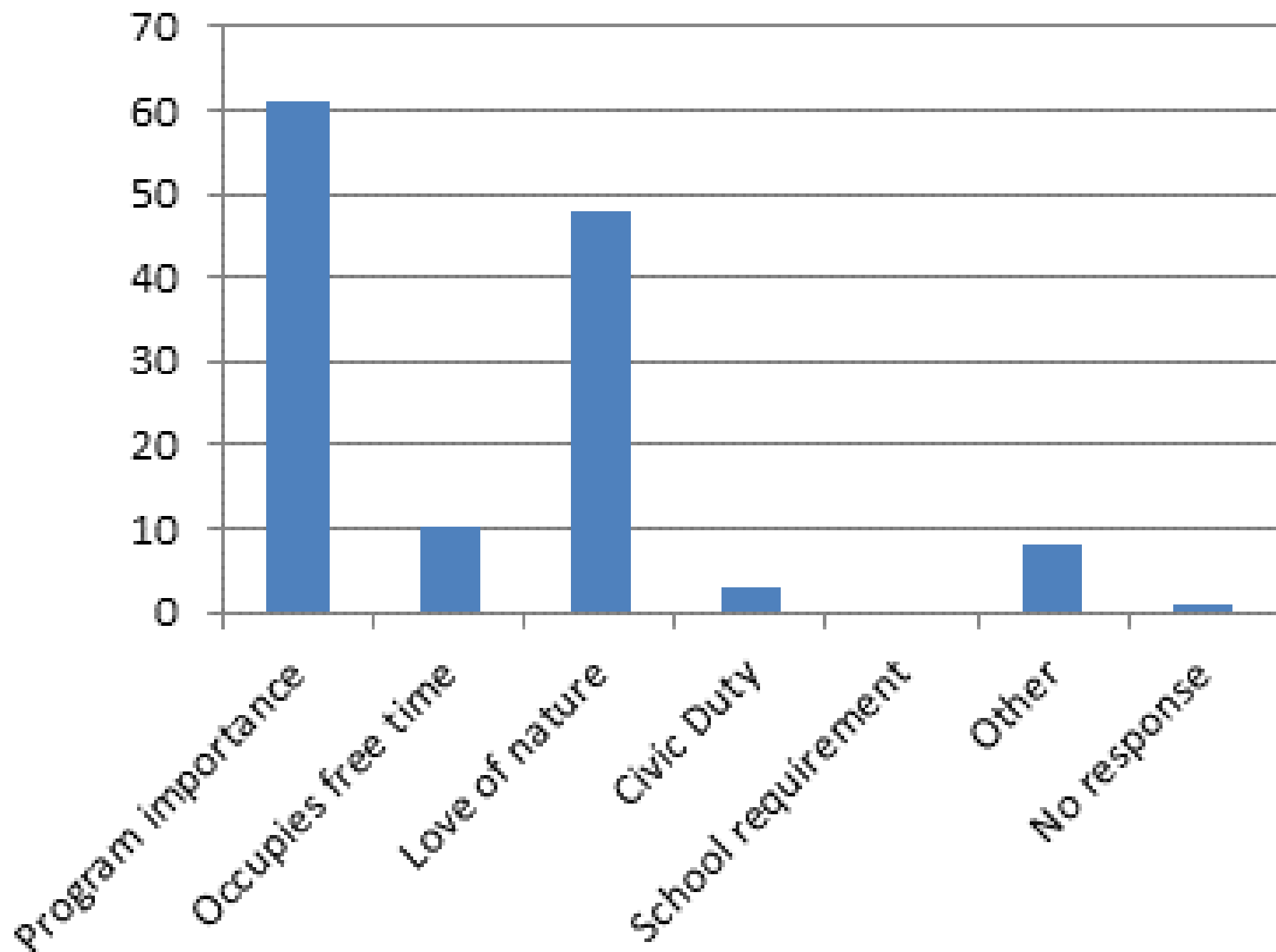
### Additional Notes

Problem with Card?

Observer Notes

Transcriber Comments

 Submit





Transcribe exactly what you see!  
Click image to rotate if necessary

### Species/Name of Bird

Common or Scientific Name

AOU Number

### Observer

Observer Name

### Location

Country

State/Province

City/Town

County/Parish

Name of bird <u>Rosebreasted Grosbeak</u>						Year <u>1938</u>
Locality <u>Racine, Wisconsin</u>						
Observer <u>Lulu M. Lunn</u>						
First seen	Number seen	Next seen	Became common	Last seen	Breeds	Winters
				<u>9/13</u>	<u>tol.com.</u>	<u>no</u>

Form BI-801

GPO

8-5280

### Event Observation Data

Reference Year

	Event Month	Day	No. of Birds
First Arrival	<input type="text"/>	<input type="text"/>	<input type="text"/>
Next Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Most Common	<input type="text"/>	<input type="text"/>	<input type="text"/>
Last Seen	<input type="text"/>	<input type="text"/>	<input type="text"/>
Breeds?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Overwinters?	<input type="radio"/> Yes <input type="radio"/> No <input checked="" type="radio"/> Unknown		
Commonness?	<input type="text" value="Not Given"/>		

### Additional Notes

Problem with Card?

Observer Notes

Transcriber Comments


 Submit



Fig. 11. The two Georges, Cantwell and Willett, at Mount Pinos, May, 1937.

# Lab Session