Enslaved Data Preparation

Methodology from Transcription to the final csv file for submission to the Journal of Slavery and Data Preservation

Kathy Cabell Terlesky, Ph.D.

January 2024

Tool Selection for Managing Data on Enslaved Persons for journal submission

- Needed a genealogy tool that allows capture, maintenance, and visualization of family data.
  - Requirements:
    - Allow addition of additional sources for genealogy data elements
    - Allow adding new types of genealogy events
    - Provide data quality checks (identify and merge duplicates, data clean)
    - Sync with a genealogy on-line site such as Ancestry.com
    - Allow tailored extraction of data elements into a spreadsheet for analytics using python
    - Allow selection of subsets of people for data extraction
- RootsMagic determined to be the best product for this purpose.
  - www.rootsmagic.com

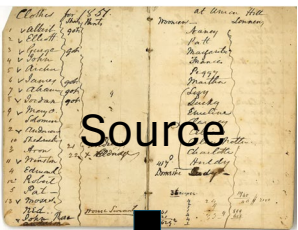# Transcribing methodology
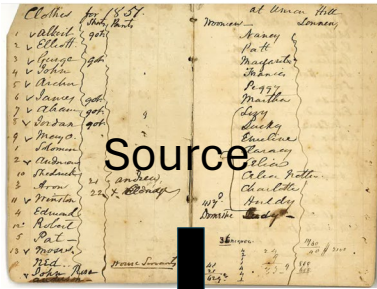
Source
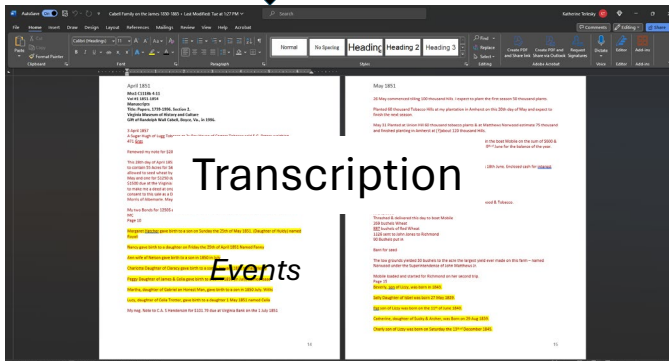
Transcription

Events

1. Visit Repository
   1. Wills, Deeds, Probate Court records – Nelson County Court house in Lovingston, Va., or Amherst County Court House in Amherst Va. Or on-line through Ancestry.com.
   2. Personal journals
      1. University of Virginia Special Collections Library
      2. William & Mary SWEM Special Collections Library
      3. Library of Virginia
      4. Virginia Museum of History & Culture
3. Used camera phone to take photos of artifacts and their call number information.
4. Downloaded and compiled photos into one PDF document by source.
5. Transcribe document on computer using word processor.
   1. Capture source document call number and repository information in document.
   2. Capture page numbers or dates on page of each journal.
   3. Use [?] when a word or letter cannot be deciphered.
   4. Capture abbreviations used by the writers and their full definition.
   5. Highlight events with enslaved persons as transcribing.

Enslaved Data Preparation for JDSP Publication

Method for moving events and names of enslaved persons from original source material such as journals, wills, deeds and newspaper clippings to a CSV file ready for publication in *The Journal of Slavery and Data Preservation* (JDSP)

Requires knowledge of RootsMagic genealogy software and python for Jupyter Notebooks

Source

Transcription

Events

'Events' spreadsheet

RootsMagic Desktop Genealogy Software

Export CSV data tables

1. Person Data by Unique ID
2. Event Data by Unique ID

Jupyter Notebooks

Launch

CSV

JDSP-ready format

Columns- Event ID, Place, Enslaver, Date, Source, Page, Description, Final 'Final' field is all data combined with semi colon separator. Added to RootsMagic Event on 'Description' field.

EVE-ALT-002;Norwood;857;1857-12-03;36;;"Negro Lucy and Negro Ann had a quarrel":
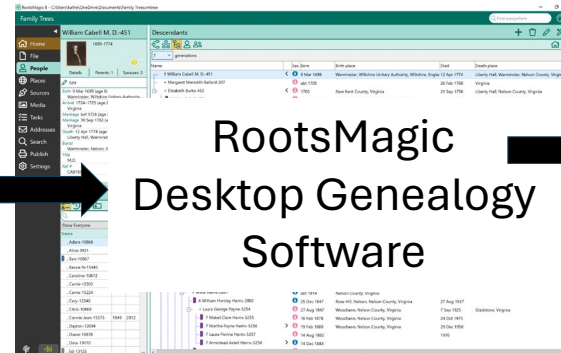
1. Add Person and data from event
2. Estimate birth date
3. Update with new events
4. Add enslaver and spouse data
5. Add overseer data
6. Add relationships
7. Update birth date estimate

Align events to person. Perform data quality checks. Add sources. Add enslaver ID overseer ID associations. Add JDSP unique ID prefix. Perform descriptive analytics. Prepare final csv file.

Compiling the Events Spreadsheet

Excel formula to concatenate across the row using a semicolon separator between column values

Events Spreadsheet

1. Create spreadsheet tab of 'Event Types, definitions and abbreviations consistent with the Enslaved controlled vocabulary recommendations. Expand as necessary during transcriptions.

Source

Transcription

Events

Event Types Master

1. Add events to 'Events' spreadsheet using successive event IDs as the Event ID in Column A.
1. Column Headers
   1. Event
   2. Place
   3. Enslaver ID (RN from Genealogy software)
   4. Date (of event)
   5. Source (document)
   6. Page (in source)
   7. Description (text from source)
   8. Combined combine the first 7 cells in row into one column with ";" separator.
   9. Final column is a copy of the 'combined' with only the text- not the formula. THIS is the cell pasted into RootsMagic.

Color code the event in Column A once it has been entered into RootsMagic to denote completion.

Important: Replace all semi-colons and colons in the original text with periods. The semi-colons and colons are used to facilitate data extraction in python and cannot have this punctuation within the text of a cell in the spreadsheet.

Moving data from the Events Spreadsheet to RootsMagic

RootsMagic Genealogy Software

Manually add people starting with oldest event dates.

Manually add events and the associated data for each person.

Event types in RootsMagic with the Spreadsheet compiled information captured

Use the 'Description field' in the events of RootsMagic to paste the cell contents from the 'Final' column

Establish Groups to bin people as "Enslaved", 'Enslaver', 'Overseer', and 'Union Hill'

Using the 'Custom Report' Feature to extract Person data

Toggle who is included using established groups for Enslaved, enslavers and overseers

Publish Custom Reports
1. Person Data: 3 separate csv files
   1. EnslavedPersonData_date.csv
   2. EnslaverPersonData_date.csv
   3. OverseerPersonData_date.csv

Bring into Jupyter notebook

Jupyter Notebooks

EXCEL STEP REQUIRED
For groups that may have more than one spouse- split these out in Excel into separate columns prior to uploading into Jupyter notebooks.

Fields pulled from each person for the Person Data
This data is created as each new person is added to RootsMagic

Using the 'Custom Report' Feature to extract events by person attached to their unique RN

Toggle who is included using established groups for Enslaved, enslavers and overseers

Publish Custom Reports
1. Event Data: 2 separate reports
    1. EnslavedEventData_date.csv
    2. OverseerEventData_date.csv
2. List of people at Union Hill
    1. UnionHill_date.csv

Bring into Jupyter notebook

Jupyter Notebooks
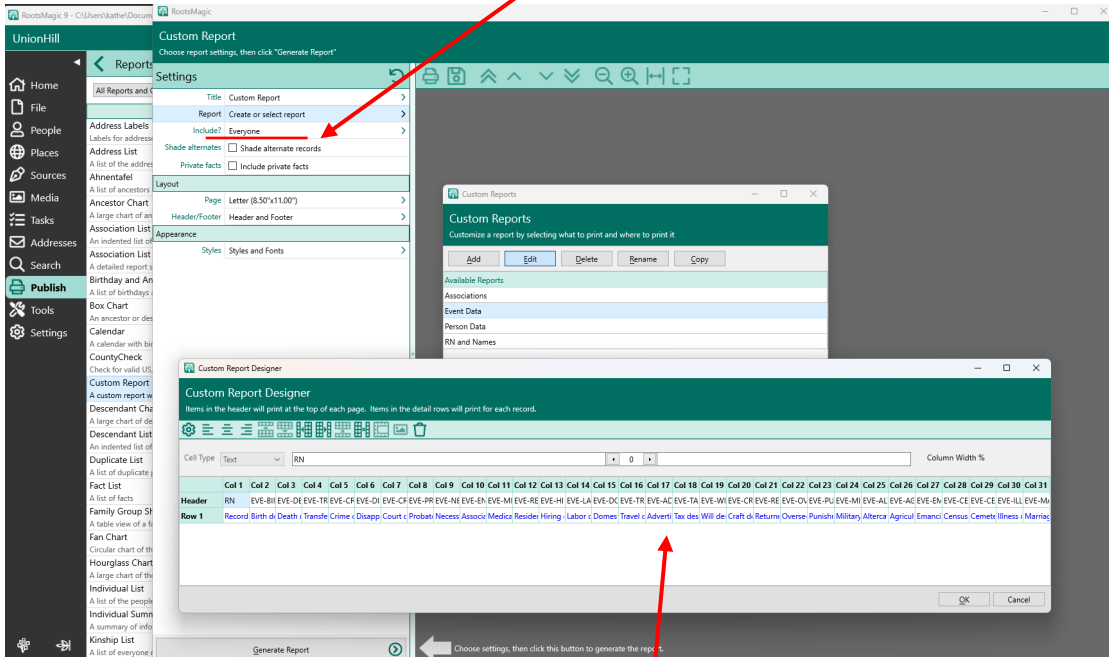
Custom Report – Event Data
Create Custom Report for Events that grabs the 'Description' field for all events.
Save as text file, then use Text Import Wizard in Excel to bring data into a csv. Going straight from RootsMagic to csv will cause misinterpretation of numbers and dates.

EXCEL STEP REQUIRED
Remove all quotation marks from the csv file before uploading to Jupyter. The quotation marks complicate manipulation in python.

First

All Enslaved Person Data by RN

↓

Data Clean up steps

↓

Calculate Child Field from Mother and Father fields

↓

Bring in list of people just at Union Hill and reduce the enslaved list down to those at Union Hill

↓

Add column for status "enslaved"

↓

Person Person DataFrame

Second

All Enslaved Event Data by Person RN One column for each event type
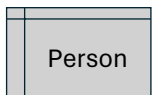
↓

Use wide-to-long to stack event types by row

↓

Separate multiple events into new columns on ";" designation at end of each event

↓

Use wide-to-long to stack unique events by row

↓

Break apart using ";" separator into new columns (Event ID', 'Place', 'Enslaver ID', 'Date', 'Source', 'Page', 'Description')

↓

Data Clean up steps
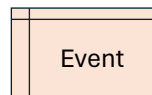
Event Event DataFrame

Third
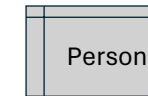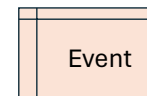
Event → Person

Bring Event data into Person data making 'Enslaved' DataFrame

↓
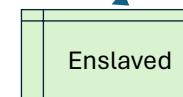
Enslaved

csv → Add source data to each row

Enslaved

↓
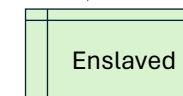
Fourth

Repeat steps 1-3 for Overseer data

↓

Fifth

Bring in Enslaver Person data

Enslaved

- Add Prefix to unique ID
- Combine children to 1 column. Spouses to 1 column.
- Add list of overseers by enslaver to each enslaved
- Bring in lat long for each place

Enslaved

→ Descriptive Analytics in python

↓

Export to csv for JDSP

Final csv file

25 columns

3100 rows

705 people total
- 568 enslaved
- 90 enslavers and spouses
- 47 overseers

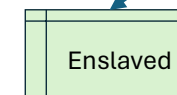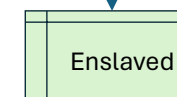| % Complete	Heading	Column	Definition
705 persons	**PERSON DATA**		
100%	Person	A	Unique Identifier for Individual
95%	Given Name	B	Given name
9%	Alternate Name	C	Alternative spellings of given name
30%	Surname	D	Surname if available
98%	Sex	E	Sex if known
100%	RN	F	Record Number in the genealogy data file
100%	Birth date	G	Exact or approximate date of birth
15%	Death date	H	Exact or approximate date of death
22%	Father	I	Unique identifer of Father
30%	Mother	J	Unique identifer of Mother
28%	Spouses	K	Unique identifer of Spouse(s)
15%	Children	L	Unique identifer of children
% Complete	Heading	Column	Definition
594 events	**EVENT DATA**		
100%	Event ID	M	Unique Identifier of an event
100%	Event Type	N	Type of event based on Enslaved.org controlled vocabulary
100%	Place	O	Place of event either plantation, town or city
97%	lat	P	latitude of event place
97%	long	Q	longitude of event place
100%	Status	R	status of individual (enslaved, free, enslaver, overseer)
100%	Enslaver ID	S	Unique ID of enslaver for event
100%	Date	T	Date of event
100%	Description	U	Description of event using narrative directly from source
78%	Overseers	V	Unique Identifier of overseers under the enslaver that may have been associated with the enslaved person
79%	Source Description	W	Source Reference
44%	Page	X	Page in source document for event
79%	Repository	Y	Repository name and location