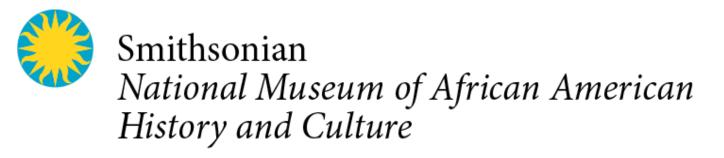
the Taming of the Unstructured Data

Rayvn Manuel



Agenda

- About....
- Brief History of the Bureau
- Congressional Mandate
- Datasets
- Searchability
- Next Steps & Don't //TODOs
- Qs?

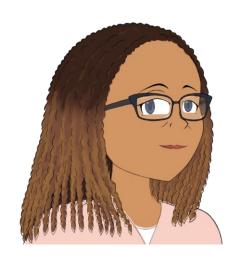


About...

Me

& why am I standing up here speaking to you

Sworn in on 9 January 2017, I serve as the **Senior Application Developer and DevOps engineer** for NMAAHC.



I am the principle developer responsible for the planning, design and implementation of some of the hands-on interactives scattered throughout the museum.

Taming of the Unstructured Data is a tale of transforming two (2) unruly and somewhat feral datasets into a searchable, web-based application and kiosk interactive.

The interactive was designed for the enjoyment of the museum's visitors as well as a research tool for genealogists and historians to explore the digitized archive of the Freedmen's Bureau Records.





Brief History of the Bureau

the "Bureau"



The Bureau of Refugees, Freedmen and Abandoned Lands (Freedmen's Bureau) was established in 1865 by Congress to help millions of emancipated slaves and poor whites in the South in the aftermath of the Civil War.

https://images.app.goo.gl/t56fy6m1Y5diRuKY9

During its years of operation, the Freedmen's Bureau fed millions of people, built hospitals and provided medical aid, negotiated labor contracts for ex-slaves and settled labor disputes. It also helped former slaves legalize marriages and locate lost relatives, and assisted black veterans.



the "Bureau" cont'd

Intended to be a temporary organization to last the duration of the war and one year afterward (1865 - 1872), the bureau failed to provide long-term protection or ensure any real measure of racial equality. In the summer of 1872, Congress dismantled the bureau due to a shortage of funds and personnel, along with the politics of race and Reconstruction.







Congressional Mandate

the Freedmen's Bureau Preservation Act

In 2000, the U.S. Congress passed the Freedmen's Bureau Preservation Act, which directed the National Archivist to preserve the extensive records of the Bureau on microfilm, and work with educational institutions to index the records.

Announced on June 19th 2015, the Freedmen's Bureau Project was created as a partnerships between FamilySearch International and the National Archives and Records Administration (NARA), the National Museum of African American History and Culture, the Afro-American Historical and Genealogical Society (AAHGS), and the California African American Museum.



Datasets

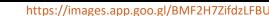
the Freedmen's Bureau Records

https://images.app.goo.gl/hob3RjmpcCd5bhSWA

The records constitute a major source of documentation on the operations of the Bureau, political and social conditions in the Reconstruction Era, and one of the few recorded genealogies of freed people.

NARA provided access to the microfilms to the Smithsonian and the Church of Latter Day Saint's FamilySearch to facilitate and fulfill the record preservation and indexing mandate.

Women and children consisted and wellter the Ship Indenture made the fourteenth where history the first ship of the send to fourteenth where history the first ship to send to the fourteenth of the first ship to send to the first ship to the first ship to send to the first ship to the first shi







the Bureau Records | Smithsonian

[{"title":"North Carolina Assistant Commissioner, Letters Received, Entered in Register 1, S-Y, Part 3", "timestamp": "Tue Oct 24 12:01:20 EDT 2017", "lastTimeUpdated": "Mon Nov 20 12:40:04 EST 2017", "status": 0, "projectName": "North Carolina Assistant Commissioner, Letters Received, Entered in Register 1, S-Y, Part 3", "projectId": "11193", "assetName": "NMAAHC-004567393 00675", "assetStatus": 8, "assetCompleted": false, "transcriptData": "^[[[[underline]] 1516 [[/underline]]]]\r\n\r\n[[preprinted]] \r\nWar Department, \r\nBureau of Refugees, Freedmen and Abandoned Lands, Washington, ^[[October 22\"]] 186^[[6]]. \r\n[[/preprinted]] \r\n\r\nBvt. Maj. Genl. J. C. Robinson, \r\nAsst. Comr. B. R. F & A. L. \r\nState of North Carolina.\r\nRaleigh, N. C. \r\n\r\nGeneral,\r\n\r\nBy direction of the Maj. Genl. Com'r. I transmit by to days mail, forms \"Application of discharged soldier for additional bounty\" and copies of Circular No. 6 from the Claim Division of this Bureau, promulgating extracts from Act of Congress approoved July 28 - 1866, granting additional bounty, and rules and regulations governing the payment of same. Please acknowledge receipt.\r\n\r\nI am General \r\nVery Respectfully \r\nYour Obt. Servant \r\n\r\n [[signature]] W^m Fowler [[/signature]]\r\nA.A. Gen'l.\r\n\r\nL. S.\r\n\r\n ","transcriptDataSplit":["^[[[[underline]] 1516 [[/underline]]]]","[[preprinted]]","War Department, ", "Bureau of Refugees, Freedmen and Abandoned Lands, Washington, ^[[October 22\"]] 186^ [[6]].","[[/preprinted]]","Bvt. Maj. Genl. J. C. Robinson,","Asst. Comr. B. R. F & A. L.","State of North Carolina.","Raleigh, N. C.","General,","By direction of the Maj. Genl. Com'r. I transmit by to days mail, forms \"Application of discharged soldier for additional bounty\" and copies of Circular No. 6 from the Claim Division of this Bureau, promulgating extracts from Act of Congress approoved July 28 - 1866, granting additional bounty, and rules and regulations governing the payment of same

Servant","[[si

24 12:01:20 ED

[*ILLE-*Borth Gredina Assistant Commissioner, Letters Received, Intered in Register 1, 5-V, Part 3*, "insestamp." The Oct. 24 12/81/20 107 2017.

**Satishand Commissioner, Cetters Received, Entered in Register 1

5-V, Part 3*, "projectId-*"1197", "assettame*" "NAWAGE.00056789,00057", "assetStatus '8, "assetCompleted **laise, "transcriptOtat=*"[[[underline]] 1516 [[] Underline]] "Invitational Properties of Verification of the Received Foreign and Abandomed Lands, Unshington, "[(Otober 22V*]] Bio([[0])] Noth[/[preprinted]] \text{Print} Print Pr

"Instrinsedpates" "Non Now 20 32:05:05 157 2017, "status:10, projectNames" "North Carolina Assistant Commissioner, Letters Received, Entered in Register 1.
5-Y, Part 3", projectEd": "INJ37", assettames", "IMPANCE-00557033,00550", "assetStatus" 8, assetCompleted "false, "transcriptOtas"; [[pre-printed]] Departmenthyndureau of Refugees, Freedmen, and Albandomed Lands, hyndrabinington, [[pre-printed]] Sept. 24th 1860/hynkylaMit. Maj. Genl. 3.C.
Redissonlynlands. Commissioner-brinkladigh N.C. hynhylbonengi-lynland and strict when freedmen have been victims or party, according to the official record.
Vinhylbring Delget is to test the truth of statements that have been publicly made, touching this matter, and to enable General Howard to give the facts.
Vinhylbryn RespectfullykynWour Obedient ServantylaA.P. Retchamlyndbt. Maj. & AM General", "transcriptOtaSplit":[[pre-printed]] Mar Department",
"Baleigh N.C.", "General", "I am directed by the commissioner to request you to report (without giving the names of persons concerned) the number of
numbers that have occurred in your district when freedmen have been victims or party, according to the official record., "The object is to test the truth
of statements that have been publicly made, touching this matter, and to enable General Howard to give the facts.", "Yery Respectfully", "Your Obedient
Servant", "A.D. & AAM General"])

["Stills": North Carolina Assistant Commissioner, letters Received, Entered in Register 1, 5-Y, Bart 3", "timestamp": Two Oct 24 12:01:20 EU 2017", 12:01-20 EU 2018", 12:01-20 EU 2018"

DataSet #1: The Smithsonian Transcription center is a crowdsourcing project that allows volunteers [shameless CALL TO ACTION plug] to access microfilmed records which need to be transcribed into digital content.

To include the Bureau records, the Transcription center is tasked with digitizing all of the Institution's historic documents and collection records.

- Data captured & transcribed: The complete record.
- Digital Storage: Internal Enterprise Digital Asset Network [EDAN]
- Digital Image Storage: Internal Digital Asset Management System [DAMS]



the Bureau Records | FamilySearch Intl

DataSet #2: A nonprofit family history organization run by the Church of Jesus Christ of the Latter-day Saints, FamilySearch International relies on its members to transcribe genealogy records as a free service to their members and service subscribers.

A	В	С	D	E	F	G	Н	- 1	J K	L	М	N	0	P	_
FB_DATE	FB_DAY	FB_MONTH	FB_LOCATION	EVENT_TYPE	FB_YEAR	FS_BATCH_LOCALITY	_		FB_IMAGE FS_IMAGE_TYP		_	_	_		_
15 May 1867		.5 May	Calhoun, Georgia, U			United States	4139606	1574230	168	004139606_00168	_		2 004139606		
29 Mar 1867	2	9 Mar	Albany, Dougherty,	Residence	1867	United States	4139606	1574230	157	004139606_00157	English	15-0061;	2 004139606	000000000	Unit
			Georgia, United Sta	Unspecified		United States	4139606	1574230	162 Blocked	004139606_00162	English	15-0061;	2 004139606	000000000	Unit
			Georgia, United Sta	Unspecified		United States	4139606	1574230	147 No Extractable	1004139606_00147	English	15-0061;	2 004139606	000000000	Unit
			Georgia, United Sta	Unspecified		United States	4139606	1574230	224 Blocked	004139606_00224	English	15-0061;	2 004139606	000000000	Uni
11 Mar 1867	1	1 Mar	Savannah, Chathan	Residence	1867	United States	4139606	1574230	225	004139606_00225	English	15-0061;	2 004139606	000000000	Uni
13 Sep 1866	1	.3 Sep	Atlanta, Fulton, Ge	c Residence	1866	United States	4139606	1574230	225	004139606_00225	English	15-0061;	2 004139606	000000000	Uni
31 May 1867	3	1 May	Marietta, Cobb, Ge	c Residence	1867	United States	4139606	1574230	216	004139606_00216	English	15-0061;	2 004139606	000000000	Uni
02 May 1867		2 May	McDonough, Henry	, Residence	1867	United States	4139606	1574230	218	004139606_00218	English	15-0061;	2 004139606	000000000	Uni
20 May 1867	2	0 May	Marion, Georgia, U	n Residence	1867	United States	4139606	1574230	220	004139606_00220	English	15-0061;	2 004139606	000000000	Uni
23 May 1867	2	3 May	Atlanta, Fulton, Ge	c Residence	1867	United States	4139606	1574230	231	004139606_00231	English	15-0061;	2 004139606	000000000	Uni
23 May 1867	2	3 May	Atlanta, Fulton, Ge	c Residence	1867	United States	4139606	1574230	231	004139606_00231	English	15-0061;	2 004139606	000000000	Uni
23 May 1867	2	3 May	Atlanta, Fulton, Ge	c Residence	1867	United States	4139606	1574230	231	004139606_00231	English	15-0061;	2 004139606	000000000	Uni
01 Apr 1867		1 Apr	Atlanta, Fulton, Ge	c Residence	1867	United States	4139606	1574230	231	004139606_00231	English	15-0061;	2 004139606	000000000	Uni
31 May 1867	3	1 May	Savannah, Chathan	Residence	1867	United States	4139606	1574230	200	004139606_00200	English	15-0061;	2 004139606	000000000	Uni
28 May 1867	2	8 May	La Grange, Troup, G	Residence	1867	United States	4139606	1574230	206	004139606_00206	English	15-0061;	2 004139606	000000000	Uni
			Georgia, United Sta	tUnspecified		United States	4139606	1574230	171 Blocked	004139606_00171	English	15-0061;	2 004139606	000000000	Un
01 May 1867		1 May	Columbus, Muscog	e Residence	1867	United States	4139606	1574230	177	004139606_00177	English	15-0061;	2 004139606	000000000	Un
Nov 1866	3	1 Nov	Columbus, Muscog	e Residence	1866	United States	4139606	1574230	179	004139606_00179	English	15-0061;	2 004139606	000000000	Un
01 May 1864		1 May	Columbus, Muscog	e Residence	1864	United States	4139606	1574230	179	004139606_00179	English	15-0061;	2 004139606	000000000	Un
01 May 1867		1 May	Georgia, United Sta	1 Residence	1867	United States	4139606	1574230	183	004139606_00183	English	15-0061;	2 004139606	000000000	Uni
			Georgia, United Sta	Unspecified		United States	4139606	1574230	186 Blocked	004139606_00186	English	15-0061;	2 004139606	000000000	Uni
31 May 1867	3	1 May	Macon, Monroe, Ge	Residence	1867	United States	4139606	1574230	212	004139606_00212	English	15-0061;	2 004139606	000000000	Uni
31 May 1867	3	1 May	Macon, Monroe, Ge	Residence	1867	United States	4139606	1574230	212	004139606_00212	English	15-0061;	2 004139606	000000000	Uni
			Georgia, United Sta	Unspecified		United States	4139606	1574230	43 Blocked	004139606_00043	English	15-0061;	2 004139606	000000000	Un
			Georgia, United Sta	tUnspecified		United States	4139606	1574230	49 Blocked	004139606_00049	English	15-0061;	2 004139606	000000000	Un
			Georgia, United Sta	tUnspecified		United States	4139606	1574230	54 Blocked	004139606_00054	English	15-0061;	2 004139606	000000000	Uni
13 May 1867	1	3 May	Albany, Dougherty,	Residence	1867	United States	4139606	1574230	111	004139606_00111	English	15-0061;	2 004139606	000000000	Uni
31 May 1867	3	1 May	Albany, Dougherty,	Residence	1867	United States	4139606	1574230	113	004139606_00113	English	15-0061;	2 004139606	000000000	Uni
			Georgia, United Sta	Unspecified		United States	4139606	1574230	114 Blocked	004139606_00114	English	15-0061;	2 004139606	000000000	Uni
01 Oct 1866		1 Oct	Atlanta, Fulton, Ge	Residence	1866	United States	4139606	1574230	118	004139606_00118	English	15-0061;	2 004139606	000000000	Uni
01 Feb 1867		1 Feb	Atlanta, Fulton, Ge	Residence	1867	United States	4139606	1574230	123	004139606_00123	English	15-0061;	2 004139606	000000000	Uni
16 Mar 1867	1	.6 Mar	Atlanta, Fulton, Ge	Residence	1867	United States	4139606	1574230	123	004139606_00123	English	15-0061;	2 004139606	000000000	Uni
23 Apr 1867	2	3 Apr	Georgia, United Sta	t Residence	1867	United States	4139606	1574230	123	004139606_00123	English	15-0061;	2 004139606	000000000	Uni
28 May 1867	2	8 May	Augusta, Columbia	Residence	1867	United States	4139606	1574230	102	004139606_00102	English	15-0061;	2 004139606	000000000	Uni
01 May 1867		1 May	Georgia, United Sta	1 Residence	1867	United States	4139606	1574230	105	004139606_00105	English	15-0061;	2 004139606	000000000	Uni
			Georgia, United Sta	Unspecified		United States	4139606	1574230	143 Blocked	004139606_00143	English	15-0061;	2 004139606	000000000	Uni
01 Sep 1866		1 Sep	Georgia, United Sta	1 Residence	1866	United States	4139606	1574230	149	004139606_00149	English	15-0061;	2 004139606	000000000	Uni
01 Mar 1867		1 Mar	Georgia, United Sta	Residence	1867	United States	4139606	1574230	149	004139606_00149	English	15-0061;	2 004139606	000000000	Uni
4 1	TB 2515868 Pers		(+)					1 4							

- Data captured & transcribed:
 Demographic Info
- Digital Content Storage: Internal Database
- Digital Image Storage: Web Accessible URL





Data Consolidation Challenges

FamilySearch Data: Data received from FamilySearch was captured & stored as thirteen (13) commadelimited files with a combined record count of six (6) million entries.

Unedited and viewed in spreadsheet format uncovered that across all the data not one collection set shared a single piece of common data (::groan::: no primary key).

Also revealed - a number of "what not to do" with tabular & normalized data

- Blank cells & null values versus filler zeros.
- Transcription formatting wonkiness
- Multiple pieces of information captured in a single cell
- Problematic field names
- Special characters mixed with the data
- Inclusion of metadata information
- Varied date formats

SI Transcription Center Data: Data received from SI is served from EDAN in JSON format but not accessible via client URL.



Searchability

FB Digital Record Search



Learn more about this project at the SI Transcription Center

Egerton. Jos. J. | C.149 Eaton. W.A. | W.356-H.249 Eddy Jas | D.118 Elyard Danl. | W.361 Evans Amy | W.386.-387 Epps James | W.248 Evans Edward | N.85 Evans Miss | C.169 Records of the Assistant Commissioner for the State of North Carolina, Bureau of Refugees, Freedmen and Abandoned Lands, 1865-1870

Registers of letters received, name index (4) to register 2

Danl Elyard

RECORDS OF:

Assistant Commissioner
| North Carolina

PUBLICATION NUMBER:

M843

ASSET NAME:

NMAAHC-004567389_00223

The Project received financial support from a donor to create a place within the museum – Family History Center – where visitors could sit and search the records for family members who may have been displaced during this era.

This required that the two datasets be combined and the records be matched to cross-reference (as much as possible) each other.

The purpose of the combined datasets was to provide researchers & visitors with a richer and more complete representation of the records and family history.

Additional Benefit: Mapping the demographic information with the complete transcript and stored images could possibly uncover trends during the Reconstruction Era e.g. diseases among the population at that time.





Feature set: Adding Findability



Supporting Technology:

- SOLR 7.4 with custom schema
- Application Language: Angular 7.x

Extensive understanding of the content in order to make judgement calls during cleanup: Consultation with Genealogist & SME Curator

Dream-worthy amount of Data Clean-up

Sifting through each record to verify data consistency; did I mention there are 6 million records?

- Splitting multiple pieces of data
- Standardizing column headers** & making these the indexable fields in SOLR
- Creating a composite key: autoincremented ID column + contrived CollectionID
- Removing beginning spaces





Made it so...

```
FB COLLECTION EDAN IMAGE ID
                                            EDAN ASSET NAME FB DAMS IMAGE URL TC PROJECT TITLE
                                                                                                             TC PROJECT NAME FB IMAGE FB PUB TITLE
                                                                                                                                                                  6191870 NMAAHC-004567385 00037 004567385 00037
                                                               https://ids.si.edu/ids/de Letters sent: Name index t Letters sent, name in 004567385 Records of the Assistant C Assistant Commissioner
                                                                                                                                                                                               North Carolina
          6191870 NMAAHC-004567385 00037 004567385 00037
                                                               https://ids.si.edu/ids/deLetters sent: Name index t Letters sent, name in 004567385 Records of the Assistant C Assistant Commissioner
                                                                                                                                                                                               North Carolina
          6191870 NMAAHC-004567385 00037 004567385 00037
                                                               https://ids.si.edu/ids/deLetters sent; Name index t Letters sent, name in 004567385 Records of the Assistant C Assistant Commissioner
                                                                                                                                                                                               North Carolina
          6191870 NMAAHC-004567385_00037 004567385_00037
                                                               https://ids.si.edu/ids/de Letters sent: Name index t Letters sent, name in 004567385 Records of the Assistant C Assistant Commissioner
                                                                                                                                                                                               North Carolina
          6191870 NMAAHC-004567385_00037 004567385_00037
                                                               https://ids.si.edu/ids/de Letters sent: Name index t Letters sent, name in 004567385 Records of the Assistant C Assistant Commissioner
                                                                                                                                                                                              North Carolina
          6191870 NMAAHC-004567385_00037 004567385_00037
                                                               https://ids.si.edu/id
                                                                                     :--- NMAAHC - Freedmens Bureau Records Fields -->
          6191870 NMAAHC-004567385 00037 004567385 00037
                                                               https://ids.si.edu/ids
          6191870 NMAAHC-004567385_00038 004567385_00038
                                                               https://ids.si.edu/ids
          6191870 NMAAHC-004567385 00038 004567385 00038
                                                               https://ids.si.edu/ids
                                                                                     field name="FB FULL NAME" type="string" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00038 004567385 00038
                                                               https://ids.si.edu/id
                                                                                     field name="FB FIRST NAME" type="string" multiValued="true" indexed="true" stored="true"/>
11
          6191870 NMAAHC-004567385 00044 004567385 00044
                                                               https://ids.si.edu/ids
                                                                                     field name="FB LAST NAME" type="string" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00044 004567385 00044
                                                               https://ids.si.edu/id
                                                                                     field name="FB OCCUPATION" type="text nmaahc" multiValued="true" indexed="true" stored="true"/
13
          6191870 NMAAHC-004567385 00044 004567385 00044
                                                               https://ids.si.edu/id
                                                                                     field name="FB RACE" type="text nmaahc" multiValued="true" indexed="true" stored="true"/>
14
          6191870 NMAAHC-004567385 00044 004567385 00044
          6191870 NMAAHC-004567385 00045 004567385 00045
                                                               https://ids.si.edu/ids
15
                                                                                     field name="FB GENDER" type="string" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00045 004567385 00045
                                                                                    <field name="FB YEAR" type="string" multiValued="true" indexed="true" stored="true"/>
17
          6191870 NMAAHC-004567385 00045 004567385 00045
                                                               https://ids.si.edu/id
                                                                                     field name="FB DATE" type="tdate" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00045 004567385 00045
                                                               https://ids.si.edu/id
18
                                                                                     field name="FB DAY" type="string" multiValued="true" indexed="true" stored="true"/>
19
          6191870 NMAAHC-004567385_00045 004567385_00045
                                                               https://ids.si.edu/ids
20
          6191870 NMAAHC-004567385 00045 004567385 00045
                                                               https://ids.si.edu/id
                                                                                     field name="FB MONTH" type="tdate" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385_00038 004567385_00038
                                                               https://ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids.si.edu/ids
                                                                                     field name="FB COUNTY" type="string" multiValued="true" indexed="true" stored="true"/>
22
          6191870 NMAAHC-004567385_00038 004567385_00038
                                                               https://ids.si.edu/id
                                                                                     field name="FB LOCATION" type="string" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00038 004567385 00038
                                                               https://ids.si.edu/i
                                                                                     field name="FB IMAGE APID" type="string" multiValued="false" required="true" indexed="true"
                                                               https://ids.si.edu/id
          6191870 NMAAHC-004567385 00038 004567385 00038
25
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                               https://ids.si.edu/id
                                                                                    stored="true"/>
26
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                               https://ids.si.edu/id
                                                                                     field name="FB IMAGE ID" type="string" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                                                     (field name="FS IMAGE PAL" type="string" multiValued="true" indexed="true" stored="true"/>
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                               https://ids.si.edu/i
28
                                                                                     field name="text q" type="text general" indexed="true" stored="true" />
29
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                               https://ids.si.edu/id
                                                                                    field name="text s" type="string" indexed="true" stored="true" />
30
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                               https://ids.si.edu/ids
31
          6191870 NMAAHC-004567385 00039 004567385 00039
                                                               https://ids.si.edu/ids
32
          6191870 NMAAHC-004567385_00040 004567385_00040
                                                               https://ids.si.edu/ids
                                                                                     field name="BR FULL NAME" type="string" multiValued="true" indexed="true" stored="true"/>
33
          6191870 NMAAHC-004567385_00041 004567385_00041
                                                               https://ids.si.edu/ids
                                                                                     (field name="BR FIRST NAME" type="string" multiValued="true" indexed="true" stored="true"/
34
          6191870 NMAAHC-004567385 00041 004567385 00041
                                                               https://ids.si.edu/ids
                                                                                     field name="BR LAST NAME" type="string" multiValued="true" indexed="true" stored="true"/>
                                                                                     field name="GR FULL NAME" type="string" multiValued="true" indexed="true" stored="true"/>
                                                                                      field name="GR FIRST NAME" type="string" multiValued="true" indexed="true" stored="true"/>
                                                                                     field name="GR LAST NAME" type="string" multiValued="true" indexed="true" stored="true"/>
                                                                                     field name="FB MARRIAGE PLACE" type="string" multiValued="true" indexed="true" stored="true"/>
```





Next Steps & Samp; //TODOs

BugFixes, Refactors & Enhancements of my

NEXT Steps

Creating an open-source API to make the csv records & EDAN data available to the public without exposing internal bits.

- Use case: Regional Chamber of Congress or Museum who wants limit records for a specific state.
- Use case: Family historian who wants to limit records for a specific surname
- And other use cases that I cannot fathom

//TODOs 2 GET /there

BugFixes

- Address performance issues
- Search is not working
- Wonky mobile view
- Adding SSL certs
- Additional security to SOLR

Refactor

- Change up back-end [Nodejs + GraphQL]
- Adding a back-end data store [redis/mongoDB]

Enhancements

- Pretty-fying UI while maintaining branding
- Adding more whiz-banginess
- Add keyword search





Questions??

the Taming of the Unstructured Data

Rayvn Manuel

