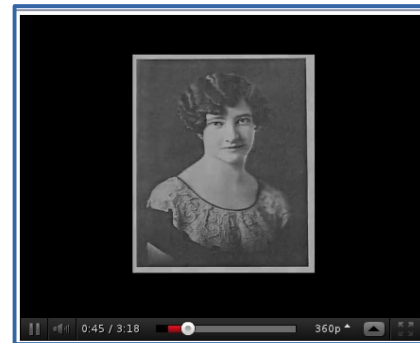
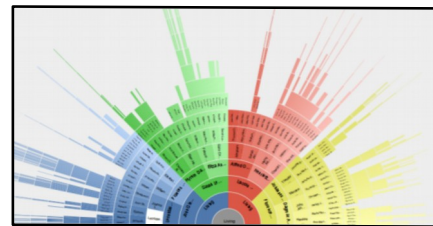


Enabling Emerging Family History Technologies through Open Data



Doug Kennard
dougk@opengendata.org



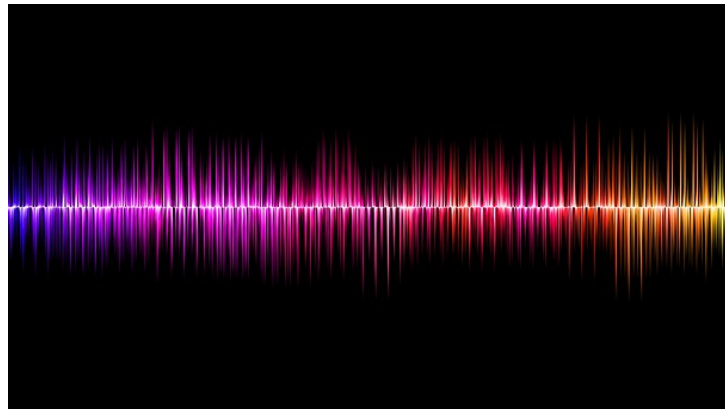
FHTW 2019 (Developer Talk)
Provo, Utah, 26 Feb 2019

Open Genealogy Data (opengendata.org) is
a project sponsored by Historic Journals LLC

Huge Advancements This Decade

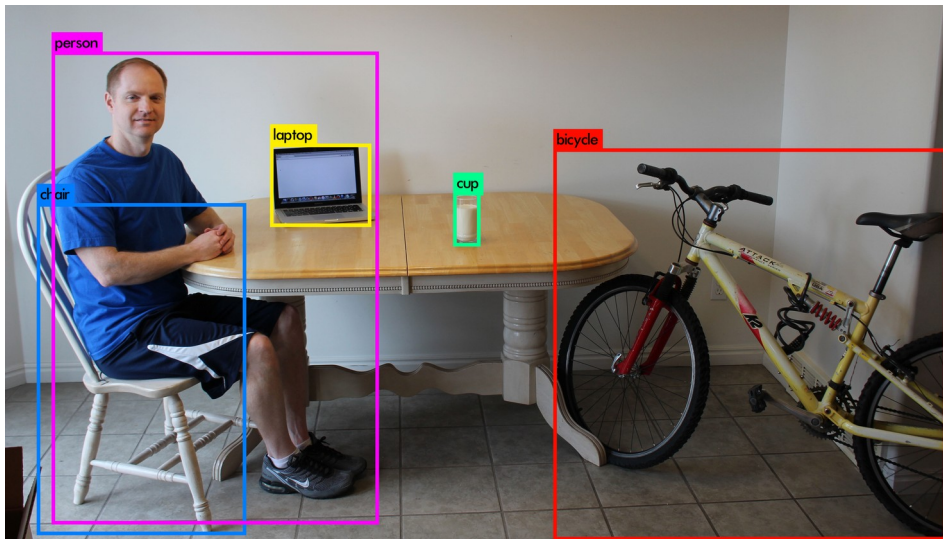


Data Linkage
Handwriting Recognition
Translation
Natural Language Understanding
Text Generation



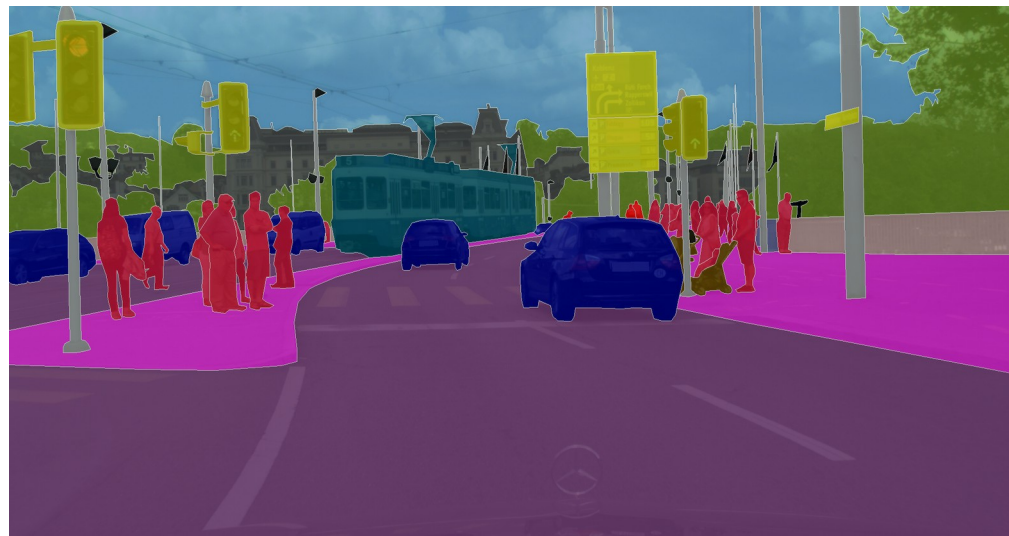
Speech Recognition
Speaker ID
Voice cloning
Text-To-Speech

Huge Advancements This Decade



Redmon, et al. 2018 "YOLOv3: An Incremental Improvement"

Smarter Cameras
Content-Aware and Generative Tools
Auto-Colorization
Face Recognition
Object Classification
Caption Generation



www.cityscapes-dataset.com

"Ditto" for videos
Object Tracking
Video Generation
Deep Video Portraits / Deep Fakes
Vision Systems for Cars
3D, Augmented Reality

Huge Advancements This Decade

Medicine,
Advertising,
Recommendations,
Defense,
Law-enforcement,
Business Strategy,
Economics,
Banking,
Finance,
Fraud Detection,
Drones,
Monitoring Systems,
Robotics,

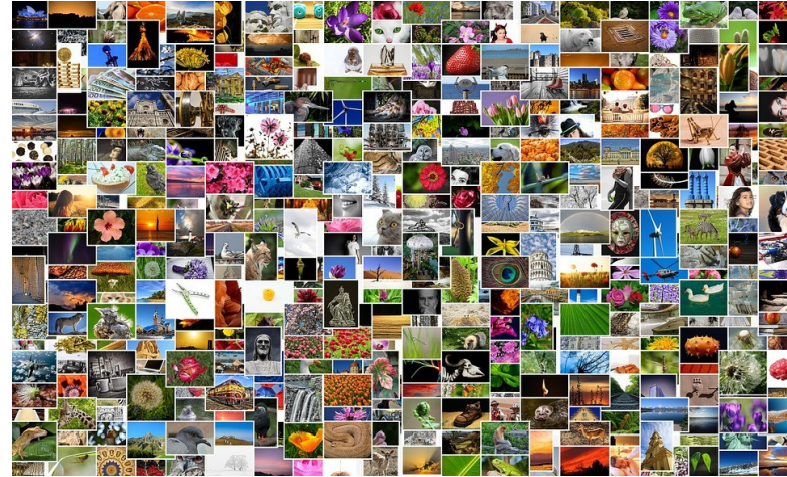
Meteorology,
Brain-Computer Interfaces
Games / Entertainment,
Sports,
Pharmaceuticals,
Education,
Radiology,
Music,
Art,
Geology,
Mining,
Design Optimization,
Astronomy,

... and plenty more

Advancements Are Data-Driven



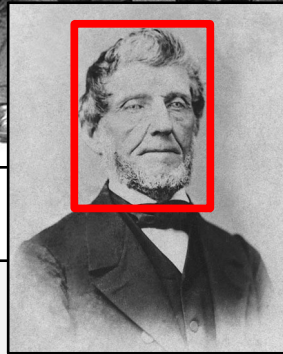
Faces from "Labeled Faces in the Wild" dataset (Huang, et al. 2007)



- Big Companies (Google, FB, Baidu, Amazon)
- Large, domain-specific public datasets
(ImageNet, YouTube-8M, Common Voice, etc.)

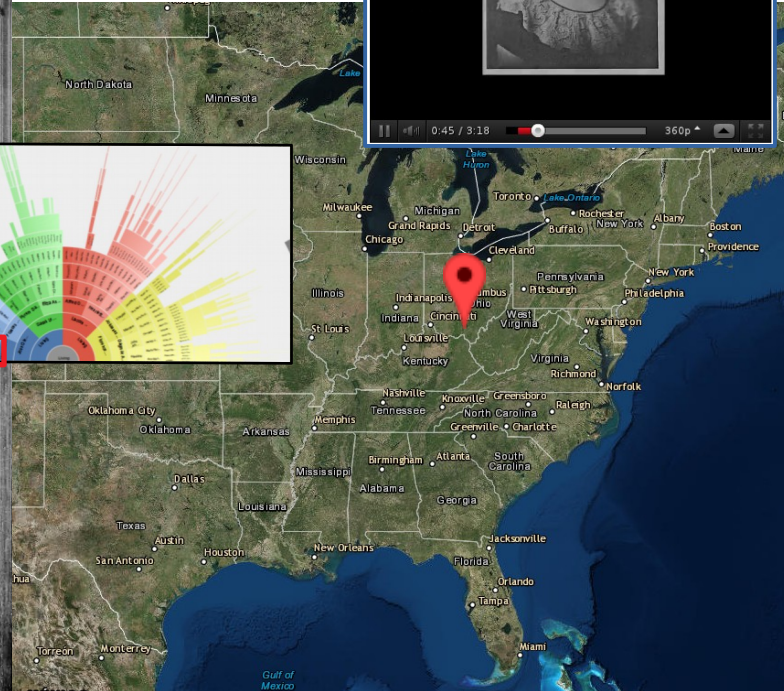
Family History Technology

ted the sick
the past week work
e new honey
e meetings continue
house the evening
to hear Joseph Young
ch
at on an arbitra
between John Shackley
Stovell & N. Steward
d prayer Meeting at
ouse filled to overflow
he Spirit of God
not with us we had
ad Meeting
med minding & preached
afternoon
day Sister Sarah Thomson
living at my house the
d 3 new women. I hope
the sea my work



10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100



10 and adjourned until
meeting of worship. Asked Br Joseph Young to speak.
we adjourned until afternoon, when we met and
asked Br Joseph Young if he would like to act as capt. of
his friends as he wished to travel with, as I had previusly
te him. He said no. He did not want to be capt. of 10 nor

Foundational Data is “Owned” not “Open”

- Limited Access (webpage, API)
- Personal, non-commercial
- Don't Compete/Redistribute
- Convince the Gate-Keeper



Open Genealogy Data (opengendata.org)

Enable, streamline Family History Technology by eliminating hurdles of gaining access to data “owned” by those with business incentives to restrict how it is used.

First Project – US Census

(Kennard, JCDL 2018)

Crowd-sourcing + HR (by Censcript AI)

All volunteer transcriptions to be Public Domain

Dataset for HR research (images + labels)

Open Genealogy Data

You dedicate this transcript to the Public Domain under [CC0 1.0](#)

Indexing as Guest ([login](#) to track your stats)

Roll 0396 Image 719 (ID) Row 8

[Save](#)
[Decline page](#)
[Names complete](#)
[View full image](#)
[Previous](#)
[Next](#)

[Hide instructions](#)

Instructions for names: Please transcribe names as they are written. Try to use the same spelling, capitalization, and punctuation. **If the surname is not written** on a particular row, do not copy it down from a previous row. Instead, type a single hyphen (minus sign) as a place-holder, followed by a single space, then type the rest of the name. **If you are unsure** of a name, please enter your best guess and click the ? (question mark) button to the right to turn it red. Do not type "wildcard" characters (such as * or ?) within the name to indicate uncertainty. If no name is present, skip that line. **When you finish** transcribing all names (or if the page is blank), click "Names complete." If there are problems with the image or field alignment, or if you decide you do not want to transcribe it, click "Decline page" and it will be given to someone else.

Name	Name	Relation	Sex	Age	AgeM
<input type="text" value="Church, Ray A."/> ?	Church, Ray A.	Head	M	29	25
<input type="text" value="- Thursa D."/> ?	Thursa D.	Wife-H	F	23	19
<input type="text" value="- Avrie R"/> ?	Avrie R	Son	M	2 1/2	
<input type="text" value="Campbell, Buck"/> ?	Campbell, Buck	Head	M	37	23
<input type="text" value="- Helen G."/> ?	Helen G.	Wife-H	F	33	19
<input type="text" value="- Virginia E."/> ?	Virginia E.	Daughter	F	14	
<input type="text" value="- Naomi I."/> ?	Naomi I.	Daughter	F	12	
<input type="text" value="-"/> ?	Ray L.	Son	M	10	
<input type="text" value=""/> ?	R. L. & E.	Son	M	8 1/2	

HR (name)

~93% Character Accuracy
(on a sample of confident
human transcriptions)

Kennard, JCDL 2018

Only 5000 labeled names!
Raw output (no post yet)!

HR allows:

Search early in process
Transcribe own family
Error-checking / Hints
Most fields HR-only

106 Results Found

Page 1 of 2 (search took 0.48 seconds)

Year	State	Roll	Image	Row	Last	Given	Relation	Sex	Age	AgeM	
1930	UT	2414	115	32	Davidson	C W	Head	M	39	22	Image Part
1930	UT	2414	115	33	Davidson	Fae B	Wife H	F	36	9	Image Part
1930	UT	2414	117	23	Davidson	John A	Head	M	30	24	Image Part
1930	UT	2414	117	24	Davidson	Ruby	Wife H	F	25	18	Image Part
1930	UT	2414	117	25	Davidson	Myles	Son	M	6		Image Part
1930	UT	2414	117	26	Davidson	Naomi	Daughter	F	5		Image Part
1930	UT	2414	117	27	Davidson	Betty	Daughter	F	3 ⁸ / ₁₂		Image Part
1930	UT	2414	117	28	Davidson	Marie	Daughter	F	/12		Image Part
1930	UT	2414	716	9	Davidson	Russell T	Head	M	25	24	Image Part
1930	UT	2414	716	10	Davidson	Edna	Wife-H	F	26	23	Image Part
1930	UT	2414	716	11	Davidson	Edna	Daughter	F	3 ¹ / ₁₂		Image Part
1930	UT	2414	778	35	Davidson	Mary L	Head-H	F	60	24	Image Part
1930	UT	2414	779	20	Davidson	Herbert	Head	M	26	26	Image Part
1930	UT	2414	779	21	Davidson	Herbert L	Son	M	40		Image Part
1930	UT	2414	779	22	Davidson	Vincint L	Son	M	25		Image Part
1930	UT	2414	914	37	Davidson	E Myrtle	Sister In Law	F	39		Image Part
1930	UT	2415	221	22	Davidson	Rose	Head-H	F	25		Image Part
1930	UT	2415	226	19	Davidson	Iney	Head-H	F	49	31	Image Part

US Census - Geolocation

Township or other division of county

Ward of city

Block No.

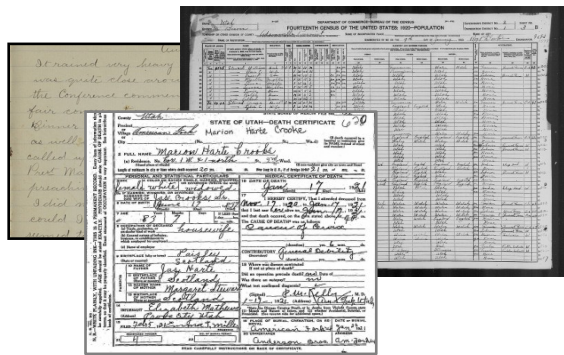
Township or other division of county

Unincorporated place

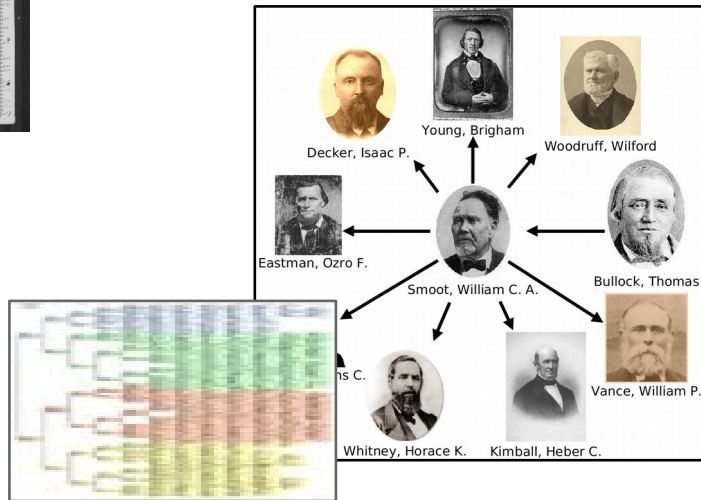
PLACE OF ABODE				NAME				RELATION				HOME DATA				PERSONAL DESCRIPTION				EDUCATION	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
31				15	Brown, Frank	Head	R	52	M	W	43	M	21	no	age	no	age	no	age	no	age
32					— Mattie	Wife	H														
33				16	Marcell, Joseph	Head	R	49	M	W	48	M	21	no	age	no	age	no	age	no	age
34					— Mattie	Wife	H														
35				17	McNamee, Alexander	Head	R	54	M	W	53	M	21	no	age	no	age	no	age	no	age
36					— James S.	Wife	H														
37				18	Finley, Walter C.	Head	R	50	M	W	48	M	21	no	age	no	age	no	age	no	age
38					— Oliver R.	Wife	H														
39				19	Conner, Charles	Head	R	45	M	W	43	M	21	no	age	no	age	no	age	no	age
40	16	3	20	20	Boily, William	Head	R	43	M	W	42	M	21	no	age	no	age	no	age	no	age
41					— James W.	Wife	H														
42					— Sarah	Son															
43	9	21	21	21	Smith, Andrew	Head	R	47	M	W	46	M	21	no	age	no	age	no	age	no	age
44					— Martha P.	Wife	H														
45					— John S.	Son															
46					— Anna	Son															
47					— Mary	Son															
48					— William	Son															
49	17	12	22	22	Johnson, Frank	Head	R	40	M	W	39	M	21	no	age	no	age	no	age	no	age
50					— Mary	Wife	H														
51					— William	Son															
52	13		23	23	Johnson, Frank	Head	R	40	M	W	39	M	21	no	age	no	age	no	age	no	age
53	11	22	22	22	Paul, William	Head	R	40	M	W	39	M	21	no	age	no	age	no	age	no	age
54					— Elizabeth	Wife	H														
55					— Martha C.	Son															
56				24	Johnson, William	Head	R	41	M	W	40	M	21	no	age	no	age	no	age	no	age
57					— Mary	Wife	H														
58	14	12	25	25	Smith, Andrew	Head	R	42	M	W	41	M	21	no	age	no	age	no	age	no	age
59					— William	Son															
60					— Caroline	Son															
61					— William W.	Son															
62					— Sarah J.	Son															
63					— James	Son															
64	18	13	26	26	Johnson, Andrew B.	Head	R	47	M	W	46	M	21	no	age	no	age	no	age	no	age
65					— William	Wife	H														
66					— Elizabeth	Son															
67					— Marie	Son															
68					— John A.	Son															
69					— Robert L.	Son															
70	21	14	27	27	Reese, Joseph	Head	R	43	M	W	42	M	21	no	age	no	age	no	age	no	age
71					— J. Paul	Son															
72					— Mary	Son															
73					— Helen	Son															
74					— William Henry	Son															
75	14	15	28	28	Johnson, William	Head	R	45	M	W	44	M	21	no	age	no	age	no	age	no	age
76					— Daniel	Wife	H														
77					— Joseph	Son															
78	13	16	29	29	Taylor, William	Head	R	48	M	W	47	M	21	no	age	no	age	no	age	no	age
79					— Helen	Wife	H														
80					— William	Son															



Future Open Databases



Records and Documents



Global Genealogies, Links,
Historical Social Networks



Media
(People, Places, Events)

Other Open Data Initiatives

- **Organizations** (Reclaim the Records, Free UK Genealogy)
- **Libraries** (Allen Co. Public Library, NYPL, British Library)
- **Museums** (Met Museum of Art, Getty Museum, Nat'l Gallery Art)
- **Archives** (NARA, LOC, Internet Archive)
- **Governments** (US, some others, various State and Local)

Invitation to Open Data

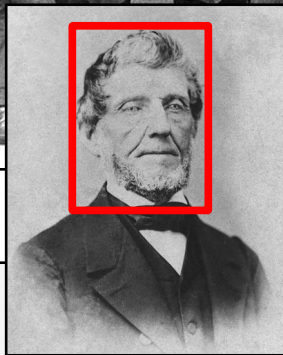
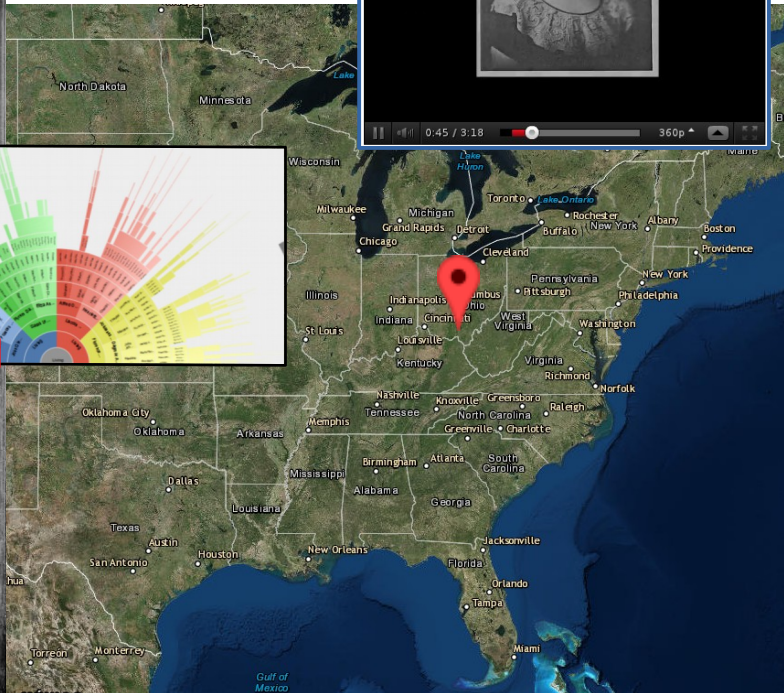
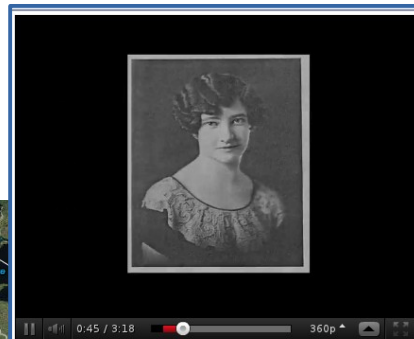
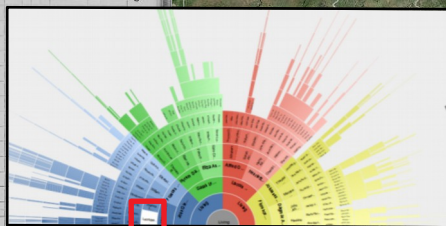


- Public Domain (or Attribution)
 - avoid “Non-Commercial”
 - avoid “Share-Alike”
- Data snapshots (vs. API)
 - cheaper (develop, maintain, host, use)
 - easier, flexible, full-access
 - permanent (mirror/archive)

Download



ted the sick
the past week work
new honey
meetings continue
house the evening
to hear Joseph young
ch
of at an arbitrate
between John Mackey
Gorvil & Stewart
d prayer Meeting at
house filled to overflow
the spirit of God
not with us we had
ed Meeting
ned meeting & preached
afternoon
day Sister Sarah Thomson
living at my house the
young woman I had
with her very much

[illegible]

University Harold B. Lee Library Special Collections VMss 227 v2

and B. Joseph Young to speak.

Joseph Young if he would like to act as capt. of

is friends as he wished to travel with, as I had prevsily
te him. He said no. He did not want to be capt. of 10 nor

Thank You!

Doug Kennard

dougk@opengendata.org

Open Genealogy Data (opengendata.org) is
a project sponsored by Historic Journals LLC