



**Drawers, shelves & boxes full of data:
Status of analog life sciences data and
solutions for the future**

**Shannon L. Farrell,
Kristen L. Mastel &
Julia A. Kelly**



What about older, non-digital data?

Impetus for this project was faculty approaching us for help with 100-year old potato breeding records

- What to do with this data was not clear: Does it belong to Archives? Should it stay in the department?
- Horticulture faculty wanted the Libraries to have it



We know there is analog data on campus





Who have we talked to?

Department of Applied Economics



Minnesota Landscape
ARBORETUM



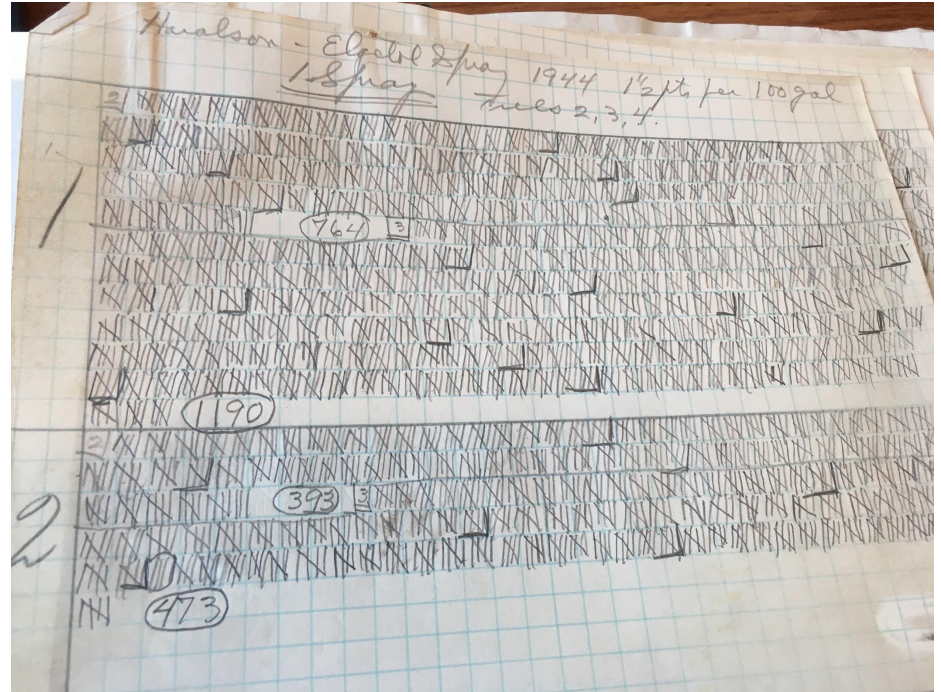
BELL MUSEUM
OF NATURAL HISTORY





Project in University Archives

BRITISH NORTH BORNEO		COLL.	DISTRIBUTION
MT KINABALU Lat. 6° 03' N Long. 116° 32' E			
February 7, 1962			
No. 415 = Kaul 2031			
No. 416 <i>Gordonia</i> aff. <i>excelna</i> King fide Sinclair	BB	K A B S	
Small to large trees along trail from Kambaranga to Paka Cave. Petals white, numerous stamens dark yellow, flowers ca. 2-3 in. diam., Alt. above 7,000 ft.			
February 8, 1962			
417 <i>Trachymene saniculaefolia</i> Stapf fide J.A.R. Anderson	BB	K A U M B S	
Small lax plants in clearing approx. 1 mi. above Kambaranga, on trail to Paka Cave. with <i>Dipteris</i> . Stems red, flowers pink. Alt. ca. 8,000 ft.			
No. 418 <i>Diplazium pittosporifolia</i> J. J. Sm. var. <i>punctiloba</i> Sleumer fide J.A.R. Anderson	BB K	G B S	
Collected in forest above Kambaranga. Alt. ca. 7500'.			
No. 419 <i>Lindera rufa</i> Gamble fide J. A. R. Anderson	BB K	K A G U M E B S	
Tree. Vicinity of Kambaranga. ca. 7050'.			
No. 420 <i>Pygeum cocarpam</i> Stapf fide J.A.R. Anderson	BB K	K A G U M E B S	
Small tree. Flowers pinkish. Leaves leathery, shiny above and beneath. Veins reddish beneath. Stems of new growth abundantly rusty pubescent. Above Kambaranga, ca. 7500'. (=Abbe 10179)			
No. 421 <i>Quercus havilandii</i> Stapf fide Wm. Meijer fide J.A.R. Anderson	M BB		
In forests above the Kambaranga Radio Station. Alt. ca. 7000-9000'. Flowering inflorescences pickled.			
February 10, 1962			
No. 422 <i>Utricularia orbiculata</i> Wall. fide Sinclair	BB M	B	
Growing among mosses on wet rocks along stream at base of water fall, in gorge to west of Kambaranga radio station. Elev. ca. 6150'. Petersen & meander.			





Inventory elements

- Collection name, # of boxes
- Box number and folder/bound item name
- Physical format of data
- Creator/author of data
- Description of topic of data
- Location of where data was taken
- Numeric or descriptive
- Amount of raw data (pages, etc.)
- Are variables decipherable?
- Is there description of methods?
- Does summary data exist?
- Could you repeat the work?



Name of collection	Person or unit (P or U)	Total # of boxes	# of boxes pulled	# of boxes w/ data	# of data items found	Numeric data only	Descriptive data only	Both numeric and descriptive data	Date range	Geographic location
Botany researcher	P	9	8	8	40			X	1939-1964	MN, outside of US
Wildlife management researcher	P	3	3	3	53			X	1939-1969	MN, other US states
Plant pathology researcher	P	51	8	5	49			X	1948-1991	MN, other US states, outside of US
Ecology researcher	P	11	3	2	20	X			1956-1998	other US states
Plant pathology researcher	P	1	1	1	15			X	1921-1952	MN, other US states, outside of US
Agronomy researcher	P	8	2	1	1	X			1961-1962	Unknown
Fisheries researcher	P	21	9	9	137			X	1947-1978	MN, outside of US
Ornithology researcher	P	1	1	0	0	-		-	-	-
Forestry Center	U	12	4	4	30			X	1911-1956	MN, other US states
Department of Agricultural Biochemistry	U	9	3	2	7	X			1933-1950	Unknown
Department of Botany	U	186	22	16	184			X	1879-1996	MN, other US states, outside of US
Department of Horticultural Science	U	35	3	3	97			X	1890-1953	MN
Biological Field Station	U	10	1	1	31			X	1960-1961	MN
TOTAL	8 P, 5 U	357	68	55	664	3	0	9		

Data items per decade



Location	# of Datasets
MN	357
Other US states	78
Outside the US (every continent except Antarctica)	113



Reusability of data

- Over 49,000 pages of data in 672 items
- 140 or 21% describe their methods
- 239 or 36% could probably be repeated
- 628 or 94% had data of potential value
- 88% were at least partly handwritten



Survey purpose

- Explore how much analog data exists in laboratories, offices, departments, research centers, and field stations
- Discover how researchers are currently using, storing, and planning for the future of the data



Survey participants

- Focused on agricultural and life sciences
- Faculty from College of Biological Sciences & College of Food, Agricultural, & Natural Resources Sciences
- Sent to 774, 108 responded (14% response rate)



Questions Asked

- How much?
- What time frame?
- What subject?
- Future plans for the data?
- Where stored?
- Who is responsible for it?
- Still being used? Added to?
- Raw or processed?
- Sensitive, confidential or proprietary?
- Been scanned? Who has access?
- Been made machine readable?
- Been asked to share?
- Feelings about sharing?
- Written documentation exists?
- Ever reused or incorporated into a new study?
- Related to physical specimens that are held somewhere?

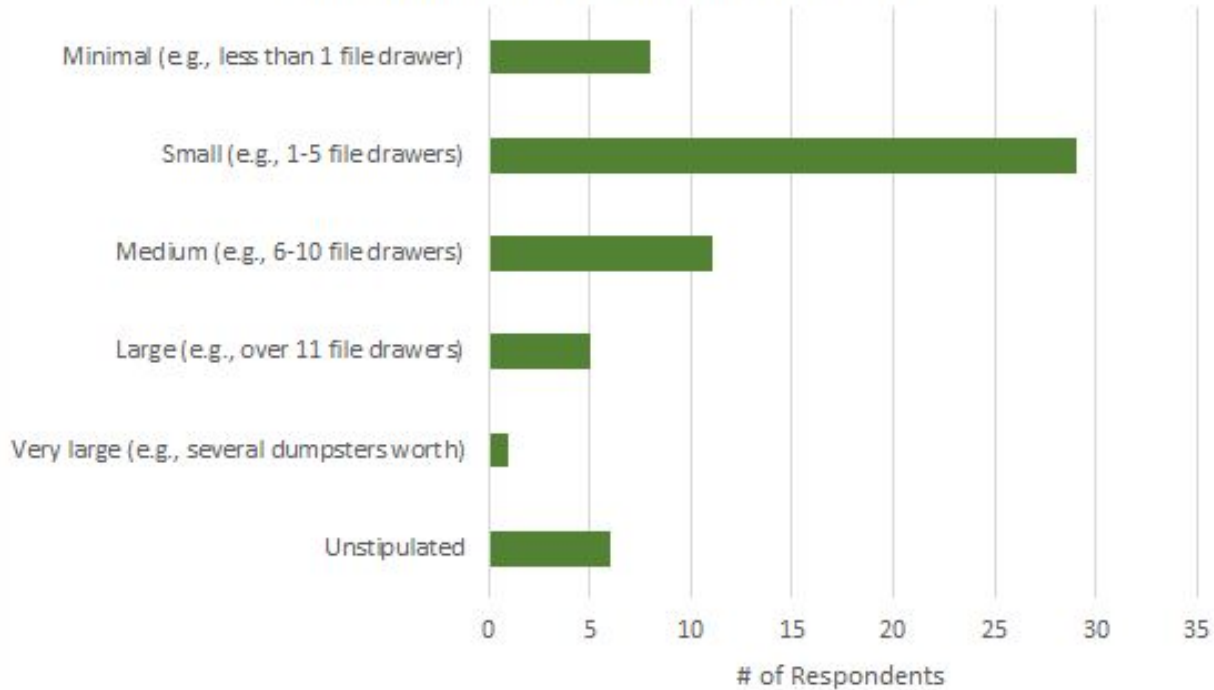


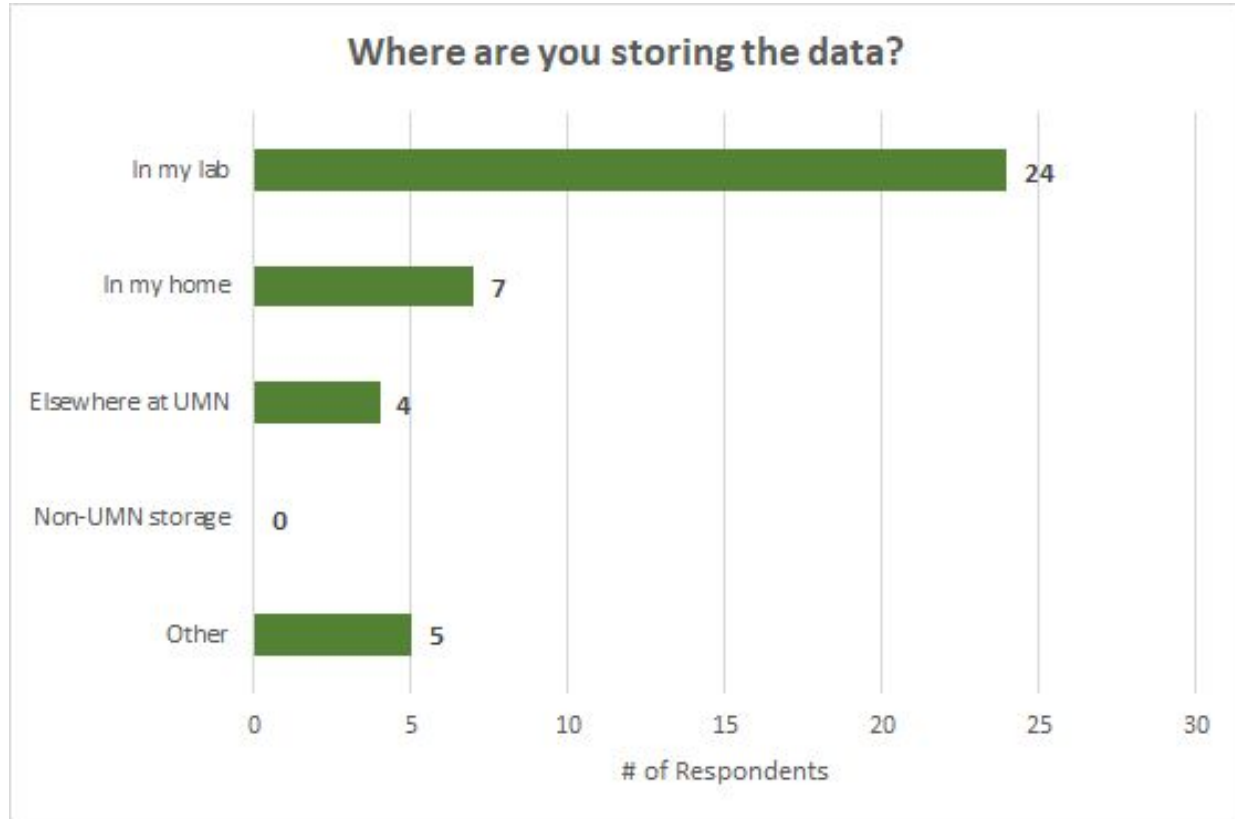
Definition

“**analog data** is defined as **non-digital data**, primarily in print, such as field books, lab notebooks, ledgers, data sheets, photographs, maps, drawings/sketches, slides, and so on. It **does not include physical specimens** (e.g., soil samples, tissue samples, herbarium specimens, organisms), but would include any analog data describing physical specimens”



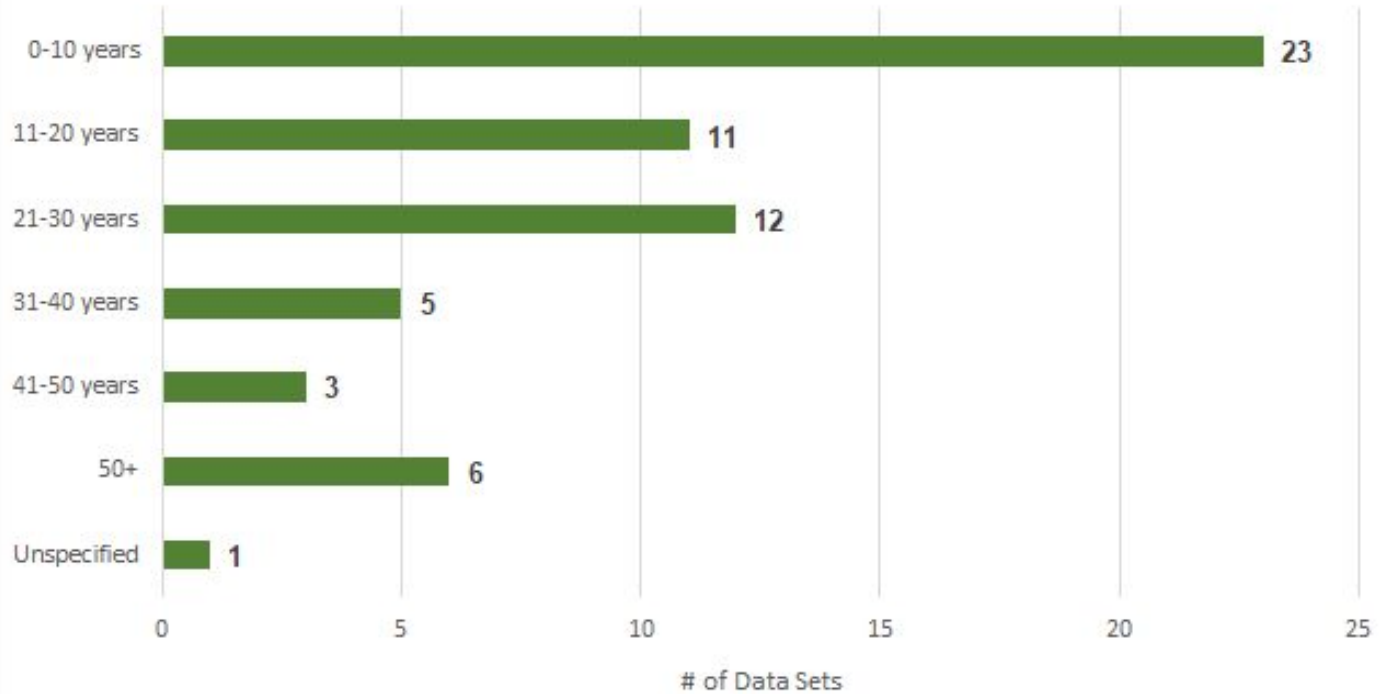
How much analog data do you have?







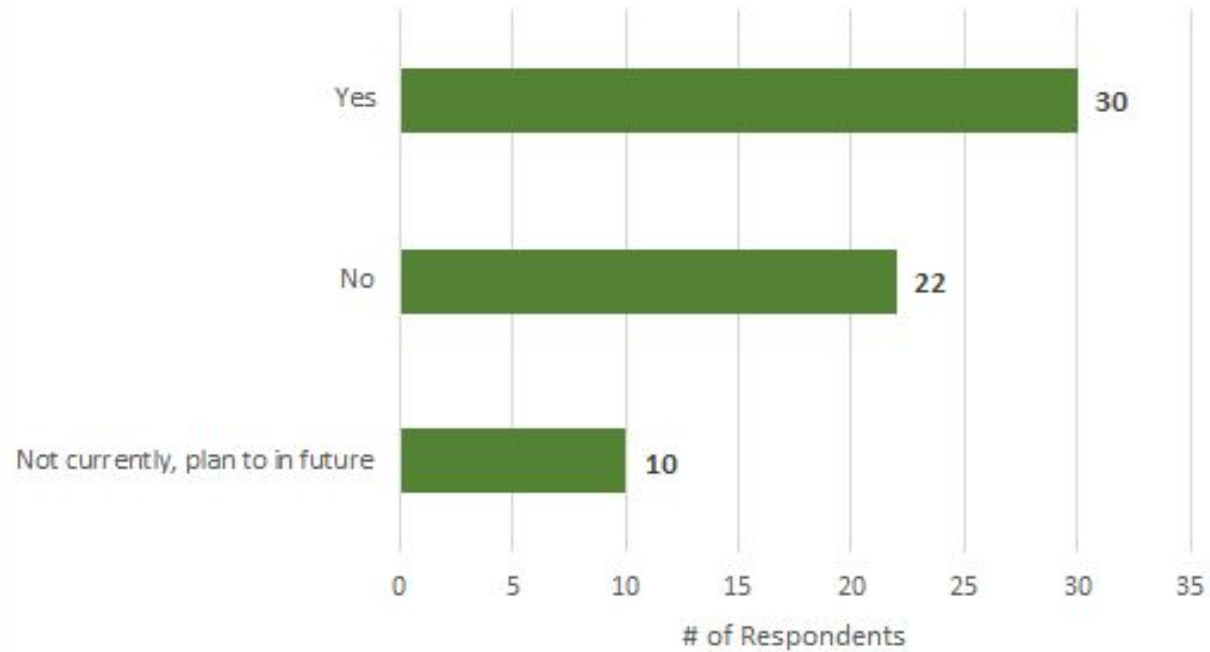
What time span does the analog data cover?



population distributions residential breeding
notebooks including water greenhouse
Some presence/absence dot ecology
Horticulture invasive studies Primarily
Museum Nutrition sensory Aquatic
quality production species walking apple Forestry
tourism cameras taxonomies crop stuides crops foreign management
sales RNA farmers etc growth Soil books
mussels business identification animals plant projects infrared birds
incubating wheat associated non near yield Bird dynamics natural multiple
Cropping sheets lab evolution travel experimental horticultural Life
Agronomy theory biology many notes past Landscape
selection theory resources Fruit Agriculture
information urban Science about spectroscopy systematics
related invertebrate various forest Statistics clean
inventory experiments various forest safety
avian

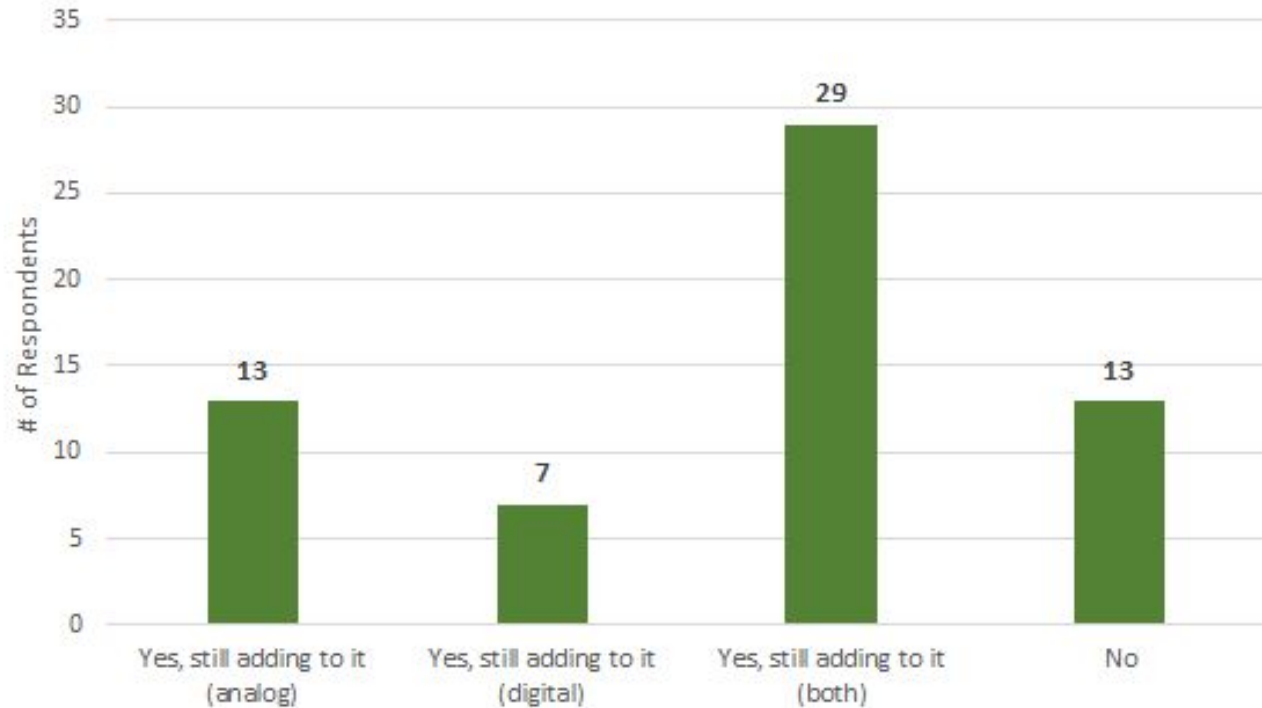


Are you using the data in analog form?



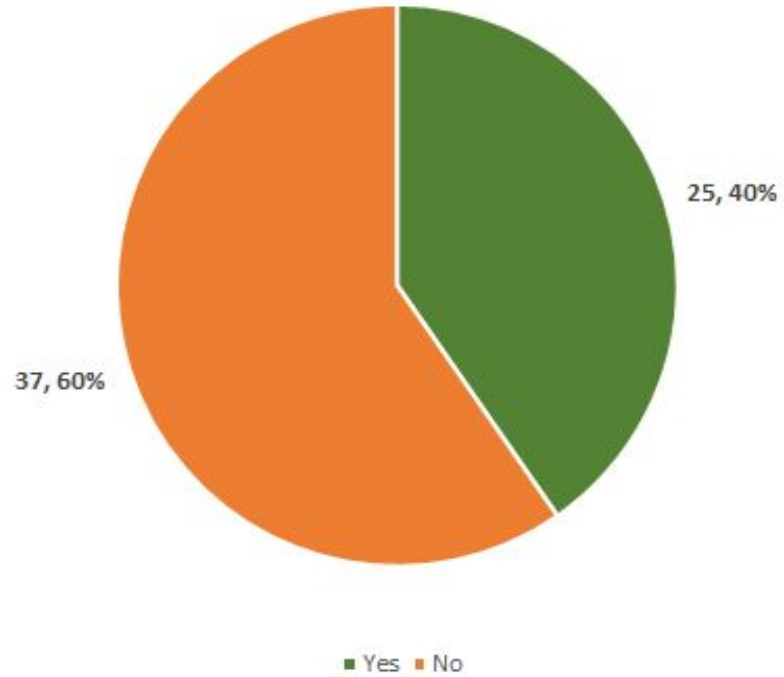


Are you still adding to the data?



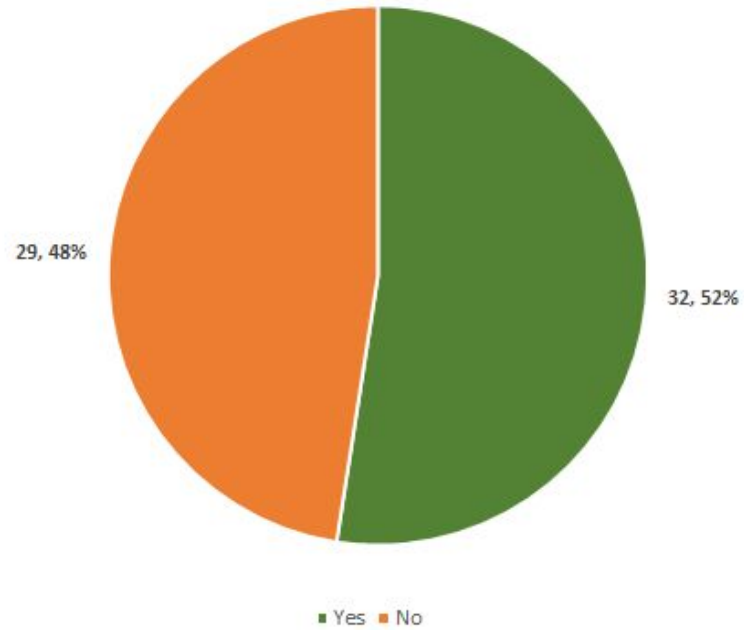


Has any of the analog data been scanned?





Have you made any of the analog data machine readable?





What happens when you leave?

I hope that it will be archived.

It will be destroyed.

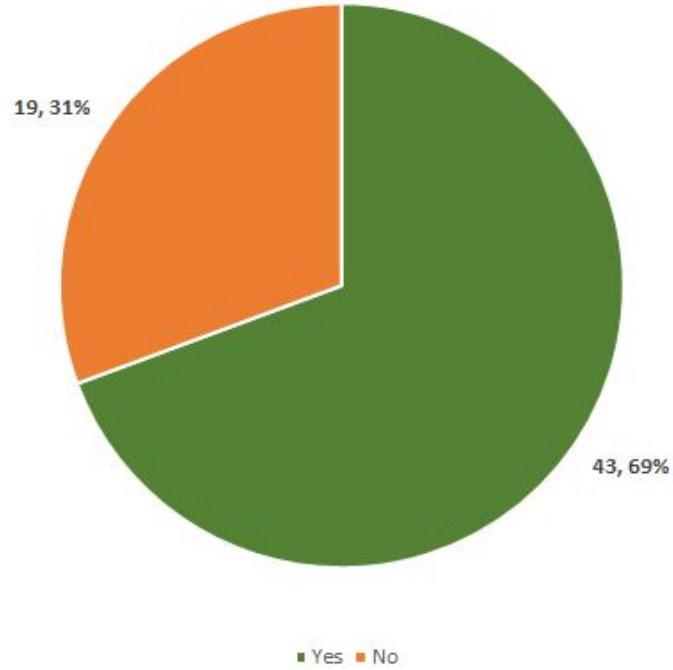
As the primary investigator, it will come with me to wherever I go next if I move to another institution. If I retire, I don't know what will happen to it.

It will stay physically associated with the collections.

I don't know.
We have bookshelves in the lab where we hold/archive the notebooks. I'm not sure if my successor will keep them.

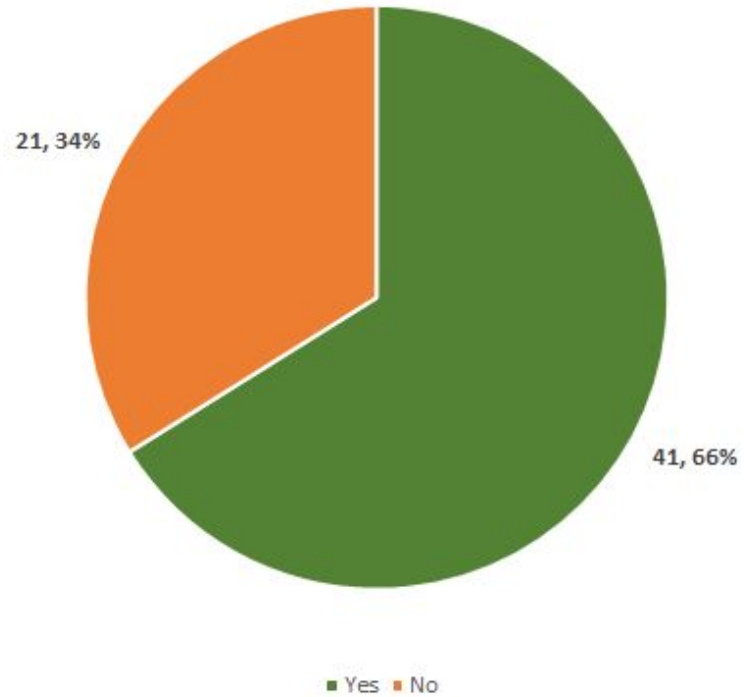


Is there written documentation about the data?



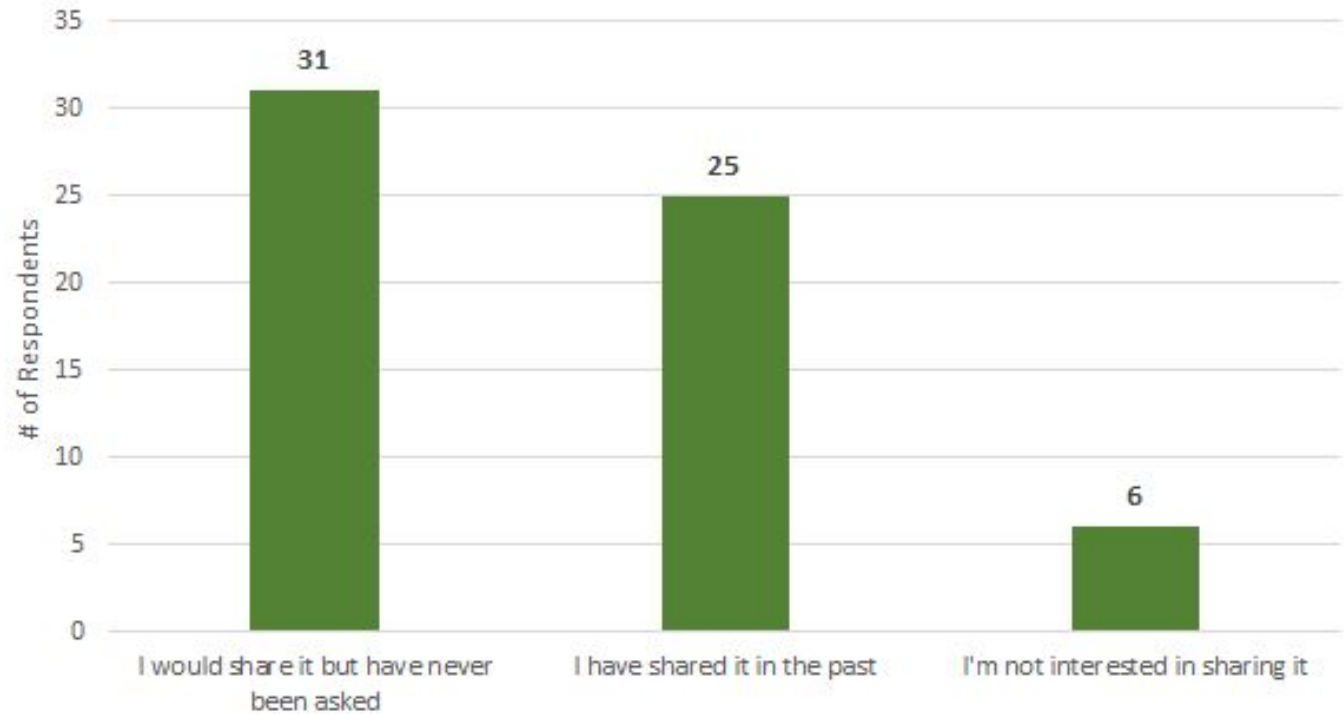


Has anyone in your lab/dept reused this data?



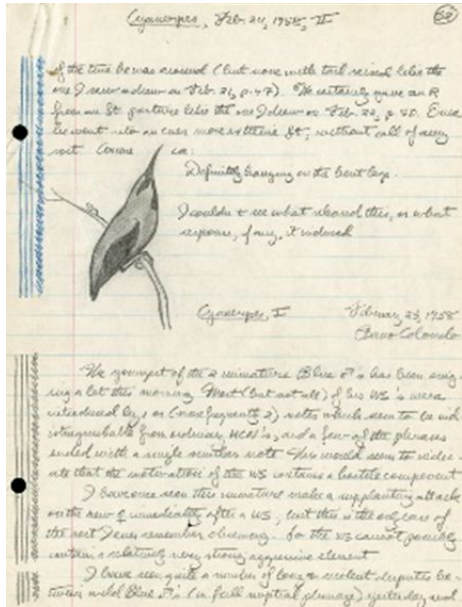


How do you feel about sharing your data with others?





What we've learned



- There is a lot of analog data on campus
- Data is at a risk of being lost
- Archives may not be the appropriate location to preserve analog data
- Researchers are still using (& adding to) their analog data
- Researchers want to share their data



Next steps

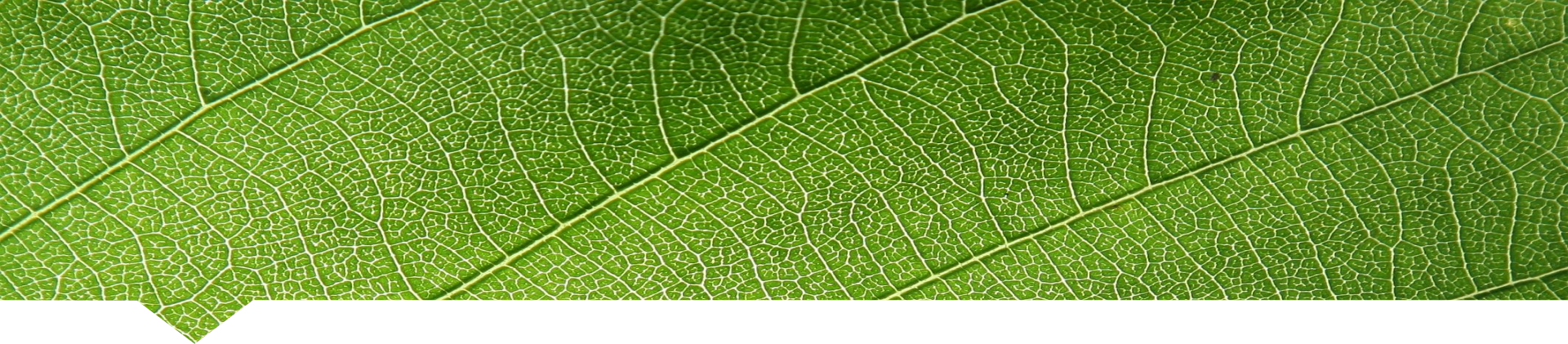
- Conduct interviews with faculty to:
 - discuss concerns they have about analog data
 - potential services they would want
 - thoughts about reuse
- Examine solutions to increase access/discoverability
- Raising awareness among scientists





This is not an isolated problem

- Our campus is not unique
- Analog data abounds but it is not increasing exponentially
- Is there interest from other research universities to look at their analog data?



Contact Us

Shannon L. Farrell
sfarrell@umn.edu

Julia A. Kelly
jkelly@umn.edu

Kristen L. Mastel
meye0539@umn.edu