← slides

Towards Categorical Metadata for Unreduced Climate Observations

Colton Grainger

University of Colorado Boulder

July 31, 2019

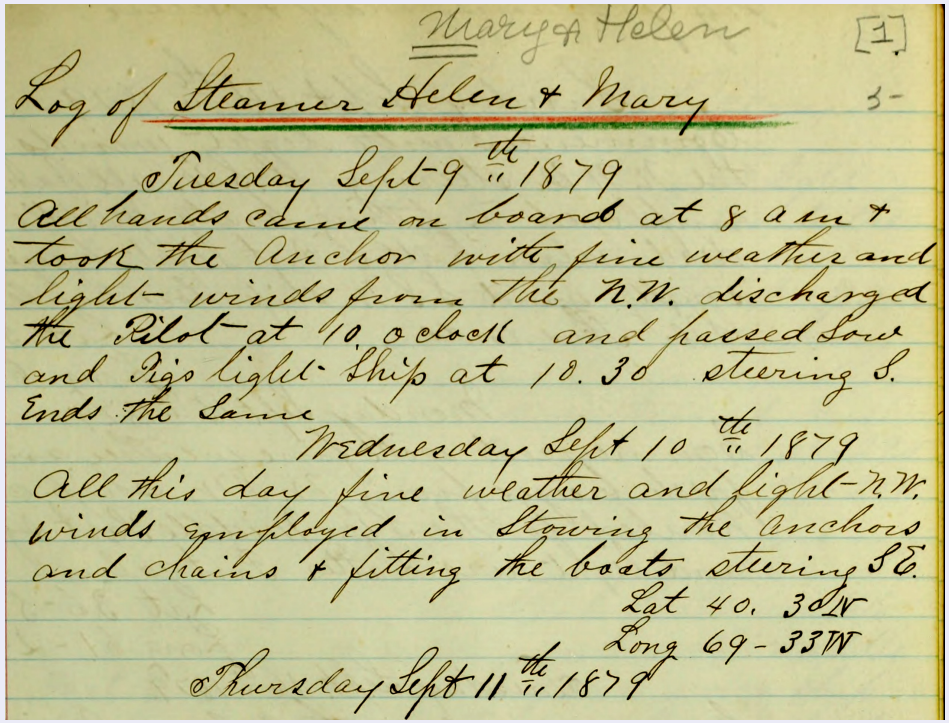
NCAR CISL

University of Colorado Boulder

NSF

unreduced data from ~100 Tb collection of images (ACRE)

marine logbooks 1870–1950



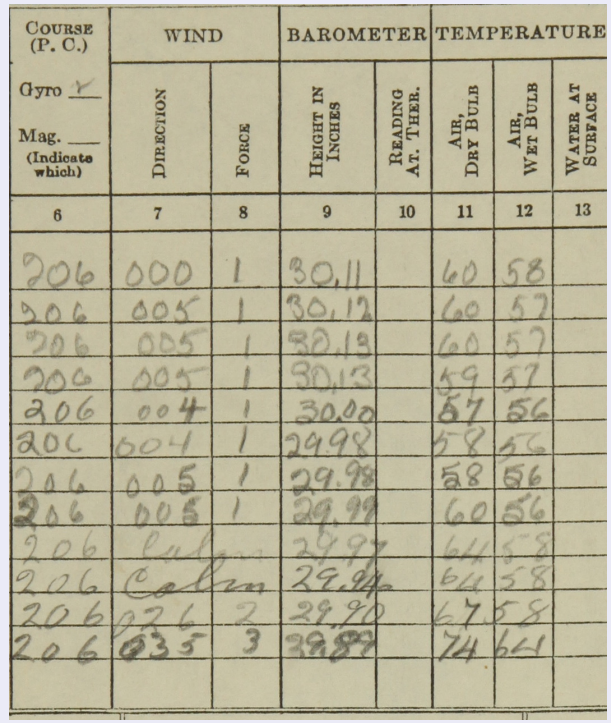
Mary & Helen [1]

Log of Steamer Helen & Mary 3-

Tuesday Sept 9th " 1879
All hands came on board at 8 am & took the Anchor with fine weather and light winds from the N.W. discharged the Pilot at 10 o clock and passed Sow and Pigs light Ship at 10.30 steering S. Ends the Same

Wednesday Sept 10th " 1879
All this day fine weather and light N.W. winds employed in Stowing the Anchors and chains & fitting the boats steering SE.
Lat 40. 30 N
Long 69 - 33 W

Thursday Sept 11th " 1879

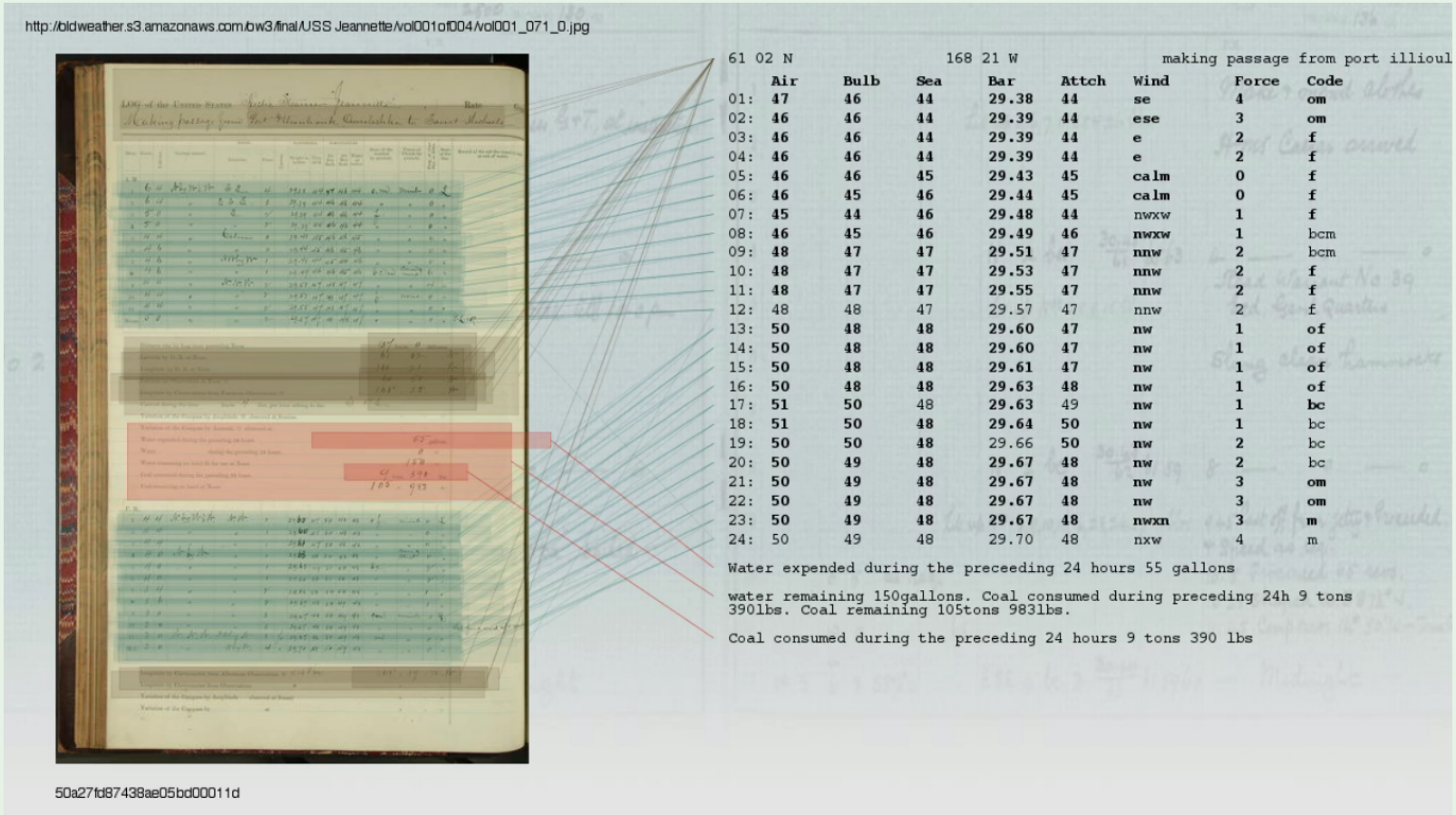


Course (P. C.) Gyro / Mag. (Indicate which)	Wind		Barometer		Temperature		
	Direction	Force	Height in Inches	Reading At. Ther.	Air, Dry Bulb	Air, Wet Bulb	Water at Surface
6	7	8	9	10	11	12	13
206	000	1	30.11		60	58	
206	005	1	30.12		60	57	
206	005	1	30.13		60	57	
206	005	1	30.13		59	57	
206	004	1	30.00		57	56	
206	004	1	29.98		58	56	
206	005	1	29.98		58	56	
206	005	1	29.99		60	56	
206	Calm		29.97		64	58	
206	Calm		29.94		64	58	
206	026	2	29.90		67	58	
206	035	3	29.87		74	64	

land stations 1870–1930

Indian Daily Weather Reports, Todd Folios, etc.

Example (reduction of 6Mb image to a 2Kb time series)



	Air	Bulb	Sea	Bar	Attch	Wind	Force	Code
01:	47	46	44	29.38	44	se	4	om
02:	46	46	44	29.39	44	ese	3	om
03:	46	46	44	29.39	44	e	2	f
04:	46	46	44	29.39	44	e	2	f
05:	46	46	45	29.43	45	calm	0	f
06:	46	45	46	29.44	45	calm	0	f
07:	45	44	46	29.48	44	nwxw	1	f
08:	46	45	46	29.49	46	nwxw	1	bcm
09:	48	47	47	29.51	47	nnw	2	bcm
10:	48	47	47	29.53	47	nnw	2	f
11:	48	47	47	29.55	47	nnw	2	f
12:	48	48	47	29.57	47	nnw	2	f
13:	50	48	48	29.60	47	nw	1	of
14:	50	48	48	29.60	47	nw	1	of
15:	50	48	48	29.61	47	nw	1	of
16:	50	48	48	29.63	48	nw	1	of
17:	51	50	48	29.63	49	nw	1	bc
18:	51	50	48	29.64	50	nw	1	bc
19:	50	50	48	29.66	50	nw	2	bc
20:	50	49	48	29.67	48	nw	2	bc
21:	50	49	48	29.67	48	nw	3	om
22:	50	49	48	29.67	48	nw	3	om
23:	50	49	48	29.67	48	nwxn	3	m
24:	50	49	48	29.70	48	nxw	4	m

Figure 1: 24 hours on the USS Jeannette (P. Brohan)

url

`twitter.com/NCAR_RDA/status/1144058111654711296`

qrcode

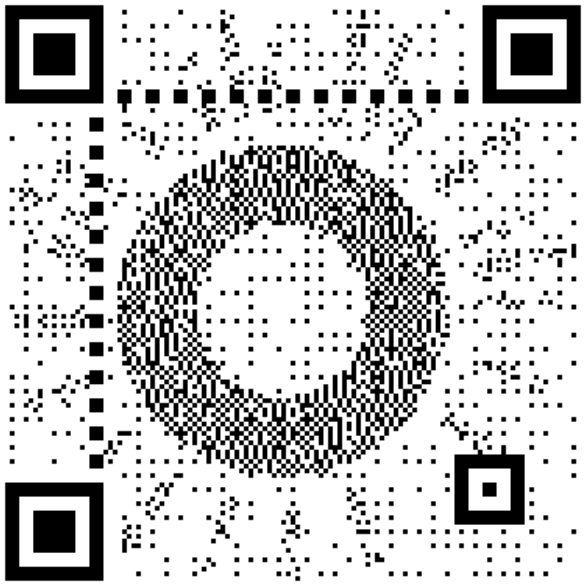


Figure 2: *CISL Seminar*, P. Brohan (2019)

Example

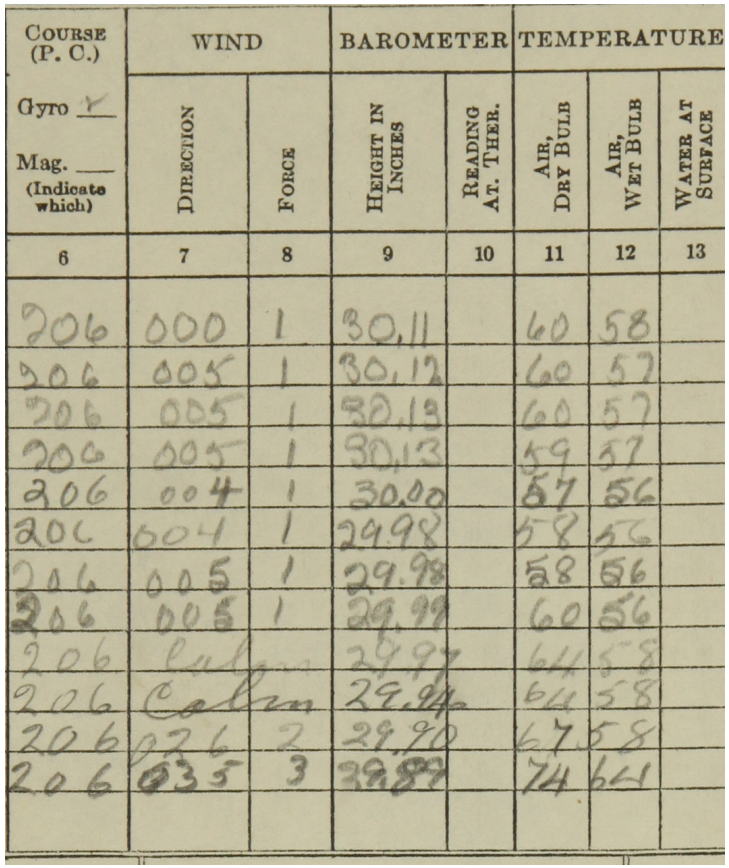
- 2^6 bytes $\sim$ 70 ASCII characters
- 2^{12} bytes $\sim 441 \times 441$ pixels
- a `qrcode` **reduces** to a `url`

```
$ exiftool brohan_qrcode.png
File Size : 2.7 kB
Image Size : 441x441
MIME Type : image/png
```

metadata

`{'key:value' for key in <schema>}`

what available metadata has low noise-to-signal ratio?



Course (P. C.) Gyro ✓ Mag. ___ (Indicate which)	Wind		Barometer		Temperature		
	Direction	Force	Height in Inches	Reading At. Ther.	Air, Dry Bulb	Air, Wet Bulb	Water at Surface
6	7	8	9	10	11	12	13
206	000	1	30.11		60	58	
206	005	1	30.12		60	57	
206	005	1	30.13		60	57	
206	005	1	30.13		59	57	
206	004	1	30.00		57	56	
206	004	1	29.98		58	56	
206	005	1	29.98		58	56	
206	005	1	29.99		60	56	
206	Calm		29.97		64	58	
206	Calm		29.94		64	58	
206	026	2	29.90		67	58	
206	035	3	29.87		74	64	

```
$ exiftool Idaho-BB-0034.JPG
File Size : 4.6 MB
Image Size : 3744x5616
MIME Type : image/jpeg
```

arc archive

doc document

img image

obs observation

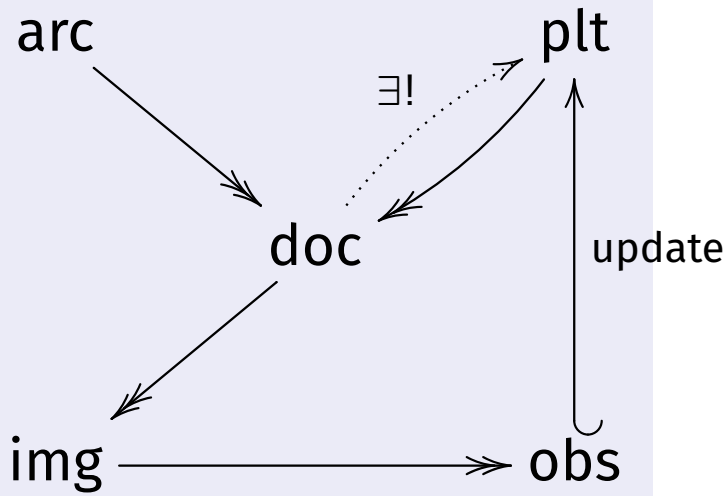
plt platform

Figure 3: May 1944, Idaho (BB-42)

- Roughly, an image in the category img approximates a time-series $\sigma\colon [t_0, t_1] \to S$, where S is the state space of meteorological variables.
- An observation is $\rho(t)$ evaluated at a time t.
- A platform in the category plt approximates a time-series $\lambda\colon [t_0, t_1] \to M$, where M is a differentiable manifold.

arc	archive
doc	document
img	image
obs	observation
plt	platform

dependencies



1. Gather images into (at least) **one repository**.
2. Establish a **common description framework** for image metadata.
3. Provide **bulk, programmatic access** to image subsets.

repo MySQL schema with Makefile. `assign_uuid()`.

framework Agnostic ingest scripts. Bundle metadata.

access `rda.ucar.edu/i/<uuid>`

repo	Conflict resolution for `assign-uuid()`.
framework	RELAX NG validator.
access	`rda.ucar.edu/i/<query>`

Figure 4: git repo