

# FUTURE-PROOFING YOUR DATA WITH



# SQUAD GOALS



mozilla

## Science Lab

Danielle Robinson, PhD

@daniellecrobins

**FUTURE-PROOFING  
YOUR DATA WITH**



**Structured!**

**Quantitative!**

**Universal Access to!**

**Data!**



Piecemeal data infrastructure

Lack of standards

Functional as long as nothing changes

Ummmm

...but then something changed

# Index of /pub/data

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 <a href="#">Parent Directory</a>		-	
 <a href="#">#rw-check</a>	08-Jun-2012 14:09	0	
 <a href="#">15min_precip-3260/</a>	17-Mar-2016 10:02	-	
 <a href="#">96300w60</a>	16-Apr-1997 15:24	5.5M	
 <a href="#">109020/</a>	07-May-2015 14:20	-	
 <a href="#">ASOS Station Photos/</a>	03-Apr-2014 15:00	-	
 <a href="#">EngineeringWeatherData_CDROM/</a>	21-Mar-2014 07:12	-	
 <a href="#">Impact/</a>	21-May-2015 12:41	-	
 <a href="#">Videoclip_50years_MCSS.wmv</a>	08-Sep-2014 06:43	388M	
 <a href="#">access.del/</a>	21-Mar-2014 12:38	-	
 <a href="#">aewc-v1/</a>	16-Sep-2014 15:17	-	
 <a href="#">airsea/</a>	19-Aug-2014 14:58	-	
 <a href="#">annualreports/</a>	20-Jun-2014 09:18	-	
 <a href="#">anomalie/</a>	06-Sep-2012 06:28	-	
 <a href="#">anomalies/</a>	27-May-2014 14:27	-	
 <a href="#">asos-fivemin/</a>	03-Feb-2017 07:00	-	
 <a href="#">asos-onemin/</a>	03-Feb-2017 07:23	-	
 <a href="#">blizzard/</a>	27-Jun-2014 12:56	-	
 <a href="#">ccd-data/</a>	26-Sep-2016 14:26	-	
 <a href="#">cdmp/</a>	02-Jul-2014 12:12	-	
 <a href="#">cdo/</a>	20-Jun-2016 10:10	-	
 <a href="#">cirs/</a>	18-Aug-2014 09:29	-	
 <a href="#">climgrid/</a>	19-Apr-2016 10:05	-	
 <a href="#">cmb/</a>	09-May-2016 11:45	-	

No machine readable metadata  
anywhere in this file structure

No agency pages link to this site

Only users know what/where

It's functional as long as nothing  
changes

https://www1.ncdc.noaa.gov/pub/data/



# Data on Internet be like

A large, messy stack of papers and folders on a wooden desk, illustrating the overwhelming volume and disorganization of data. The stack is composed of numerous yellowed, cream-colored papers and folders, some of which are spiral-bound. The papers are piled haphazardly, with some edges sticking out and others buried. The stack is resting on a light-colored wooden surface. In the background, a red object, possibly a bag or a piece of clothing, is partially visible.

No complete list of data, dataset locations, data volume

No guarantee of dataset backup

No way to assess changes over time

No way to ask questions of the national or global volume of data

Server Error

**404 - File or directory not found.**

The resource you are looking for might have been removed,

Server Error

**404 - File or directory not found.**

The resource you are looking for might have been removed,

Server Error

**404 - File or directory not found.**

The resource you are looking for might have been removed,

Server Error

**404 - File or directory not found.**

# Avoiding data disaster

“The internet is a terribly unstable way to keep information available”

– Laurie Allen

Assistant Director for Digital Scholarship

PENN Libraries

DataRefuge

Need wide adoption of **METADATA** standards

... and documentation/mirroring of data now

**What do we want?  
/ When do we  
want it?**



# What do we want? / When do we want it?



Wide adoption of **MACHINE-READABLE METADATA** standards

**DOCUMENTATION AND MIRRORING** of datasets



**Think Data, Think  
METADATA**



**SQUAD GOALS**

**Structured!**

**Quantitative!**

**Universal Access to!**

**Data!**



**Three things everyone loves...**

♥ **CATS**  
♥ **LASERS**  
♥ **SPACE**

**And NASA has all three.**

#earthrightnow

Our CATS – the Cloud-Aerosol Transport System – is a small instrument ready to do big science on the space station after launch on January 6. CATS will use lasers to study clouds and tiny particles like dust and air pollution.



**EARTH *RIGHT* NOW**

[www.nasa.gov/earthrightnow](http://www.nasa.gov/earthrightnow)

Come to Data Reuse session tomorrow to learn more about METADATA and the CATS dataset!



