UC3
UC Curation Center

University of California
CDL
California Digital Library

Services for Big Data… and the Long Tail

From Flickr by rahen z

Patricia Cruse
Stephen Abrams
Carly Strasser
Perry Willett

University of California Curation Center
California Digital Library
http://www.cdlib.org/uc3

UC3
UC Curation Center

University of California
CDL
California Digital Library

Big Data

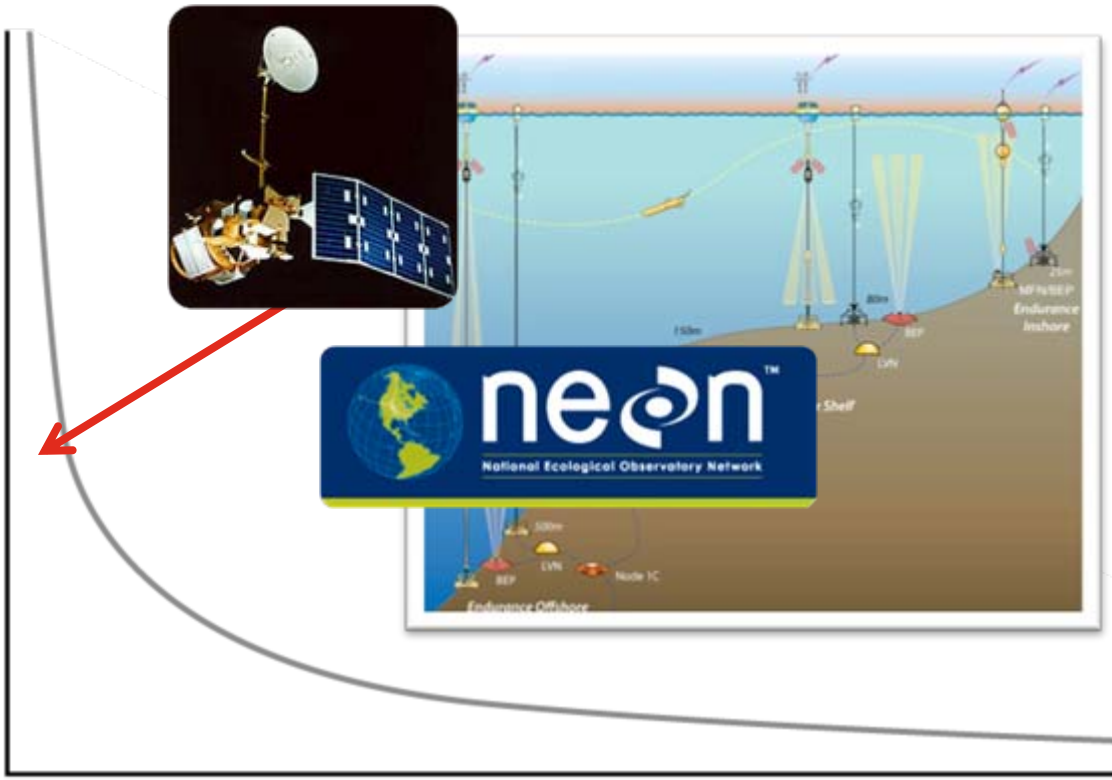
The Little Guys?

From Flickr by jason tinder

UC3
UC Curation Center

University of California
CDL
California Digital Library

The Long Tail



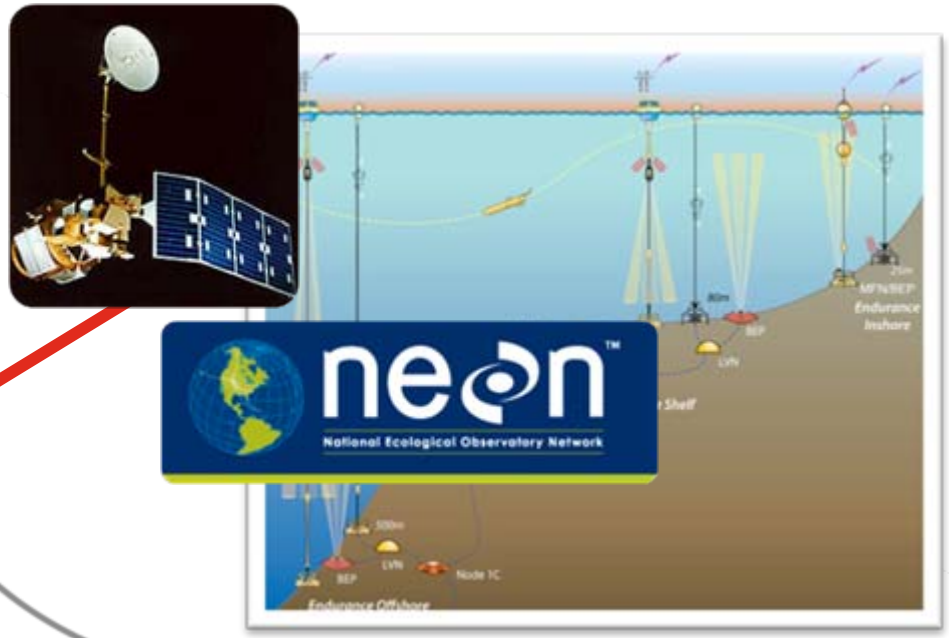
Hampton et al. *Forthcoming* 2012

UC3
UC Curation Center

University of California
CDL
California Digital Library

The Long Tail

Size of dataset



~~# datasets~~

 # researchers

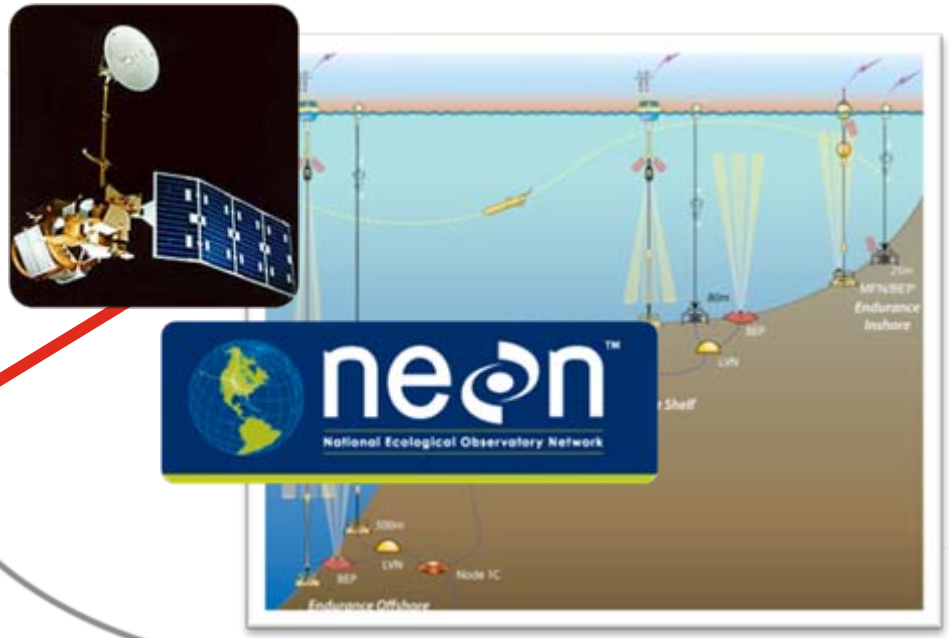
Hampton et al. *Forthcoming* 2012

UC3
UC Curation Center

University of California
CDL
California Digital Library

The Long Tail

Size of dataset



~~# datasets~~

~~# researchers~~

 # grants

Hampton et al. *Forthcoming* 2012

UC3
UC Curation Center

University of California
CDL
California Digital Library

The Long Tail

Size of
~~dataset~~
grant ($)

~~# datasets~~

~~# researchers~~
grants

Hampton et al. *Forthcoming* 2012

UC3
UC Curation Center

University of California
CDL
California Digital Library

UGLY TRUTH

Many (most?) researchers…

are not taught data management

don't know what metadata are

can't name data centers or repositories

don't share data publicly or store it in an archive

aren't convinced they should share data

UC3
UC Curation Center

University of California
CDL
California Digital Library

The Fallout: Where data end up

Recreated from Klump et al. 2006

UC3
UC Curation Center

University of California
CDL
California Digital Library

Barriers

- Cost: time, personnel, software & hardware
- Confusion about standards, how to start
- Disparate datasets
- Lack of training
- Fear of lost rights/benefits
- No incentives

From Flickr by Christina Ann VanMeter

From Flickr by bthomso

UC3
UC Curation Center

University of California
CDL
California Digital Library

Two Services to Mitigate Long Tail Challenges

1. DMPTool
   Build your Data Management Plan

2. DataUp

UC3
UC Curation Center

University of California
CDL
California Digital Library

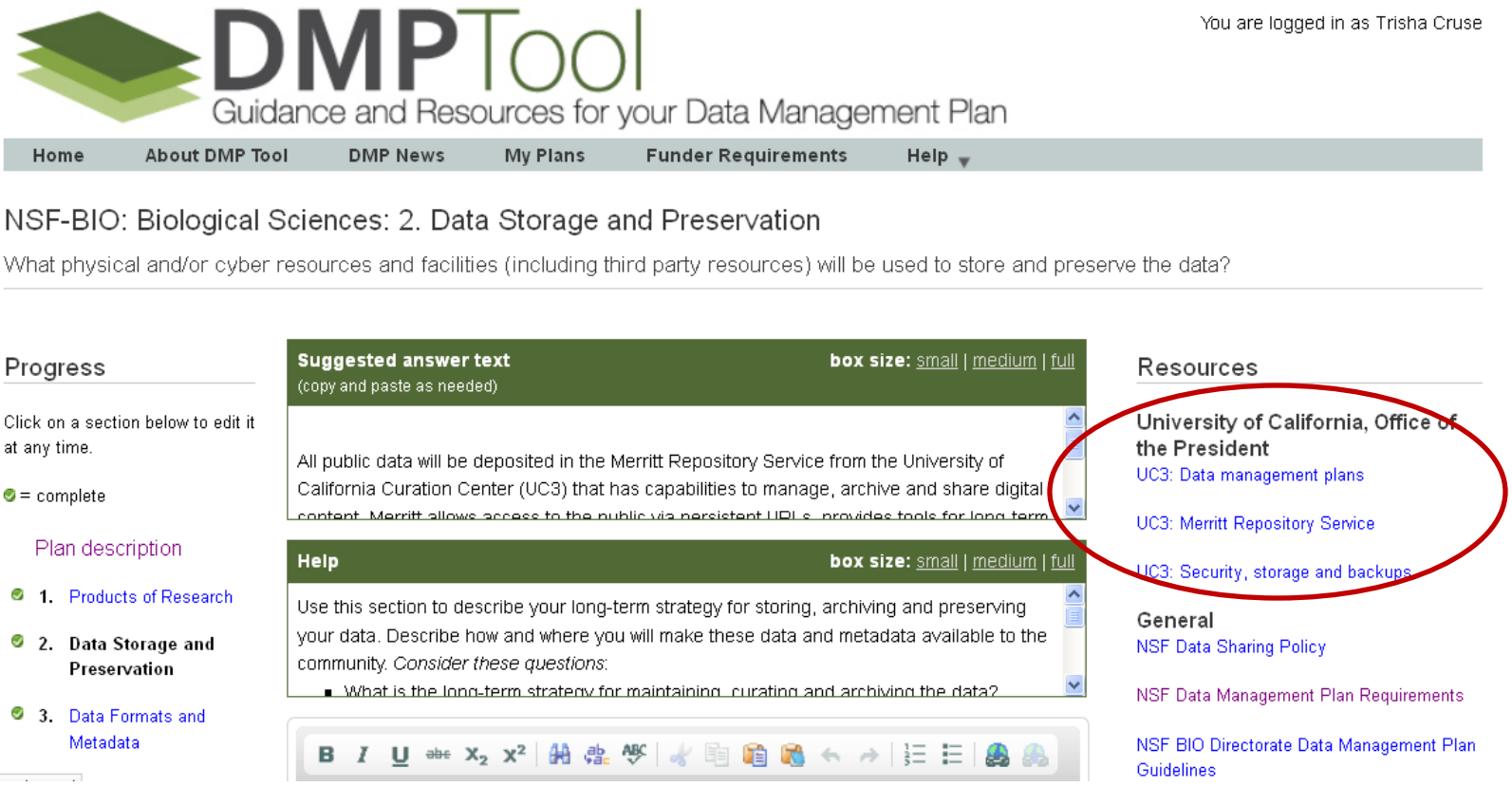
Data Management Planning – DMPTool

Meeting funding agencies data management plan requirements

- Seven institutions
- Helps researchers meet funding agency requirements
- Guide researchers through the process of creating a data management plan
- Connects researchers to local resources
- Open to everyone

http://dmptool.org

Customize resources for your institution



Data Management Plan Tool | https://dmp.cdlib.org

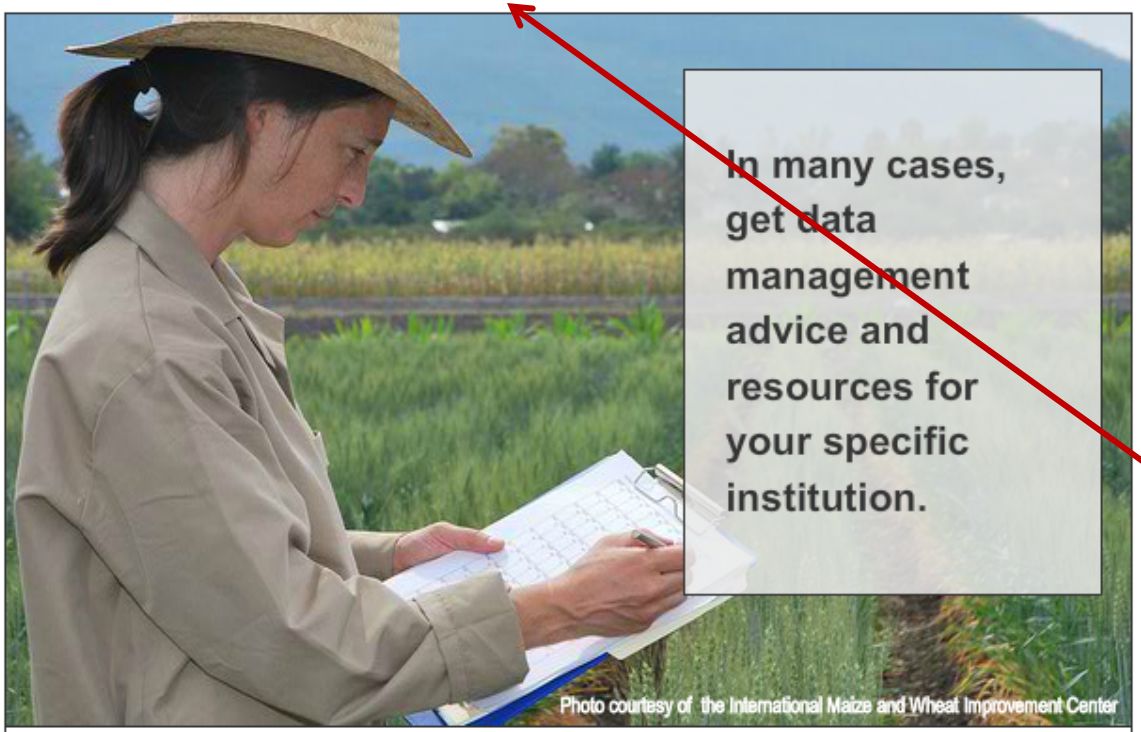
Bookmarks | General Interest | Other bookmarks

Contact Us | Get Started | Login

DMPTool

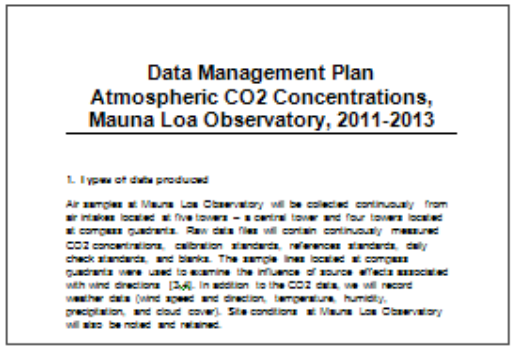
Guidance and Resources for your Data Management Plan

Home | About DMP Tool | DMP News | My Plans | Funder Requirements | Help



The DMP Tool allows you to: 1 2 3 4

Get Started!


Data Management Plan
Atmospheric CO2 Concentrations, Mauna Loa Observatory, 2011-2013

1. Types of data produced

Air samples at Mauna Loa Observatory will be collected continuously from air intakes located at five towers – a central tower and four towers located at compass quadrants. Raw data files will contain continuously measured CO2 concentrations, calibration standards, references standards, daily check standards, and blanks. The sample lines located at compass quadrants were used to examine the influence of source effects associated with wind directions (July). In addition to the CO2 data, we will record weather data (wind speed and direction, temperature, humidity, precipitation, and cloud cover). Site conditions at Mauna Loa Observatory will also be noted and retained.

See a plan created with the DMP Tool

Recent DMP News

DMPTool workshop at the DLF Fall Forum

DMPTool demo: Wed Oct 19

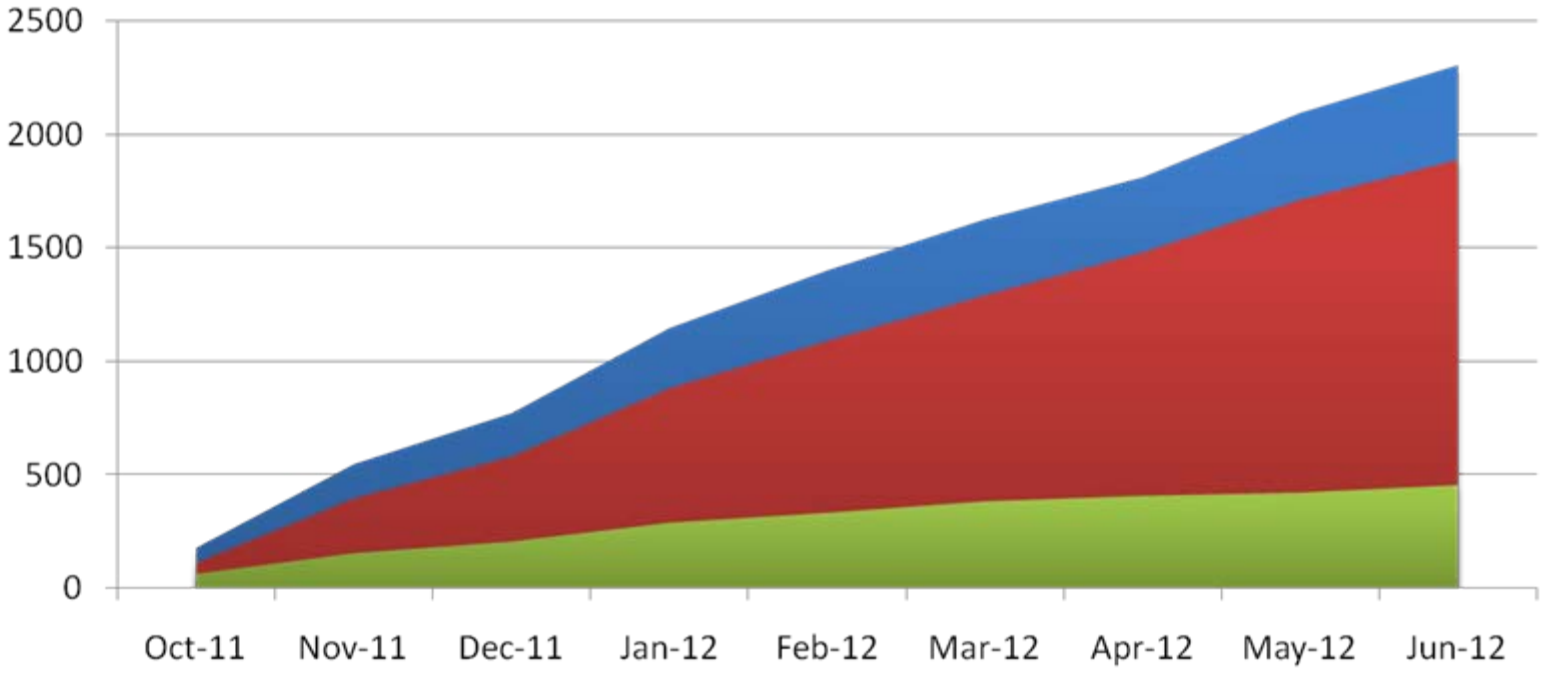
Importance of Data Management Education

More news >

UC3
UC Curation Center

University of California
CDL
California Digital Library

Growth rate since launch in Oct 2011



	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12
Unique Users	180	548	772	1146	1399	1624	1809	2092	2302
Plans	114	398	582	882	1090	1291	1479	1711	1887
Institutions	65	155	206	288	331	381	405	419	452

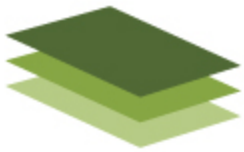
DMPTool
Build your Data Management Plan

UC3
UC Curation Center

University of California
CDL
California Digital Library

New additions

- Launched an open organization to build a community around the data management issues!
- On the horizon
  - Adding new functionality
  - Tighter community integration and educational components
  - Integration with other systems and initiatives
  - Routing of DMPs to support personnel
  - Add analytics



DMPTool
Build your Data Management Plan

http://dmptool.org

UC3
UC Curation Center

University of California
CDL
California Digital Library

DataUp

Open Source Tool

Add-in & Web Application

Facilitate sharing, preserving, using, citing, re-using tabular data

?

GORDON AND BETTY MOORE FOUNDATION

Microsoft Research

UC3
UC Curation Center

DataUp

University of California
CDL
California Digital Library

Features

Best practices check

Generate metadata

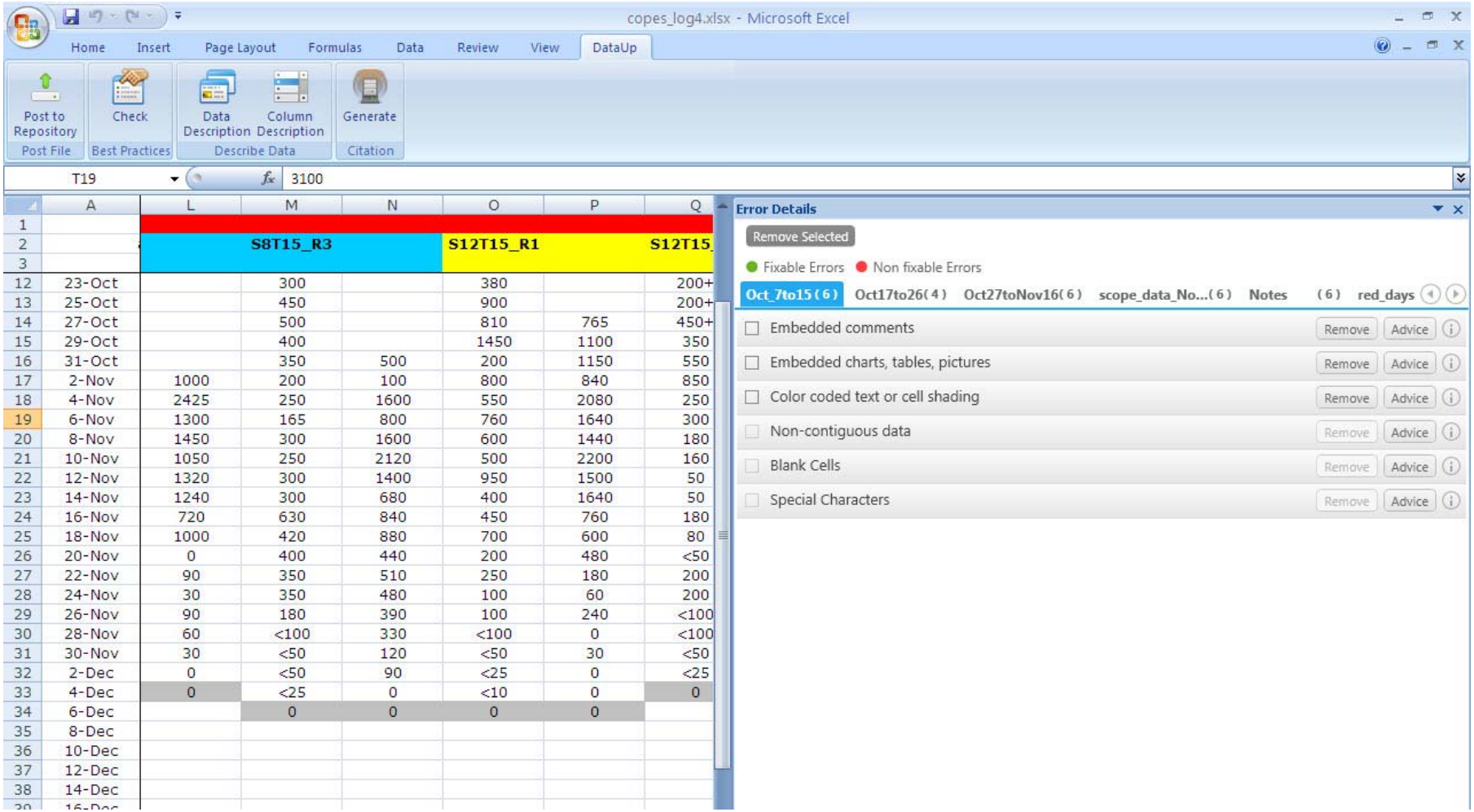
Generate citation

Post data to repository

UC3
UC Curation Center

University of California
CDL
California Digital Library

Best Practices Check

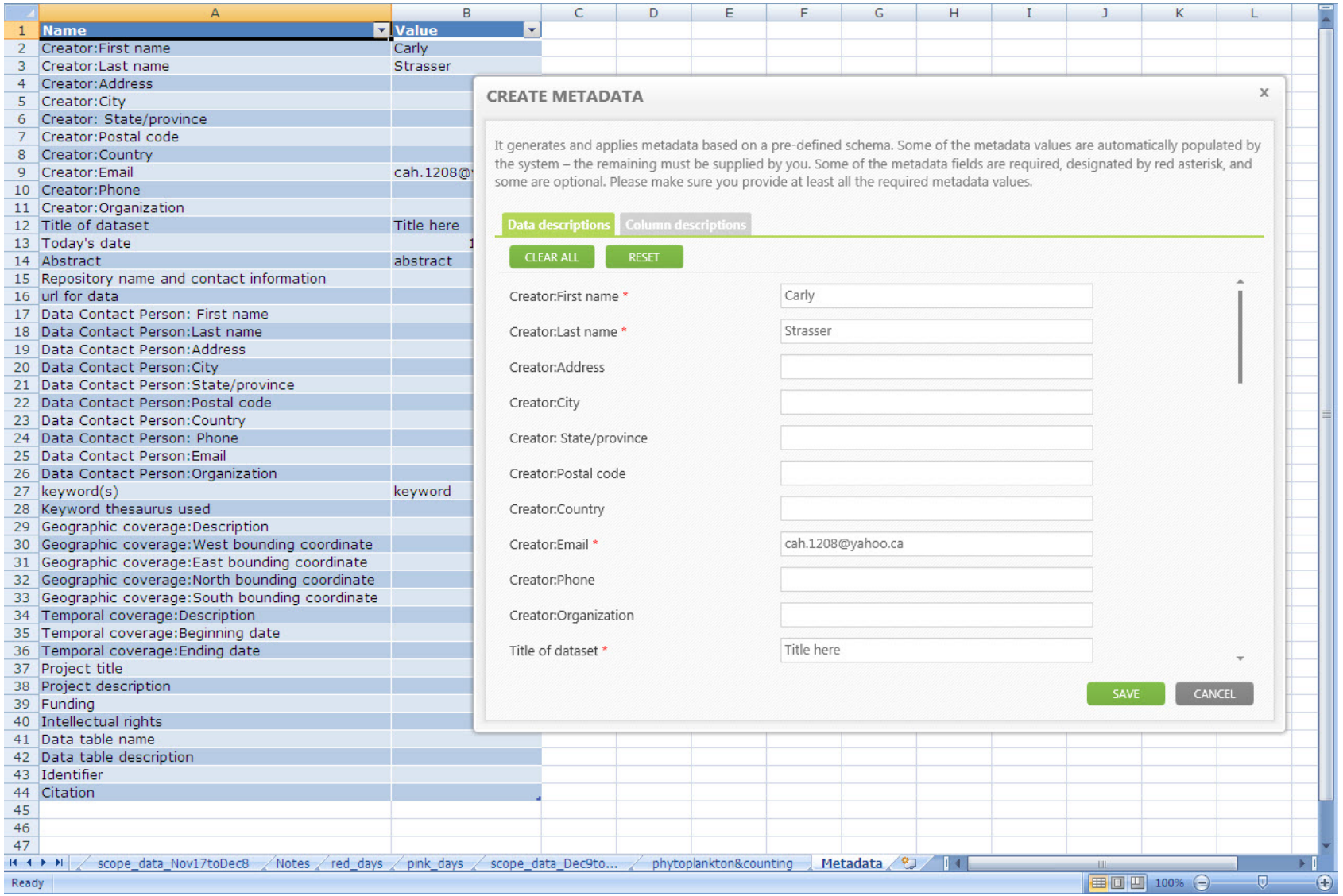


	A	L	M	N	O	P	Q
1							
2			S8T15_R3		S12T15_R1		S12T15_
3							
12	23-Oct		300		380		200+
13	25-Oct		450		900		200+
14	27-Oct		500		810	765	450+
15	29-Oct		400		1450	1100	350
16	31-Oct		350	500	200	1150	550
17	2-Nov	1000	200	100	800	840	850
18	4-Nov	2425	250	1600	550	2080	250
19	6-Nov	1300	165	800	760	1640	300
20	8-Nov	1450	300	1600	600	1440	180
21	10-Nov	1050	250	2120	500	2200	160
22	12-Nov	1320	300	1400	950	1500	50
23	14-Nov	1240	300	680	400	1640	50
24	16-Nov	720	630	840	450	760	180
25	18-Nov	1000	420	880	700	600	80
26	20-Nov	0	400	440	200	480	<50
27	22-Nov	90	350	510	250	180	200
28	24-Nov	30	350	480	100	60	200
29	26-Nov	90	180	390	100	240	<100
30	28-Nov	60	<100	330	<100	0	<100
31	30-Nov	30	<50	120	<50	30	<50
32	2-Dec	0	<50	90	<25	0	<25
33	4-Dec	0	<25	0	<10	0	0
34	6-Dec		0	0	0	0	
35	8-Dec						
36	10-Dec						
37	12-Dec						
38	14-Dec						

UC3
UC Curation Center

University of California
CDL
California Digital Library

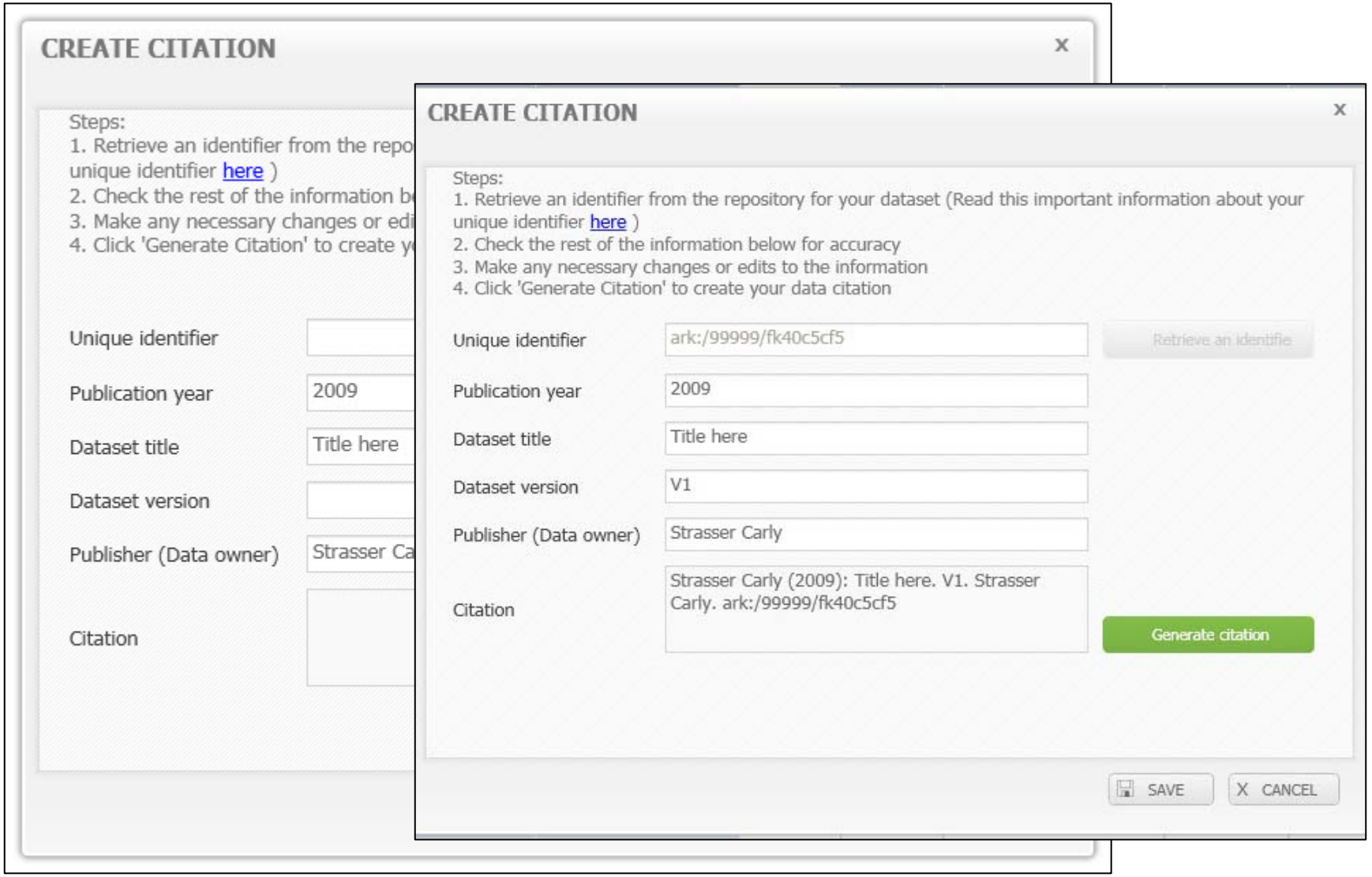
Generate Metadata



UC3
UC Curation Center

University of California
CDL
California Digital Library

Create a Data Citation



UC3
UC Curation Center

University of California
CDL
California Digital Library

Upload to a Repository

UC3
UC Curation Center

University of California
CDL
California Digital Library

Meeting the Long Tail Challenge

- ✓ Engage with the research community early in the lifecycle?
- ✓ Deploy simple and flexible infrastructures and services that can be used in diverse ways
- ✓ Start small: provide simple solutions that build up to solve more complex problems
- ✓ Collaborate now more than ever!

UC3
UC Curation Center

University of California
CDL
California Digital Library

UC3's Service Landscape

UC Curation Center
http://www.cdlib.org/uc3
uc3@ucop.edu

Stephen Abrams
Patricia Cruse
Scott Fisher
Erik Hetzner
Greg Janée
John Kunze
David Loy
Mark Reyes
Joan Starr
Tracy Seneca
Carly Strasser
Marisa Strong
Adrian Turner
Perry Willett

DMPTool
Guidance and Resources for your Data Management Plan

UC3 DataUp

UC3 EZID

UC3 Merritt

UC3 WAS