Preserving the agricultural data story at the Ontario Agricultural College

IASSIST 2018

MICHELLE EDWARDS

CAROL PERRY

Objectives

Demonstrate approaches used in teaching RDM to grads, research staff and faculty working in agricultural and environmental research

Ontario Agricultural College (OAC)

Department of Animal Biosciences

Department of Food, Agricultural and Resource Economics

Department of Food Science

Department of Plant Agriculture

Ridgetown Regional Campus
◦ 15 Research Stations

School of Environmental Design and Rural Development
◦ Landscape Architecture
◦ Rural Planning and Development
◦ Capacity Development and Extension
◦ Rural Studies

School of Environmental Sciences
◦ Earth & Atmospheric Sciences
◦ Ecosystem Science and Biodiversity
◦ Plant and Environmental Health

OAC Centres and Institutes

Aquaculture Centre

Arrell Food Institute

Bioproducts Discovery and Development Centre (BDDC)

Campbell Centre for the Study of Animal Welfare (CCSAW)

Canadian Research Institute for Food Safety (CRIFS)

Centre for Agricultural Renewable Energy and Sustainability (CARES)

Centre for Genetic Improvement of Livestock (CGIL)

Centre for Land and Water Stewardship

Centre for Nutrition Modelling

Controlled Environment Systems Research Facility (CESRF)

Gosling Research Institute for Plant Preservation

Guelph Centre for Urban Organic Farming

Guelph Turfgrass Institute

Honey Bee Research Centre

Institute for the Advanced Study of Food and Agricultural Policy

OMAFRA – UG Partnership

2018 Renewed partnership

University of Guelph - Ontario Ministry of Agriculture, Food & Rural Affairs (Provincial)

$713 million 10 year agreement

$71.3 million per year

Research themes
◦ Agri-Food & Rural Policy
◦ Bioeconomy – Industrial Uses
◦ Emergency Management
◦ Environmental Sustainability
◦ Food for Health
◦ Products & Value Chains
◦ Production systems – Plants and Animals

Other major funding

Food from Thought Research Project - $76.6 Million over 7 years

- Federal funding – Canada First Research Excellence Fund (CFREF)

Other sources: $32 Million in 2016-2017

OAC Research Scope

Extremely broad college research scope

- ◦ Social science data – SEDRD
- ◦ Health science data – CRIFS, ABSC, collaborative research with OVC departments
- ◦ Environmental data – SES
- ◦ Agriculture science data – PlantAg

Our goal is for college researchers to incorporate RDM into their daily research activities

Our agenda

Help OAC faculty, staff and students learn about
◦ RDM best practices
◦ Metadata vital for data preservation, discovery and reuse

Highlight the importance of documentation and preservation for the future
◦ Teaching 'metadata' without scaring them off

Researchers just want their data safe, secure, and some want it searchable

Workshop series – Fall 2017, Winter 2018

Scenario based training – focused on end product
◦ 4 x 2 hour sessions
◦ I session/week

Working through research data life cycle
◦ Applying RDM best practices
◦ Gathering metadata throughout the process
◦ Talking "their" language

Goals:
◦ Completed DMP
◦ Completed metadata template for data deposit

What is RDM?

A sound strategy and best practices used to………

- ◦ Organize
- ◦ Document
- ◦ Store
- ◦ Analyze
- ◦ Secure
- ◦ Preserve/Share/Reuse

………Your data

Week 1 – Organize & document

Overview of RDM & DMPs
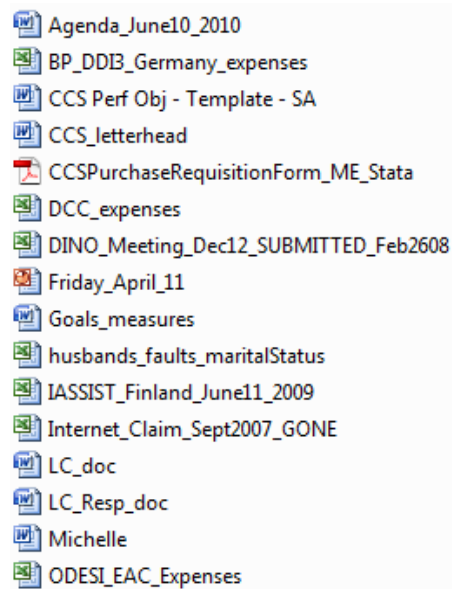
Introduce best practices to:
- ◦ structure research project data
- ◦ gather information
- ◦ collect data

Exercises
- ◦ Creating project directory structure based on scenario
- ◦ Exploring file naming conventions

Week 1 Exercise: Organizing your project files

DOES THIS LOOK FAMILIAR?

- Agenda_June10_2010
- BP_DDI3_Germany_expenses
- CCS Perf Obj - Template - SA
- CCS_letterhead
- CCSPurchaseRequisitionForm_ME_Stata
- DCC_expenses
- DINO_Meeting_Dec12_SUBMITTED_Feb2608
- Friday_April_11
- Goals_measures
- husbands_faults_maritalStatus
- IASSIST_Finland_June11_2009
- Internet_Claim_Sept2007_GONE
- LC_doc
- LC_Resp_doc
- Michelle
- ODESI_EAC_Expenses

OR DOES THIS LOOK FAMILIAR?

PC > Documents > Workshops > SAS > W18 >

Name	Date modified	Type	Size
20180118	2018-01-25 4:33 PM	File folder	
20180201	2018-02-05 8:55 A...	File folder	
20180215	2018-02-23 8:09 PM	File folder	
20180308	2018-03-12 10:44 ...	File folder	
Ridgetown_20180227	2018-02-23 3:33 PM	File folder	
Ridgetown_20180313	2018-03-12 10:15 ...	File folder	

Week 2 – Store & analyze

Review statistical packages

Review use of Excel in analysis

Discuss storage & backup practices

Exercises:
◦ Data collection practices
◦ Variable naming & labelling

Week 2 - Exercise

RDM: Best Practices for entering your Research Data using Excel
Dataset A:

Year: 2017

Measures taken: Starting weight of the horse (kg), the amount of feed consumed over the weeks on trial (kg), and weights during the trial (kg).

Natasha Figueroa, owner of ABC11111, NPQ13131, and AFS45456, made $312,400 in 2016 and spent approximately $7,000 in feed.

Start date of 2017 Trial: May 15, 2017

Horse ID	Starting Weight	Feed Consumed Week 1	Weight Week 1	Feed Consumed Week 2	Weight Week 2
ABC11111	1160	66.7	1159	69.7	1154
NPQ13131	896	79.4	894	84.4	903
AFS45456	976	72.3	989	.	.

Week 3 – Secure, preserve, share & reuse

Introduce best practices to preserve research data legacy

- DATA ACCESS, DMP and the Practicality of it all!

Pull together all of the information collected throughout the project

Prepare codebooks, user guides, scripts etc. for sharing

Prepare data files for deposit

Review repository options

Exercises:
◦ Basics of anonymization
◦ Filling in metadata template
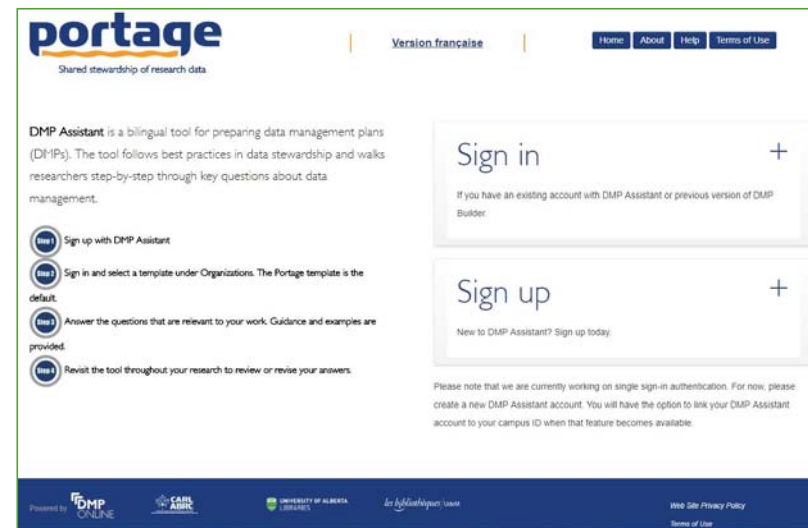
Photo by Adam Cohn. Used under CC-BY-NC-ND license from:
https://www.flickr.com/photos/adamcohn/34578872301/sizes/l

Week 3 - Exercise

	A	B	C	D	E	F	G
1	Farm ID	Horse Registration Number	Owner	Owner Age	Owner Gender	Owner Income 2015	Owner Feed Expenses
2	101	ABC11111	Natasha Figueroa	73	2	$289,100.00	$ 6,879.00
3	101	ABC12121	Faye Lowe	69	2	$256,900.00	$ 6,056.00
4	101	ABC13131	Jo Bass	56	1	$398,700.00	$ 4,735.00
5	102	DEF11111	Gregg Richardson	65	1	$587,400.00	$ 4,330.00
6	102	DEF12121	Tiffany Simmons	35	2	$ 87,600.00	$ 7,179.00
7	102	DEF13131	Rudy Mckenzie	48	1	$197,600.00	$ 4,204.00
8	103	GHJ11111	Gregg Richardson	65	1	$587,400.00	$ 7,240.00
9	103	GHJ12121	Max Gill	64	1	$187,300.00	$ 7,844.00
10	103	GHJ13131	Monique Chandler	75	2	$314,500.00	$ 5,517.00

Week 4 - Review

Review workshop content

Exercises:
- Creating the Data Management Plan using DMP Assistant
  - https://assistant.portagenetwork.ca/

- Complete the metadata template for data deposit
  - 19 required elements

Metadata element	Description	Fill in information as applicable
Title	Full title of the dataset	
Author (s)	Person, corporate body, or agency responsible for the work's intellectual content. Include names, affiliations, and email addresses (e.g. [NAME], [DEPARTMENT/SCHOOL], [COLLEGE], [UNIVERSITY].	
Production Date	The date the final version of the dataset being deposited in the repository was created. This is the date that will be used in the dataset citation.	
Funding agency	The source(s) of funds for the production of the work	
Time period covered	Time period covered by the data	
Date of collection	The date(s) when the data were actually collected	
Keywords	Words or phrases that describes the data collection's content (please provide at least 5 keywords)	
Description	Summary describing the purpose and scope of the data collection, and what questions the investigators attempted to answer.	

AGRI-ENVIRONMENTAL RESEARCH DATA REPOSITORY

Agri-environmental research data repository Dataverse (University of Guelph)

Scholars Portal Dataverse > University of Guelph Dataverse > **Agri-environmental research data repository Dataverse**

✉ Contact ⟳ Share

The Agri-environmental Research Data Repository was established by the University of Guelph Library in 2012 with the assistance of the Ontario Ministry of Agriculture, Food and Rural Affair's (OMAFRA) Knowledge Translation and Transfer (KTT) Funding Program. The purpose of this repository is to preserve and provide access to agricultural and environmental data produced during research projects undertaken at the University of Guelph. Special emphasis will be placed on research funded by OMAFRA.
Please note that the University of Guelph Research Data Repository houses research data created by researchers from all other disciplines at the University of Guelph.

To have your data included in this repository please contact us at: lib.research@uoguelph.ca

| Search this dataverse... | **Q Find** | Advanced Search |

☑ ⅄ **Dataverses (14)**
☑ 🗎 **Datasets (64)**
☐ 🗎 Files (2,175)

Publication Date
2016 (17)
2017 (16)
2012 (12)
2013 (11)
2015 (11)

1 to 10 of 78 Results

⇅ Sort ▾

Soil properties and seedling growth data for: "Soil properties as constraints to seedling regeneration beyond alpine treelines in the Canadian Rocky Mountains" 🗎
Apr 11, 2018 - Department of Geography Dataverse

Davis, Emma L., 2018, "Soil properties and seedling growth data for: "Soil properties as constraints to seedling regeneration beyond alpine treelines in the Canadian Rocky Mountains"", doi:10.5683/SP/SUTR1T, Scholars Portal Dataverse, V1, UNF:6:0TLCWgmv5XyQeK2i0a/ZJQ==

These data were produced as part of a study investigating the effects of soil properties on tree seedling germination beyond existing geographic ranges. Engelmann spruce seeds were grown in a greenhouse and growth chamber in soils collected from four study areas in the Canadian R...

Feedback Survey Results

Overall results

Increased awareness of RDM Best Practices

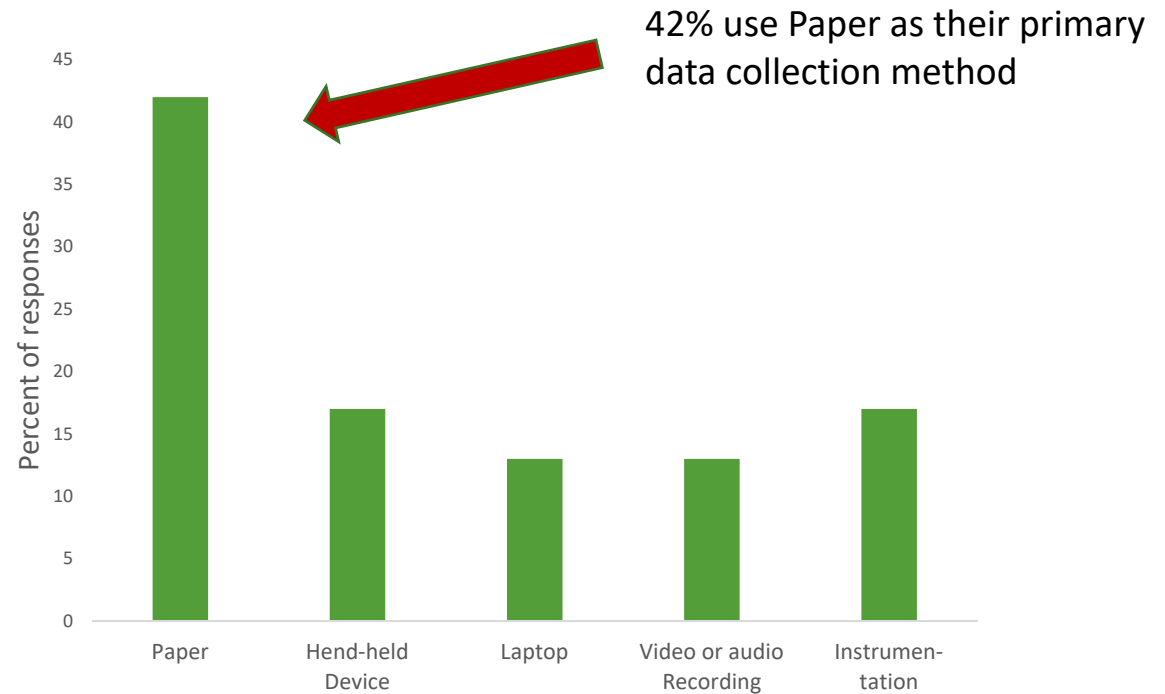Increased awareness or a Reminder of research project best practices – surrounding data

Very excited about availability of these types of workshops

Learned new ways or we reminded them of good research practices when it came to data!

Feedback Survey Results

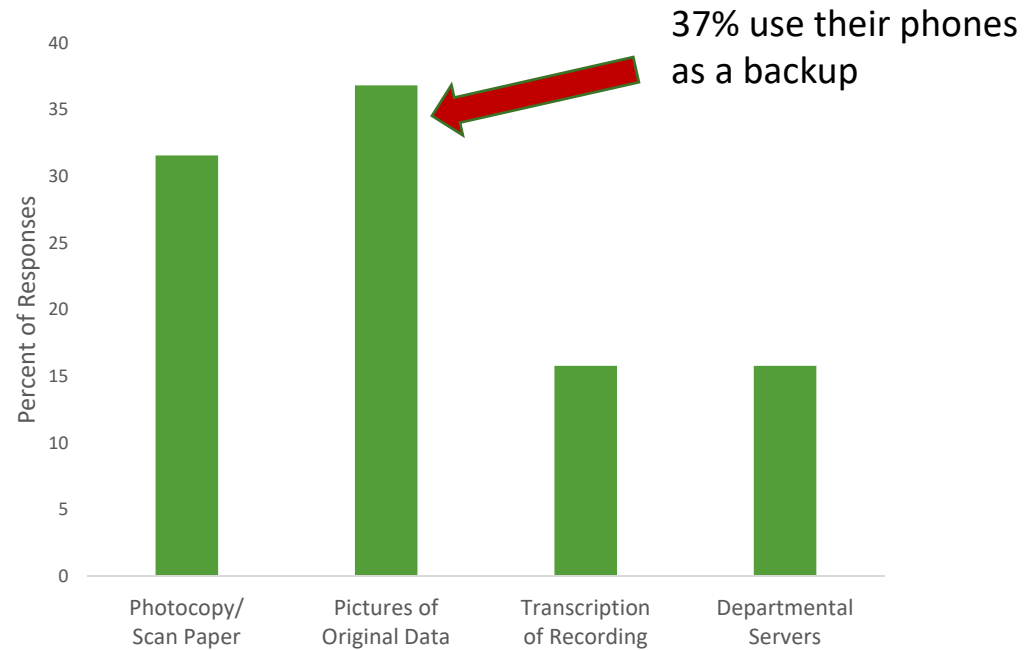Working with the Agricultural research community

Data Collection Methods:

42% use Paper as their primary data collection method

Feedback Survey Results

Working with the Agricultural research community

Raw data Back-up:

37% use their phones
as a backup

Workshop Takeaways from the Student

"Don't procrastinate, Get organized!"

Data preservation

README files

Plan how to manage data before project starts

Data organization

Students requesting more information to present to their own labs and friends

Students saying that this workshop should be a requirement for all incoming Grad students and should be taken at the start of their program.

Workshop Takeaways from the Instructors

We didn't teach then anything new!

We reminded them how important it is to organize your data

Strongly encouraged students to DOCUMENT!!

Very successful set of workshops

Requests for in-class sessions

Faculty attending

Continuing an updated series this summer and into the 2018-2019 Academic year

Contacts

A. Michelle Edwards

Statistical Consultant, OAC

edwardsm@uoguelph.ca

Carol Perry

Research & Scholarship Librarian

carolp@uoguelph.ca

Credits

Slide 1:
Finnhorse stallions: Retrieved under CC BY-SA 2.0 license from
https://www.flickr.com/photos/smerikal/5532818874/in/photostream

Cutting silage at Folsetter: Retrieved under CC BY SA 2.0 license from
https://upload.wikimedia.org/wikipedia/commons/9/93/Cutting_silage_at_Folsetter_1.jpg

Slide 15:
Photo by Adam Cohn. Used under CC-BY-NC-ND license from:
https://www.flickr.com/photos/adamcohn/34578872301/sizes/l

Slide 17:
Portage home page. Used under CC-BY-NC-Share 4.0 from:
https://portagenetwork.ca/

All other slide images:
Copyright University of Guelph