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
- 1 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?

OR DOES THIS LOOK FAMILIAR?
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**

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 ABC1111, NPQ13131, and AFS45456, made $312,400 in 2016 and spent approximately $7,000 in feed.

Start date of 2017 Trial: May 15, 2017

<table>
<thead>
<tr>
<th>Horse ID</th>
<th>Starting Weight</th>
<th>Feed Consumed Week 1</th>
<th>Weight Week 1</th>
<th>Feed Consumed Week 2</th>
<th>Weight Week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC11111</td>
<td>1160</td>
<td>66.7</td>
<td>1159</td>
<td>69.7</td>
<td>1154</td>
</tr>
<tr>
<td>NPQ13131</td>
<td>896</td>
<td>79.4</td>
<td>894</td>
<td>84.4</td>
<td>903</td>
</tr>
<tr>
<td>AFS45456</td>
<td>976</td>
<td>72.3</td>
<td>989</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td><strong>Farm ID</strong></td>
<td><strong>Horse Registration Number</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>101</td>
<td>ABC111111</td>
<td>Natasha Figueroa</td>
<td>73</td>
<td>2</td>
<td>$289,100.00</td>
<td>$6,879.00</td>
</tr>
<tr>
<td>3</td>
<td>101</td>
<td>ABC12121</td>
<td>Faye Lowe</td>
<td>69</td>
<td>2</td>
<td>$256,900.00</td>
<td>$6,056.00</td>
</tr>
<tr>
<td>4</td>
<td>101</td>
<td>ABC131311</td>
<td>Jo Bass</td>
<td>56</td>
<td>1</td>
<td>$398,700.00</td>
<td>$4,735.00</td>
</tr>
<tr>
<td>5</td>
<td>102</td>
<td>DEF11111</td>
<td>Gregg Richardson</td>
<td>65</td>
<td>1</td>
<td>$587,400.00</td>
<td>$4,330.00</td>
</tr>
<tr>
<td>6</td>
<td>102</td>
<td>DEF12121</td>
<td>Tiffany Simmons</td>
<td>35</td>
<td>2</td>
<td>$87,600.00</td>
<td>$7,179.00</td>
</tr>
<tr>
<td>7</td>
<td>102</td>
<td>DEF131311</td>
<td>Rudy McKenzie</td>
<td>48</td>
<td>1</td>
<td>$197,600.00</td>
<td>$4,204.00</td>
</tr>
<tr>
<td>8</td>
<td>103</td>
<td>GHJ111111</td>
<td>Gregg Richardson</td>
<td>65</td>
<td>1</td>
<td>$587,400.00</td>
<td>$7,240.00</td>
</tr>
<tr>
<td>9</td>
<td>103</td>
<td>GHJ121211</td>
<td>Max Gill</td>
<td>64</td>
<td>1</td>
<td>$187,300.00</td>
<td>$7,844.00</td>
</tr>
<tr>
<td>10</td>
<td>103</td>
<td>GHJ131311</td>
<td>Monique Chandler</td>
<td>75</td>
<td>2</td>
<td>$314,500.00</td>
<td>$5,517.00</td>
</tr>
</tbody>
</table>
Week 4 - Review

Review workshop content

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

- Complete the metadata template for data deposit
  - 19 required elements
<table>
<thead>
<tr>
<th>Metadata element</th>
<th>Description</th>
<th>Fill in information as applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Full title of the dataset</td>
<td></td>
</tr>
<tr>
<td>Author (s)</td>
<td>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].)</td>
<td></td>
</tr>
<tr>
<td>Production Date</td>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Funding agency</td>
<td>The source(s) of funds for the production of the work</td>
<td></td>
</tr>
<tr>
<td>Time period covered</td>
<td>Time period covered by the data</td>
<td></td>
</tr>
<tr>
<td>Date of collection</td>
<td>The date(s) when the data were actually collected</td>
<td></td>
</tr>
<tr>
<td>Keywords</td>
<td>Words or phrases that describes the data collection’s content (please provide at least 5 keywords)</td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Summary describing the purpose and scope of the data collection, and what questions the investigators attempted to answer.</td>
<td></td>
</tr>
</tbody>
</table>
Agri-environmental research data repository Dataverse (University of Guelph)

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

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 Affairs (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... 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

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

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 them 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