

Impactful Biomedical Research:
Achieving Quality and Transparency

Research Data Management

Laure Perrier
Research Data Management Librarian
June 10, 2016



Objectives

- To describe the current issues facing researchers with regards to research data management
- To outline strategies for researchers to meet current challenges
- To provide examples of meaningful support related to research data management

Current Issues

Drivers for research data management

- **Research funder policies:** Encourage or mandate creating data management plans, deposit data in repositories
- **Journals:** Require datasets to be published or made accessible (BMJ, PLOS)

Current Issues

Drivers for research data management

Recent activity,

- International Committee of Medical Journal Editors: Sharing Clinical Trial Data (January 2016)

Current Issues

Current Obligations

CIHR

- bioinformatics, atomic, and molecular coordinate data
- retain original datasets (all data) for a minimum of 5 years after the end of the grant

Current Issues

Future Directions

CIHR

- Draft Tri-Agency Statement on Digital Data Management
 - Data management planning
 - Preservation, retention, and sharing

Current Issues

wellcome**trust**

[Our philosophy](#)

[Funding](#)

[Managing a grant](#)

[Education resources](#)

News

[Investments](#)

[About us](#)

[Media office](#)

[News feeds](#)

▶ [Press releases](#)

• [2016](#)

• [2015](#)

• [2014](#)

• [2013](#)

• [2012](#)

• [2011](#)

▶ [FAQs](#)

▶ [Image bank](#)

▶ [Press release archive](#)

Global scientific community commits to sharing data on Zika

10 February 2016



Leading global health bodies including academic journals, NGOs, research funders and institutes, have committed to sharing data and results relevant to the current Zika crisis and future public health emergencies as rapidly and openly as possible.

Organisations including the Bill and Melinda Gates Foundation, Médecins Sans Frontières, the US National Institute of Health and the Wellcome Trust, along with leading academic journals including Nature, Science and the New England Journal of Medicine, have signed a joint declaration and hope that other bodies will come on board in the coming weeks.

The statement is intended to ensure that any information that might have value in combatting the Zika outbreak is made available to the international community free of

KEEP IN TOUCH

Stay up-to-date with what we're doing



Current Issues

Why keep data / make data available?

- Find and understand data when needed
- Validate results
- Ensure research is visible and has impact
- Get credit when others cite work
- Avoid unnecessary duplication

Current Issues

- 64 RCTs (oldest at top)
- 1987 – 2002
- Positive findings: aprotinin more effective than comparative treatment
- ~4,000 participants recruited without need

Source: Fergusson D, Glass KC, Hutton B, Shapiro S. Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? Clin Trials. 2005;2(3):218-29

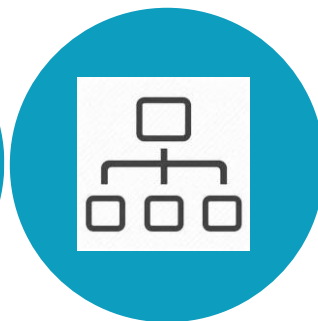
1.	Dec-87
2.	Mar-89
3.	Apr-89
4.	Sep-90
5.	Sep-90
6.	Dec-90
7.	Jun-91
8.	Sep-91
9.	Dec-91
10.	Apr-92
11.	Jun-92
12.	Jun-92
13.	Jun-92
14.	Nov-92
15.	Dec-92
16.	Jan-93
17.	Jul-93
18.	Aug-93
19.	Dec-93
20.	Jan-94
21.	Feb-94
22.	Feb-94
23.	Feb-94
24.	Apr-94
25.	Jul-94
26.	Aug-94
27.	Aug-94
28.	Oct-94
29.	Oct-94
30.	Dec-94
31.	Dec-94
32.	Feb-95
33.	Feb-95
34.	Feb-95
35.	Apr-95
36.	Jun-95
37.	Jun-95
38.	Sep-95
39.	Oct-95
40.	Oct-95
41.	Oct-95
42.	May-96
43.	Jul-96
44.	Aug-96
45.	Aug-96
46.	Oct-96
47.	Dec-96
48.	Jan-97
49.	Jan-97
50.	Aug-97
51.	Sep-97
52.	Dec-97
53.	Oct-98
54.	Oct-98
55.	Nov-98
56.	Aug-99
57.	Sep-99
58.	Mar-00
59.	Dec-00
60.	Dec-00
61.	Jan-01
62.	Sep-01
63.	Sep-01
64.	Jan-02

Strategies

4 KEY STRATEGIES



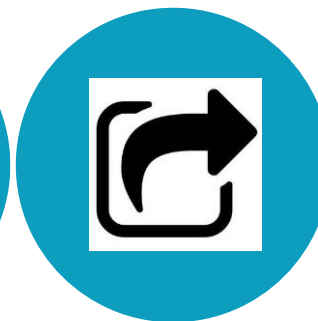
**Make
a Plan**



**Create a
System**



**Secure
Your Data**



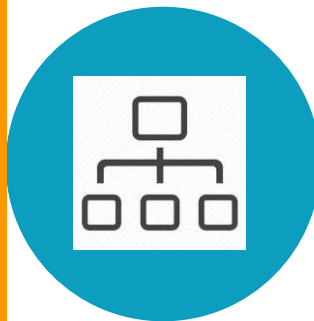
**Open File
Formats**

Strategies

4 KEY STRATEGIES



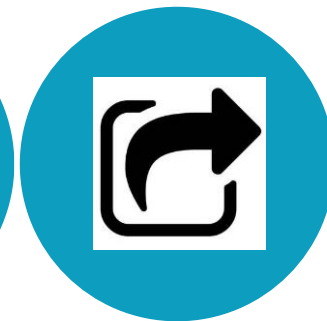
**Make
a Plan**



**Create a
System**



**Secure
Your Data**



**Open File
Formats**

Strategies: Make A Plan




Data Management Plan

- Type of data produced
- Documentation (metadata)
- Security, storage, management, and back-up of data
- Archiving and preservation
- Sharing and re-use

Strategies: Make A Plan



 UNIVERSITY OF
TORONTO

Office of the Vice-President, Research and Innovation
Human Research Ethics Program

Office Use Only
Application Number:

**ETHICS REVIEW APPLICATION FORM FOR
SUPERVISED AND SPONSORED RESEARCHERS**
(For use by graduate students, post-docs, residents, external investigators, and visiting
professors/researchers)

**SECTION A – GENERAL
INFORMATION**

1. TITLE OF RESEARCH PROJECT

Before you start, familiarize yourself
with:

[TCPS2](#)
[Application instructions](#)
[Office FAQs](#)

25. DATA SECURITY, RETENTION AND ACCESS

(a) Describe how data (including written records, video/audio recordings, artifacts and questionnaires) will be protected during the conduct of the research and dissemination of results.

(b) Explain how long data or samples will be retained. (If applicable, referring to the standard data retention practice for your discipline) Provide details of their final disposal or storage. Provide a justification if you intend to store your data for an indefinite length of time. If the data may have archival value, discuss how participants will be informed of this possibility during the consent process.

Strategies: Make A Plan



Make
a Plan

25. DATA SECURITY, RETENTION AND ACCESS

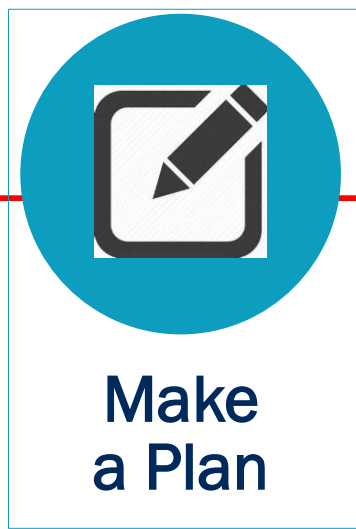
(a) Describe how data (including written records, video/audio recordings, artifacts and questionnaires) will be protected during the conduct of the research and dissemination of results.

(b) Explain how long data or samples will be retained. (If applicable, referring to the standard data retention practice for your discipline) Provide details of their final disposal or storage. Provide a justification if you intend to store your data for an indefinite length of time. If the data may have archival value, discuss how participants will be informed of this possibility during the consent process.

(c) If participant anonymity or confidentiality is not appropriate to this research project, please explain.

(d) If data will be shared with other researchers or users, please describe how and where the data will be stored and any restrictions that will be made regarding access.

Strategies: Make A Plan



- Tools available for drafting plans
 - DMP Assistant: portagenetwork.ca
 - DMP Online: dmponline.dcc.ac.uk
 - DMP Tool: dmptool.org

Strategies: Make A Plan



- Sample Data Management Plans

- Generic examples (UNC: The Odum Institute)

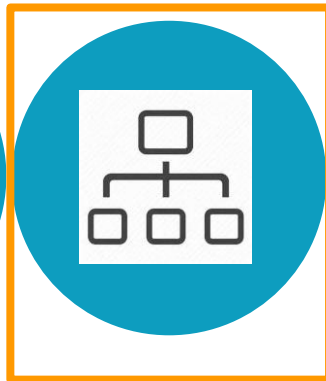
- www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=570

Strategies

4 KEY STRATEGIES



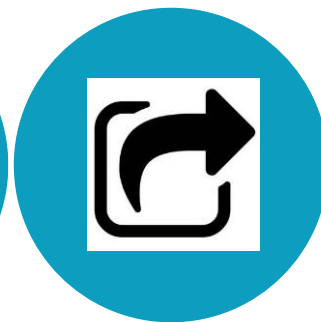
**Make
a Plan**



**Create a
System**

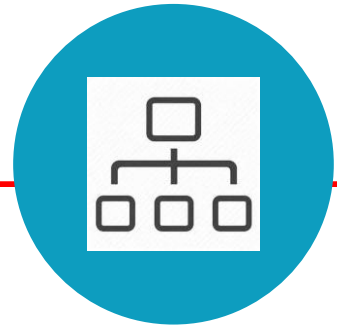


**Secure
Your Data**



**Open File
Formats**

Strategies: Create a System



Create a System

Chris Buddle @CMBuddle

#protip
NEVER name folders "stuff",
"miscellaneous", "important stuff", "misc
stuff"

Today, I realize this is what I do.
#ScienceInAction

RETWEETS 12 FAVORITES 13

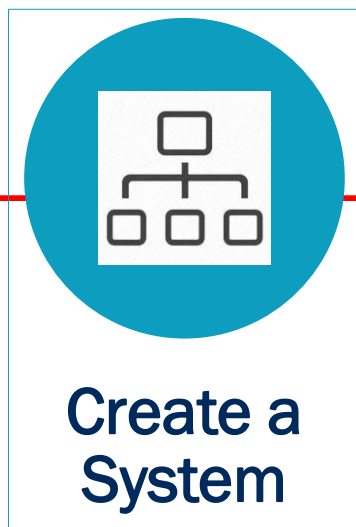
3:21 PM - 1 Sep 2015

Source: Kristin Briney.

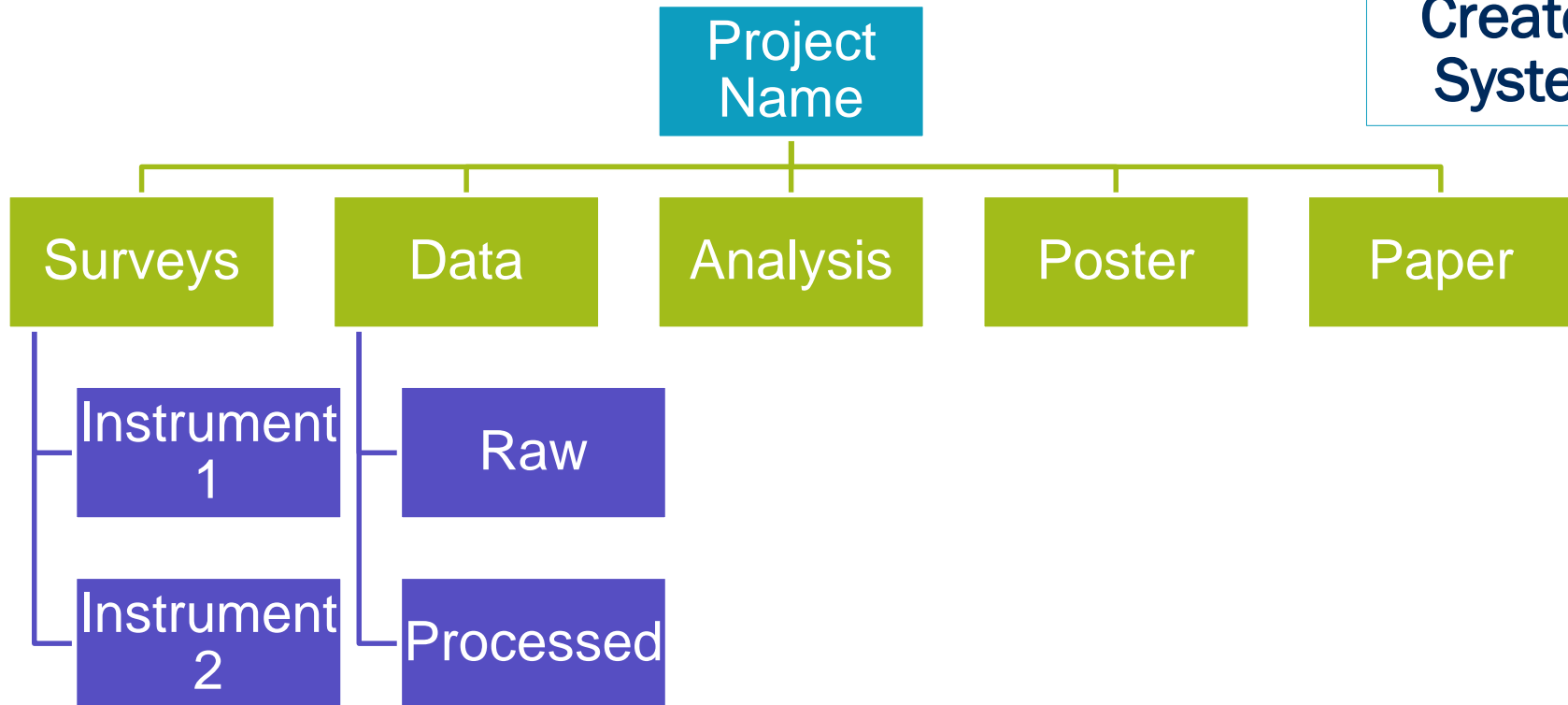
2015

Material in support of a verbal presentation,
not for interpretation as a stand-alone document: May 2016

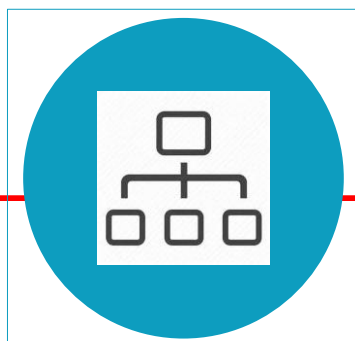
Strategies: Create a System



Folders: hierarchy



Strategies: Create a System



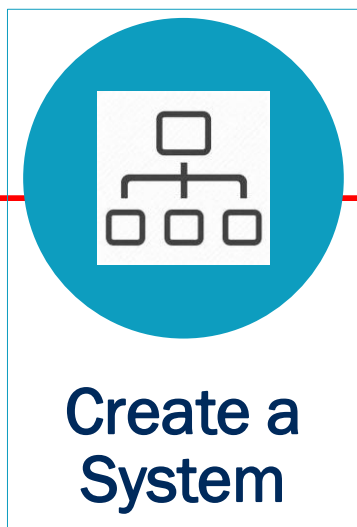
Create a
System

Folders: hierarchy

- May need to list which files belong in which folders

Strategies: Create a System

Folders: piling

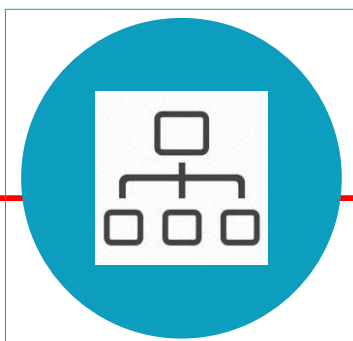


Study

Poster

Paper

Strategies: Create a System

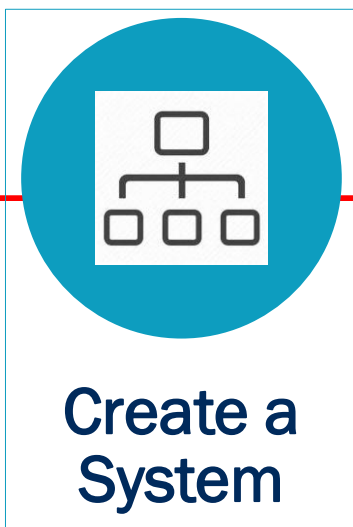


Create a
System

Folders: piling

- Less hierarchy = file names need more detail

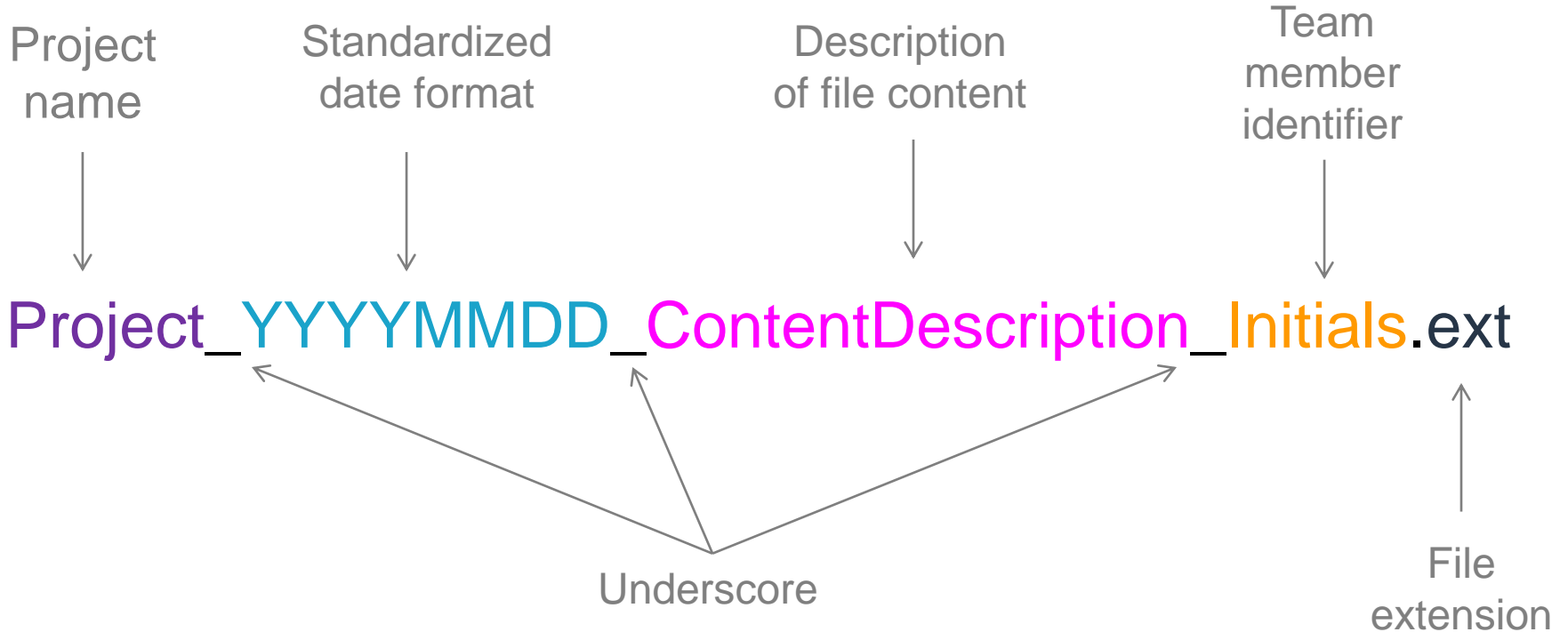
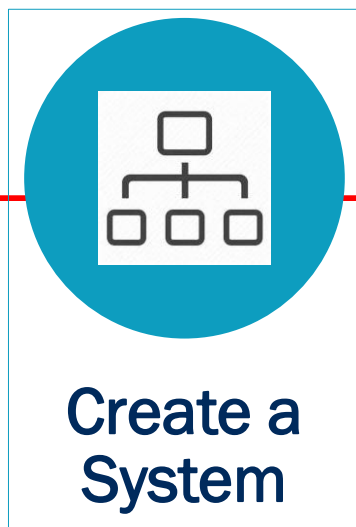
Strategies: Create a System



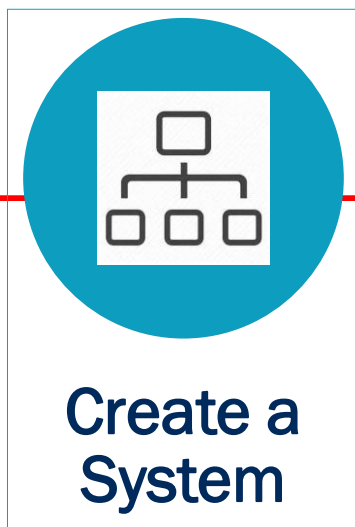
Naming Files: Best practices

- Avoid special characters (#\$%)
- Capitals or underscores (*FileName.xxx*)
- Date (ISO): YYYYMMDD
- Version information

Strategies: Create a System



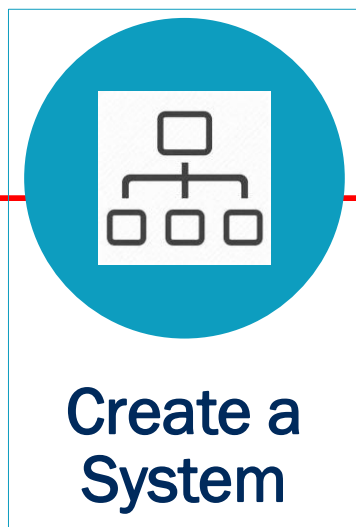
Strategies: Create a System



Naming Files: Use meaningful names

- Project/experiment name or acronym
- Location/spatial coordinates
- Researcher name/initials
- Date or date range of experiment
- Type of data

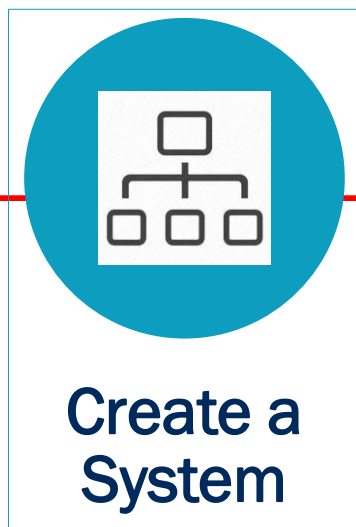
Strategies: Create a System



Documentation (metadata):

- Describes your data set
- Data documentation (metadata) helps you understand data in detail
- Helps other researchers find, use, properly cite your data

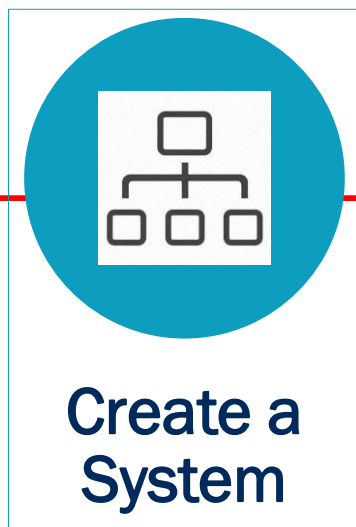
Strategies: Create a System



Documentation (metadata):

- Title
- Creator
- Dates
- Subject
- Funders
- Rights
- Language
- Location
- Methodology
etc.....

Strategies: Create a System



Documentation (metadata):

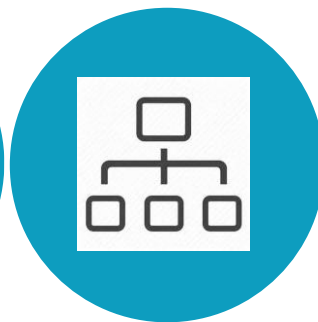
- Many standards for specific research disciplines
- Digital Curation Centre:
www.dcc.ac.uk/resources/metadata-standards

Strategies

4 KEY STRATEGIES



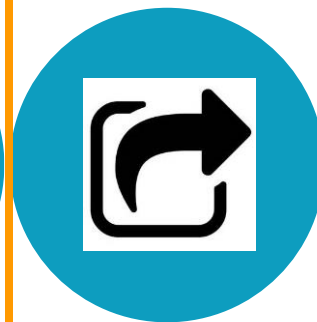
**Make
a Plan**



**Create a
System**



**Secure
Your Data**



**Open File
Formats**

Strategies: Secure Your Data



Secure
Your Data

CAMBRIDGE
news

WED 6 °C Cloudy

Site Directory Search...

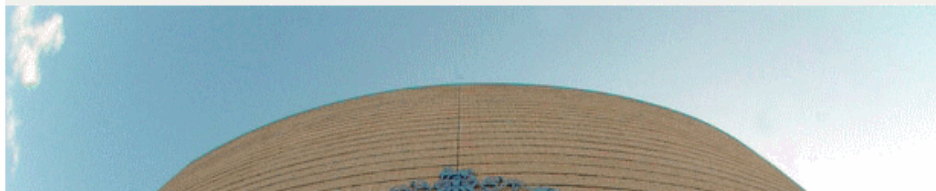
News Sport Business Lifestyle What's On Jobs Property Cars Buy & Sell Directory Notices Dating ALL

Cambridge | Cambourne | Ely | Haverhill | Huntingdon, St Ives & St Neots | Newmarket | Cambridge Live | Royston | Saffron Walden | Your Community | Community Awards | Education | International Student Guide | University News | Cycling | Pictures | Jobs news | Traffic and travel | Addenbrooke's | Help Our Hospice | Obituaries | Get Involved | E-Edition | RSS

Cambridge student 'devastated' as burglar steals the only copy of his PHD work

By Cambridge News | Posted: January 30, 2016

By Josh Thomas



house hunting?



we have hundreds of homes in your area

click here

Strategies: Secure Your Data



Secure
Your Data

3-2-1 Rule:

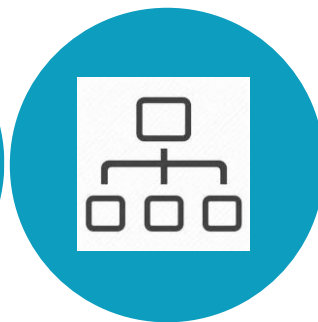
- **3** copies of your data
- **2** different locations
- More than **1** type of storage hardware

Strategies

4 KEY STRATEGIES



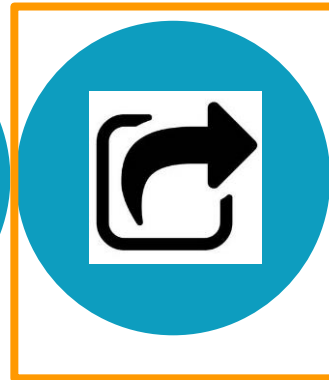
**Make
a Plan**



**Create a
System**

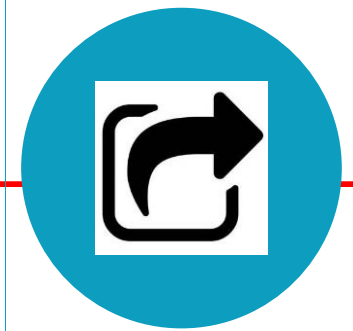


**Secure
Your Data**



**Open File
Formats**

Strategies: Open File Formats

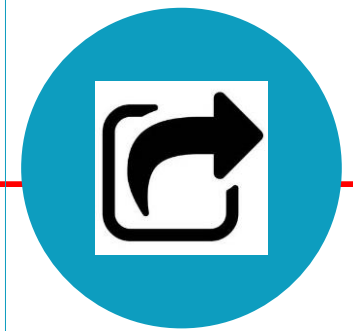


Open File
Formats

Data needs to be:

- Readable
- Accessible
- Understandable

Strategies: Open File Formats



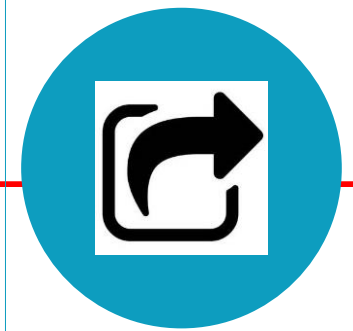
Open File
Formats

Data needs to be **Readable**

- Use non-proprietary formats

Yes	No
.txt	.docx (Word)
.csv	.xlsx (Excel)
M4a (MPEG-4)	.mov (Quicktime)
.tif	.gif or .jpg (images)
XML	RDBMS

Strategies: Open File Formats

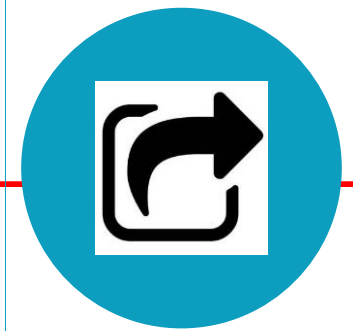


Open File
Formats

Data needs to be **Accessible**

- Move data to new media
- Average life span ~3-5 years
- If no open file format: Preserve software

Strategies: Open File Formats

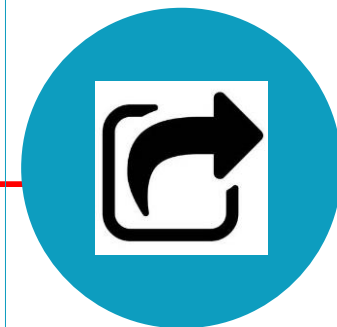


Open File
Formats

Data needs to be **Understandable**

- Data must include notes
- Include details
- Others should be able to understand it

Strategies: Open File Formats



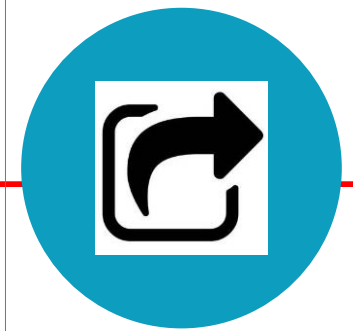
Open File
Formats

Data needs to be **Understandable**

Example: Data Dictionary (quantitative)

Variable	Variable Name	Variable Type	Variable Width	Values / Notes
Participant ID Number	ID	Numeric	3	001-900
Date of Birth	DOB	YYYY/MM/DD		1900-2010/1-12/1-31
Status	STAT	Numeric	1	1 = alive 2 = deceased
Hemoglobin	HB	Numeric	2.1	4.0 - 8.0
Urinary Iodine	UI	Numeric	4.1	0.0 – 1000.0

Strategies: Open File Formats



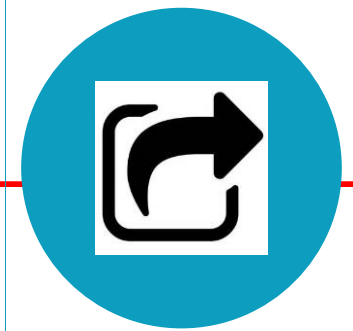
Open File
Formats

Data needs to be **Understandable**

Example: Data Dictionary (qualitative)

Code	Code Name	Explanation
Clarity	CLA	Coherence of components
Structure	STR	Arrangement between component parts
Navigation	NAV	Accurately ascertaining position and planning for movement through information
Saliency	SAL	Quality by which item stands out in relation to its neighbours
Flow	FLO	Moving along in a logical, steady manner

Strategies: Open File Formats

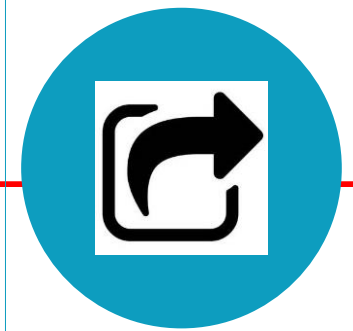


Open File
Formats

Data repositories

- Secure, long-term place for research data
- Often can impose appropriate access restrictions and /or embargoes

Strategies: Open File Formats



Open File
Formats

Data repositories at UToronto

- Dataverse

<http://dataverse.scholarsportal.info/dvn>

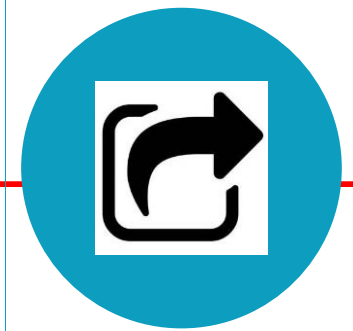
- TSpace

<https://tspace.library.utoronto.ca>

- Collections UofT (beta)

<https://collections.library.utoronto.ca>

Strategies: Open File Formats



Open File
Formats

Data repositories

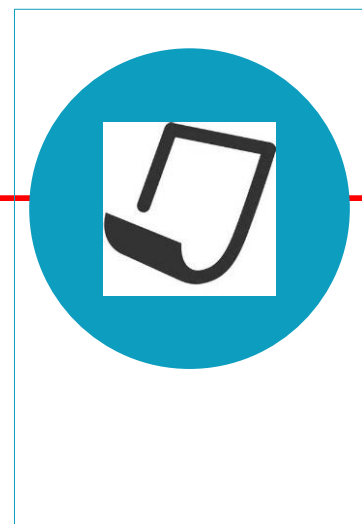
- Subject-specific

- Registry of Research Data Repositories

- www.re3data.org (see: Browse)

Finally

UToronto



onesearch.library.utoronto.ca/researchdata

The screenshot shows the top navigation bar of the University of Toronto Libraries website. The header includes the University of Toronto Library logo and a search bar. Below the header, the main content area is titled "Research Data Management". On the left, there is a sidebar menu with options: "Research Data Management", "Organizing", "Sharing", "Preserving", and "Contact Us". The main content area contains a welcome message and a list of services: "Organizing your data", "Sharing your data", "Preserving your data", and "Contact us". On the right, there is a "Ask" chat button with the text "Chat with a librarian" and a "Frequently Asked Questions" link.

UNIVERSITY OF TORONTO LIBRARY

Research Services Visit Ask About My Account UoT Search

Research Data Management

Welcome to the University of Toronto Libraries' Research Data Management resources! Good data management is an important part of the research process and is increasingly emphasized by institutions and funding agencies. University of Toronto Libraries has data management expertise and can assist you in the following areas:

- Organizing your data
- Sharing your data
- Preserving your data
- Contact us

Share this page

Ask
Chat with a librarian

About Ask a Librarian Chat

Also see our

Frequently Asked Questions

Material in support of a verbal presentation,
not for interpretation as a stand-alone document: May 2016

Questions?

Laure Perrier

l.perrier@utoronto.ca

Impactful Biomedical Research:
Achieving Quality and Transparency

Research Data Management

Laure Perrier
Research Data Management Librarian
June 10, 2016

