Responsible Conduct of Research Workshop Series, 2018-2019

Record Keeping, Data Management, and Sharing of Information

-- September 13, 2018--
Swipe in, Swipe out = validation you attended full workshop

No swipe? I cannot give you credit

Bring your MSU ID card every time!
As of Fall 2016

RCR Minimum Plan requirements

Tracked automatically in SABA =
Tracked by department in GradInfo =

Master’s plan B and grad professional:

CITI Modules Year 1 (  
- Introduction to the Responsible Conduct of Research  
- Authorship  
- Plagiarism  
- Research Misconduct  

6 hours discussion-based training (by completion of the degree)

Master’s Plan A students:

CITI Modules Year 1 (  
- Introduction to the Responsible Conduct of Research  
- Authorship  
- Plagiarism  
- Research Misconduct  

6 hours discussion-based training (by completion of the degree)

CITI Modules Year 2  
Complete 3 additional from specific list

Doctoral students:

CITI Modules Year 1 (  
- Introduction to the Responsible Conduct of Research  
- Authorship  
- Plagiarism  
- Research Misconduct  

6 hours discussion-based training (by completion of the degree)

CITI Modules Year 2  
Complete 3 additional from specific list

Year 3 forward  
3 hours of annual refresher training
The plan represents the basic university plan. Each department/program or college will develop a plan that at a minimum incorporates these university-level requirements. The Graduate School RCR Workshop series may be used to help fulfill both the annual refresher and discussion-based training requirements.

*Students who are supported by NSF, NIH, or USDA grants may be required to complete additional specific training; they must meet the timeline and content requirements of training for that grant.

*Students engaged in research involving human subjects or animal use must complete the Michigan State University training modules for those subjects before submitting IRB or IACUC approvals. These modules may be completed as part of the training requirements below, or in addition to them, depending on the department/program or college plan.
All graduate professional, master’s and doctoral students

1) Year 1
All new graduate and graduate professional students will complete 4 CITI online modules within the first year of enrollment in their program: Completion of this requirement will be tracked in SABA
• Introduction to the Responsible Conduct of Research
• Authorship
• Plagiarism
• Research Misconduct

2) Discussion-Based Training
All graduate and graduate professional students must complete a minimum of 6 hours of discussion-based training prior to receiving their degrees. These hours can be completed at any point in the graduate program, including during the first 2 years (e.g., as part of a course), or as part of the ongoing training requirement (for doctoral students). Specifics about the number of hours required, the content, and the timing of this training will be defined in the individual department/program or college plan. For master’s Plan A and PhD students completion of this requirement will be recorded by the department in GradInfo as “Initial” training.
Getting to SABA (RCR Reg and records) and CITI
What the CITI screen looks like…Use your MSU ID to log in
In addition to 1 and 2 above, master’s plan A and doctoral students will complete:

3) Year 2
Within the first 2 years of enrollment in their program, master’s plan A and doctoral students will complete 3 additional MSU online training modules, to be selected from the following list. Specific requirements for course selection may be defined in the individual department/program or college plan.

Completion of this requirement will be tracked in SABA.

- CITI Collaborative Research
- CITI Conflicts of Interest
- CITI Data Management
- CITI Financial Responsibility
- CITI Mentoring
- CITI Peer Review
- IACUC Tutorial for Animal Care Training (in http://Train.ORA.msu.edu)
- Rigor and Reproducibility Course (in production)

In addition to 1, 2 and 3 above, doctoral students will complete:

4) Annual Refresher Training
Starting in year 3, all doctoral students must complete 3 hours of annual refresher training; this can include discussion-based training and online courses beyond the 7 required in basic training. Specifics about the number of hours required, the content, and the timing of this training will be defined in the individual department/program or college plan. Completion of this requirement will be recorded by the department in GradInfo as “Annual” training.
Please indicate your current status at MSU

1. Early (first years)
2. Mid (middle years)
3. Late
4. Postdoctorate
5. Faculty
6. Other
In 2011 Tilburg University suspended Stapel, pending further investigation, for fabricating and manipulating data for his research publications. This scientific misconduct took place over a number of years and affected at least 55 publications (58) http://en.wikipedia.org/wiki/Diederik_Stapel

“The committee, which interviewed dozens of Stapel's former students and colleagues, concluded that Stapel acted alone. The report says he would discuss experimental designs in detail with collaborators and would claim to conduct the surveys at high schools and universities with which he had special arrangements. The experiments, however, never took place, and Stapel gave collaborators made-up data sets, investigators allege. In other instances, the report says, he told colleagues that he had an old data set lying around that he hadn't had a chance to analyze. When Stapel did conduct actual experiments, the committee found evidence that he manipulated results.”

“Many of Stapel's students were simply given data to analyze and graduated without having ever run an experiment, the report says. The commission writes that Stapel was “absolute lord of the data” in his collaborations. Colleagues or students who asked to see raw data told the commission they were given excuses or even threatened and insulted.”

Fraud Scandal Fuels Debate Over Practices of Social Psychology
Even legitimate researchers cut corners, some admit

By Christopher Shea

The discovery that the Dutch researcher Diederik A. Stapel made up the data for dozens of research papers has shaken up the field of social psychology, fueling a discussion not just about outright fraud, but also about subtler ways of misusing research data. Such misuse can happen even unintentionally, as researchers try to make a splash with their peers—and a splash, maybe, with the news media, too.
Even before the Stapel case broke, a flurry of articles had begun appearing this fall that pointed to supposed **systemic flaws in the way psychologists handle data.**

"If high-impact journals want this kind of surprising finding, then there is pressure on researchers to come up with this stuff," says Mr. Wagenmakers, an associate professor in the psychology department's methodology unit [Eric-Jan Wagenmakers, of the University of Amsterdam].

Bad things happen when researchers feel under pressure, he adds—and it doesn't have to be Stapel-bad: "There's a slippery slope between making up your data and torturing your data.'"

In September, in comments quoted by the statistician Andrew Gelman on his blog, Mr. Wagenmakers wrote: "The field of social psychology has become very competitive, and high-impact publications are only possible for results that are really surprising. **Unfortunately, most surprising hypotheses are wrong. That is, unless you test them against data you've created yourself.**"
He’s back: Data faker Diederik Stapel will support research at vocational university

with 17 comments

Diederik Stapel, the social psychology researcher who has had 58 papers retracted after admitting that he made up the data, has a new job: helping other researchers.

Stapel, according to BN DeStem (via Google Translate),

is going to help professors at major research projects and studies. “Outside his mistake he has been a good and thorough scientist,” says Hein van Oorschot of the Executive Board [of NHTV Breda, in the Netherlands]. “He has a vast knowledge in the field of research. He knows how the world works.”

The appointment is for one year. Van Oorschot tells BN DeStem:

He is not engaged in research and will not teach. He gets a supporting role, and he is the source of information and a helping hand for large projects.

Stapel has not responded to a request for comment. His last academic appointment, as an adjunct of sorts at Fontys Academy in Tilburg in 2014, was short-lived. Around the same time he resigned, he admitted sock puppetry in comments here at Retraction Watch.

Earlier, Stapel had settled with Dutch prosecutors and agreed to 120 hours of community service, after resigning from Tilburg University and relinquishing his PhD.
But wait...that’s not the end
No academic post for fraudster Diederik Stapel, after all

Recently, we reported that social psychologist and renowned data faker Diederik Stapel had found himself a new gig supporting research at a vocational university in the Netherlands — but it appears that was short-lived.

According to multiple news reports, NHTV Breda will not be employing Stapel, after all.

Here’s our Google translate of a portion from De Telegraaf:

“[Stapel’s] work at the NHTV have been terminated after one week. This was announced by the Breda college Tuesday.”

And this from NOS:

“The social psychologist Diederik Stapel do not get a second chance at the College of Tourism and Transport in Breda (NHTV).

Staff and the MR have too many objections to the appointment of the controversial scientist. Therefore, the school decides to reverse the appointment...”

Omroep Brabant includes Stapel’s reaction to the news:

“[Stapel] himself said Tuesday on Radio 1 that he was sad about the quick dismissal. “I was pleased with the new job...””

Stapel’s last academic appointment at Fontys Academy in Tilburg in 2014 was short-lived; he resigned around the same time he admitted sock puppetry in comments here at Retraction Watch.

Stapel has retracted dozens of papers, settled with Dutch prosecutors and agreed to 120 hours of community service, after resigning from Tilburg University and relinquishing his PhD.

Update 9/14/16 9:35 a.m. eastern: We’ve received a press release from NHTV about the announcement; here’s an excerpt, according to our Google translate:

“At the request of the Participation Council to reconsider the cooperation of the Executive Board of the university decided to terminate the cooperation with him.

Although the Board has formally acted correctly, the Participation Council, the Executive convinced that with many employees within the university insufficient confidence in the collaboration. The cooperation is not productive, and therefore it was decided the consulting agreement no further effect.”
DEFINITIONS

- **Data OWNERSHIP**: who has LEGAL rights to the data and holds onto data once project is complete.
- **Data COLLECTION**: the process of obtaining recordable, measurable endpoints.
- **Data STORAGE & PROTECTION**: how and where data are physically or virtually stored and protected from theft or tampering.
- **Data RETENTION**: how long the data need to be stored according to regulations.
- **Data ANALYSIS**: how raw data are modified for others to use and relate to.
- **Data SHARING**: communicating data – at any level – to others privately or publicly.

- Guidelines for Responsible Data Management in Scientific Research, Clin Tools, ORI FUNDED
Research Data

- **Control**
  - physical management of research data and records

- **Access**
  - conditions under which various people may inspect or use research data

- **Ownership**
  - in a legal sense, who owns the data
“Raw” data – typical research data in your discipline!

- Research records
  - Research plans as described in funding or dissertation proposals
  - Procedures documented in IRB, IACUC, or ORCBS protocols
  - Lab notebooks or other records of the implementation of your research
  - Manuscripts and other reports
Thess are Raw Data
MSU INSTITUTIONAL DATA
DATA SECURITY AT MSU

Seth Edgar, MSU CISO
Agenda

- Institutional Data
  - Definition
  - Responsibilities
  - Handling

- Information Security Program
  - Threats
  - Assets
  - Protections
  - Approach

- Questions
Institutional Data at MSU

Institutional Data are all of the data and records held by the University, *in any form or medium*, for the administration, operation, or governance of the University or any unit of the University

- Confidential
- Public

https://tech.msu.edu/about/guidelines-policies/msu-institutional-data-policy/
Confidential Data at MSU (4 Categories)

- Institutional Data that could, by itself or in combination with other such Data, be **used for identity theft** or related crimes.
- Institutional Data whose public disclosure is **restricted by law, contract, University policy, professional code, or practice** within the applicable unit, discipline, or profession.
- Records of the University’s **security measures**.
- Institutional Data whose **value would be lost or reduced by unauthorized disclosure or by disclosure in advance of the time prescribed** for its authorized public release, or whose unauthorized disclosure would otherwise **adversely affect the University financially**.

[https://tech.msu.edu/about/guidelines-policies/msu-institutional-data-policy/](https://tech.msu.edu/about/guidelines-policies/msu-institutional-data-policy/)
Confidential Data Categories (and)

- **Identity Theft**
  - SSN
  - Credit Cards
  - Bank Accounts
  - Health Insurance #’s
  - Digital Signatures

- **Restricted Disclosure**
  - Educational
  - Medical
  - NDA’s
  - Granting Agency Data
  - Human Subjects

- **Security Measures**
  - Passwords
  - Security Tools
  - Digital Keys
  - Security Brands
  - Threat Assessment

- **Publicity Devaluation**
  - Results prior to publication
  - Trade Secrets
  - Non Patentable Information
  - Computer Applications
  - Research & Scholarly Integrity
Public Data at MSU

Institutional Data that have become generally available to members of the public because a person with authority to do so has intentionally released or distributed them without restriction or limitation.

https://tech.msu.edu/about/guidelines-policies/msu-institutional-data-policy/
Responsibility

Members of the University community are individually responsible for the security and integrity of Institutional Data in their possession or control, including their proper storage and disposal.

https://tech.msu.edu/about/guidelines-policies/msu-institutional-data-policy/
Handling

- Email
  - Do not send Confidential Data via email

- Retention
  - Encrypt Confidential Data at rest and in-transit

- Disposal
  - See MSU guidance for proper disposal of data

- Data Sharing
  - FileDepot

- International Travel
  - Use VPN
  - Maintain control of your computer/files at all times
  - Encrypt your device
  - Avoid networks that require a software download first

- Cloud Computing
  - Not all cloud services are created/configured equally

- Software/Tools
  - Information Security performs risk assessments of tools at NO COST
Who do I call if I have questions?

- **Security/Data Handling Inquiries**
  - MSU IT Service Desk
    - 517-432-6200
    - ithelp@msu.edu

- **Specific Consultations**
  - Seth Edgar, MSU CISO (me)
    - ciso@msu.edu

- **Malicious/Spam Emails**
  - Mail Abuse Team
    - abuse@msu.edu
Information Security Program: Case-Study

- Campaign Type: Ransomware
- Unique features:
  - Leveraged recently-published Microsoft vulnerability
  - Automated propagation
  - Kill Switch/Poor execution
- Affected:
  - Telefonica
  - Myrck Pharmaceuticals
  - UK’s National Health Service
  - FedEx
<table>
<thead>
<tr>
<th>Loss</th>
<th>Theft</th>
<th>Exposure</th>
<th>Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data rendered irretrievable</td>
<td>Data stolen, usually motivated by profit</td>
<td>Sensitive data posted in a public forum</td>
<td>Data altered without our knowledge</td>
</tr>
<tr>
<td>Ransomware</td>
<td>SQL Injection</td>
<td>Data Dumps</td>
<td>Tampering</td>
</tr>
<tr>
<td>Deletion</td>
<td>Carding</td>
<td>Hacktivism</td>
<td>Errors</td>
</tr>
<tr>
<td>Overwriting</td>
<td>Identify Theft</td>
<td>Doxing</td>
<td>APT</td>
</tr>
</tbody>
</table>
MSU High-Value Assets

- Research
- Administrative Data
- PCI Data
- Critical Infrastructure
- FERPA Data
- HIPAA Data
Protections

- Secure High-Value Assets
  - Encryption
  - Security Baseline
  - Two-factor

- Prioritize towards best practices in all areas
  - Mail Protection
  - NIST Frameworks
  - Purchasing Risk Assessments

- Leverage existing technologies
  - Firewalls
  - Patch management
Example: NIST Security Control Families

<table>
<thead>
<tr>
<th>ID</th>
<th>FAMILY</th>
<th>ID</th>
<th>FAMILY</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Access Control</td>
<td>MP</td>
<td>Media Protection</td>
</tr>
<tr>
<td>AT</td>
<td>Awareness and Training</td>
<td>PE</td>
<td>Physical and Environmental Protection</td>
</tr>
<tr>
<td>AU</td>
<td>Audit and Accountability</td>
<td>PL</td>
<td>Planning</td>
</tr>
<tr>
<td>CA</td>
<td>Security Assessment and Authorization</td>
<td>PS</td>
<td>Personnel Security</td>
</tr>
<tr>
<td>CM</td>
<td>Configuration Management</td>
<td>RA</td>
<td>Risk Assessment</td>
</tr>
<tr>
<td>CP</td>
<td>Contingency Planning</td>
<td>SA</td>
<td>System and Services Acquisition</td>
</tr>
<tr>
<td>IA</td>
<td>Identification and Authentication</td>
<td>SC</td>
<td>System and Communications Protection</td>
</tr>
<tr>
<td>IR</td>
<td>Incident Response</td>
<td>SI</td>
<td>System and Information Integrity</td>
</tr>
<tr>
<td>MA</td>
<td>Maintenance</td>
<td>PM</td>
<td>Program Management</td>
</tr>
<tr>
<td>CNTL NO.</td>
<td>CONTROL NAME</td>
<td>PRIORITY</td>
<td>INITIAL CONTROL BASELINES</td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------</td>
<td>----------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOW</td>
<td>MOD</td>
</tr>
<tr>
<td>AC-1</td>
<td>Access Control Policy and Procedures</td>
<td>P1</td>
<td>AC-1</td>
</tr>
<tr>
<td>AC-2</td>
<td>Account Management</td>
<td>P1</td>
<td>AC-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AC-2 (1) (2) (3) (4) (5) (11) (12)</td>
</tr>
<tr>
<td>AC-3</td>
<td>Access Enforcement</td>
<td>P1</td>
<td>AC-3</td>
</tr>
<tr>
<td>AC-4</td>
<td>Information Flow Enforcement</td>
<td>P1</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-5</td>
<td>Separation of Duties</td>
<td>P1</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-6</td>
<td>Least Privilege</td>
<td>P1</td>
<td>Not Selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AC-6 (1) (2) (5) (9) (10)</td>
</tr>
<tr>
<td>AC-7</td>
<td>Unsuccessful Logon Attempts</td>
<td>P2</td>
<td>AC-7</td>
</tr>
<tr>
<td>AC-8</td>
<td>System Use Notification</td>
<td>P1</td>
<td>AC-8</td>
</tr>
<tr>
<td>AC-9</td>
<td>Previous Logon (Access) Notification</td>
<td>P0</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-10</td>
<td>Concurrent Session Control</td>
<td>P3</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-11</td>
<td>Session Lock</td>
<td>P3</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-12</td>
<td>Session Termination</td>
<td>P2</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-13</td>
<td>Withdrawn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC-14</td>
<td>Permitted Actions without Identification or Authentication</td>
<td>P3</td>
<td>AC-14</td>
</tr>
<tr>
<td>AC-15</td>
<td>Withdrawn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC-16</td>
<td>Security Attributes</td>
<td>P0</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-17</td>
<td>Remote Access</td>
<td>P1</td>
<td>AC-17 (1) (2) (3) (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AC-17 (1) (2) (3) (4)</td>
</tr>
<tr>
<td>AC-18</td>
<td>Wireless Access</td>
<td>P1</td>
<td>AC-18 (1) (4) (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AC-18 (1) (4) (5)</td>
</tr>
<tr>
<td>AC-19</td>
<td>Access Control for Mobile Devices</td>
<td>P1</td>
<td>AC-19 (5)</td>
</tr>
<tr>
<td>AC-20</td>
<td>Use of External Information Systems</td>
<td>P1</td>
<td>AC-20 (1) (2)</td>
</tr>
<tr>
<td>AC-21</td>
<td>Information Sharing</td>
<td>P2</td>
<td>AC-21</td>
</tr>
<tr>
<td>AC-22</td>
<td>Publicly Accessible Content</td>
<td>P3</td>
<td>AC-22</td>
</tr>
<tr>
<td>AC-23</td>
<td>Data Mining Protection</td>
<td>P0</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-24</td>
<td>Access Control Decisions</td>
<td>P0</td>
<td>Not Selected</td>
</tr>
<tr>
<td>AC-25</td>
<td>Reference Monitor</td>
<td>P0</td>
<td>Not Selected</td>
</tr>
</tbody>
</table>
Approach

- Defense-in-depth
- Minimized redundancy
- Principle of Least Privilege
- Iterative Approach
Defense-in-Depth Case-Study: Ransomware
Iterative Security Approach

NIST Cyber Security Framework

1. Identify
   - AM: Asset Management
   - BE: Business Environment
   - GV: Governance
   - RA: Risk Assessment
   - RM: Risk Management Strategy

2. Protect
   - AC: Access Control
   - AT: Awareness Training
   - DS: Data Security
   - IP: Information Protection
   - PT: Protective Technology

3. Detect
   - AE: Anomalies and Events
   - CM: Security Continuous Monitoring
   - DP: Detection Processes

4. Respond
   - RP: Response Planning
   - CO: Communications
   - AN: Analysis
   - MI: Mitigation
   - IM: Improvements

5. Recover
   - RP: Recovery Planning
   - IM: Improvements
   - CO: Communications
Questions?
SHARE YOUR DATA:

GETTING STARTED WITH RDM

Scout Calvert, PhD
Data Librarian
Michigan State University Libraries
calvert4@msu.edu
OBJECTIVES

• Set context for data sharing
• Understand the basic components of research data management (RDM)
• Promote good data management practices
• Provide resources for further study
UPCOMING RDM WORKSHOPS

DIGITAL SCHOLARSHIP LAB

- **Crash Course in Research Data Management**
  November 9, 3:00 pm
- **Prepare Your Data for Upload to a Repository**
  November 16, 3:00 pm
- **Writing Data Management Plans**
  Spring semester, TBA
WHY MANAGE FOR SHARING?
CONTEXT FOR RDM

- Data deluge: answer is not a bigger hard drive
- **OSTP Memo 2013** ("Increasing Access to the Results of Federally Funded Scientific Research")
- **DMPs now required for most grant applications**
- Trend toward Open Data
- **Credit for data sharing**
- Managing data for more products
OPEN DATA AND OPEN SCIENCE

Optimism that:

- data sharing will improve knowledge production and discovery
- data interoperability will lead to stronger knowledge claims (e.g., climate science)
- data openness will encourage the discovery of errors and discourage fraud
- data sharing will promote reproduction and replication in science
CITATION ADVANTAGE

- Gene expression studies with data in a public repository are cited 9% more (Piwowar 2013).
- In astrophysics, that advantage could be as much as 50% (Dorch 2012).
- Open access articles are cited 18% more (Piwowar et al. 2018).
BENEFIT TO FUTURE YOU

- Return to a project easily after time off or a set-back
- Avoid catastrophic data loss
- Improve collaborations
- Safeguard data and context loss due to team changes
PLANNING FOR RDM
GRANT REQUIREMENTS

*Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.*
NSF DATA MANAGEMENT PLAN

1. types of data . . . and other materials to be produced in the course of the project
2. standards to be used for data and metadata format and content
3. policies for access and sharing . . .
4. policies and provisions for re-use, re-distribution, and the production of derivatives
5. plans for archiving data . . . and for preservation of access to them
NIJ DATA ARCHIVING PLAN

- Data
- Cleaned data with analysis variables
- Project documents
- Syntax, code, scripts for transformed variables
- Variable labels and codes, record counts, missing data
- Documentation, including questionnaires, instruments, etc.
- Codebooks, glossaries, technical documentation, etc.
GOLD STANDARD FOR RDM

A well described data package that can be passed to another researcher in the same discipline and meaningfully reused (reproduced, replicated, reanalyzed) without additional communication.
PREPARING YOUR DATA PACKAGE
THE DATA PACKAGE

• Raw data
• Processed, cleaned data
• Metadata, documentation, description
• Code, scripts, analysis
• README, file manifest, codebooks

. . . with human readable file names, in an organized file structure, in open, long-lived file formats.
FILE NAMES AS METADATA

- Author(s)
- Project name
- Type of data file
- Type of analysis
- Date, version
FILE NAMING

Project-Code-Kindoffile-YYYYMMDD
PaperNameReviewerInitialsYYYYMMDD
HDEX_Interviews_Coded_by_SC_20130812
Docility_analysis.csv
Docility_Data_Codebook.pdf
DocilityDataBulls.csv
DocilityDataCows.csv

http://calvert4.msu.domains/presenting/shareyourdata.html#
FILE ORGANIZATION

A) Organized by File type

Example A

- Code
  - Step.1
  - Step.2
- Data
  - Processed
  - Raw
- Results
  - Figure.1
  - Figure.2
  - Models
- readme.txt

B) Organized by Analysis

Example B

- Figure.1
  - Code
  - Data
  - Results
- Figure.2
  - Code
  - Data
  - Results
  - readme.txt
FOLDER ORGANIZATION
<table>
<thead>
<tr>
<th>Name</th>
<th>Date Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Articles</td>
<td>Aug 10, 2017, 12:01 PM</td>
</tr>
<tr>
<td>Business</td>
<td>Aug 10, 2017, 12:02 PM</td>
</tr>
<tr>
<td>Coursework</td>
<td>Aug 10, 2017, 12:02 PM</td>
</tr>
<tr>
<td>CV and dossier</td>
<td>Aug 10, 2017, 12:02 PM</td>
</tr>
<tr>
<td>Papers</td>
<td>Aug 10, 2017, 12:01 PM</td>
</tr>
<tr>
<td>Personal</td>
<td>Aug 10, 2017, 12:02 PM</td>
</tr>
<tr>
<td>Presentations</td>
<td>Aug 10, 2017, 12:01 PM</td>
</tr>
<tr>
<td>Research</td>
<td>Aug 10, 2017, 12:03 PM</td>
</tr>
<tr>
<td>Health Data Exploration Project</td>
<td>Aug 10, 2017, 12:05 PM</td>
</tr>
<tr>
<td>Data</td>
<td>Aug 10, 2017, 12:05 PM</td>
</tr>
<tr>
<td>Interview questions</td>
<td>Aug 10, 2017, 12:05 PM</td>
</tr>
<tr>
<td>IRB documents</td>
<td>Aug 10, 2017, 12:04 PM</td>
</tr>
<tr>
<td>Lit review</td>
<td>Aug 10, 2017, 12:04 PM</td>
</tr>
<tr>
<td>Paper drafts</td>
<td>Aug 10, 2017, 12:04 PM</td>
</tr>
<tr>
<td>Presentations</td>
<td>Aug 10, 2017, 12:04 PM</td>
</tr>
<tr>
<td>Survey questions</td>
<td>Aug 10, 2017, 12:04 PM</td>
</tr>
<tr>
<td>Teaching</td>
<td>Aug 10, 2017, 12:02 PM</td>
</tr>
</tbody>
</table>
FILE TYPE AND LONGEVITY

- Prefer or export to non-proprietary formats
- Open, widely used file types
- Be sure to document software and versions
- **A list from UK Data Service**
DATA DOCUMENTATION

- Codebook
- Data dictionaries
- README
- Code or scripts
- Methodology
- Metadata
README.TXT

- Project name and grant number
- Contact information, list of co-authors or team
- Citation for publication (with DOI)
- Citation information for this dataset
- License information for this dataset
- Project summary
- Description of data processing and analysis
- File and folder manifest
- Notes on experimental set up
- Software and versions used for collection and analysis
- Metadata for instruments and sensors
- Citations for other data used in analysis
BACKUP PRINCIPLES

- Three copies, two local, one offsite (3-2-1) (4-2-2)
- Enterprise storage from iCER
- MSU data policy, cloud policy
- Spartan Drive

http://calvert4.msu.domains/presenting/shareyourdata.html#/
SPECIAL CONSIDERATIONS

- IRB / Human subjects data
- Permission to share?
- Deidentified, fully anonymized
- Sensitive ecological data
- Disguised or non-located
DATA REPOSITORIES
WHY PUT DATA IN A DATA REPO?

- Exposure, searchability, findability
- Reliable, independent infrastructure
- Persistent identifiers and data citation
- Ensure frictionless sharing
- **Registry of Research Data Repos (Re3data)**
COMMON FEATURES

- Licensing for sharing
- Data embargo
- Digital object identifiers (DOIs)
- Versioning datasets
- Public or private storage
- Permissions and collaboration
WHAT'S NOT A REPOSITORY?

- Some place in the cloud like dropbox
- Some storage space on your departmental server
- Space on your website
- **Spartan Drive (One Drive)**
  (But NB: we get 5 TB so include in your backup plan!)
WRAPPING UP
CITATION MANAGERS

Library offers monthly workshops on:

• Zotero
• Mendeley
• EndNote
GETTING HELP

- Contact me for RDM and DMP consultation
- Research Data Management Guidance
- DMPTool
RESOURCES

- Re3data.org
- ICPSR
- DataONE Primer on Data Management
- Ten Simple Rules for Creating a Good Data Management Plan
- Ten Simple Rules for Reproducible Computational Research
- Ten Simple Rules for the Care and Feeding of Scientific Data
QUESTIONS?

Scout Calvert
calvert4@msu.edu
GOOGLE APPS/G. SUITE AT MSU

Catherine Zhang, Information technologist
Justin Booth, Senior Manager

September 13, 2018
Agenda

- MSU Google Apps for Education
  - Overview
  - Appropriate use
  - Support
- Spartan 365
  - Spartan Drive – One Drive
- Cloud Services
  - Introduction
  - Key factors to consider
  - Risk triage
  - When you are not sure, please ask
- QAs
Google Apps/G Suite Overview

- MSU Google Apps (https://googleapps.msu.edu) for Education started in summer 2009 including 6 Core Apps.

- Serving faculty, academic faculty retirees, students, and staff
- Offers increased security and protections of intellectual property rights through integrated MSU NetIDs and passwords access ----not found in the public version.
- Use of any other Google tools is not recommended for MSU data.
- is NOT part of MSU Google Core Apps!
<table>
<thead>
<tr>
<th>Type of Data</th>
<th>ITS File Share</th>
<th>Spartan Drive</th>
<th>SharePoint 365</th>
<th>SharePoint Campus</th>
<th>Local Hard Drive</th>
<th>Google Drive</th>
<th>Spartan Skype</th>
<th>AFS</th>
<th>Departmental Server</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Data</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Confidential Data</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>FERPA Data</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>HIPPA Data</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>SSN</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>PCI &amp; SSN</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

*CONSULT* - MSU Information Security
Appropriate Use

- Appropriate user of MSU Google Apps for Education Services

- Allows Google to store MSU educational records and requires Google to comply with FERPA

- MSU agrees to follow the Google Apps Acceptable Use Policy (AUP).

- *It is generally appropriate for many uses in instruction and scholarship*, and even some “low consequence” business and administrative uses.
Family Educational Rights and Privacy Act (FERPA)

Get the Latest on FERPA at familypolicy.ed.gov

- Frequently Asked Questions
- FERPA for parents and students and school officials
- Protection of Pupil Rights Amendment (PPRA)
- Guidance and Notices

G Suite Acceptable Use Policy

Use of the Services is subject to this acceptable use policy ("AUP").

If not defined here, capitalized terms have the meaning stated in the applicable contract ("Agreement") between customer and Google.

You agree not to, and not to allow third parties or Your End Users, to use the Services:

- to generate or facilitate unsolicited bulk commercial email;
- to violate, or encourage the violation of, the legal rights of others;
- for any unlawful, invasive, infringing, defamatory, or fraudulent purpose;
- to intentionally distribute viruses, worms, Trojan horses, corrupted files, hoaxes, or other items of a destructive or disruptive nature;
- to alter, disable, interfere with or circumvent any aspect of the Services;
- to test or reverse-engineer the Services in order to find limitations, vulnerabilities or evade filtering capabilities;
Cautions and discretions should be exercised to consider:

- Where the authoritative copy of educational records is kept.

- Whether the sharing permissions on any particular document complies with MSU policies, including those regarding Access to Student Information.

- Whether privacy and confidentiality concerns around any particular piece of information or data is appropriate to inclusion in a document hosted by Google.

- **Be aware:** Google G Suite for Education is NOT appropriate for use with most other forms of MSU Confidential Data.
Appropriate Use

Detailed MSU guidelines and policies can be found in https://tech.msu.edu/about/guidelines-policies/

users are advised to fully review
Appropriate Use
Data Formats

- Export from Core Apps: most common data formats, such as Microsoft Office, PDF, and iCalendar.

- Uploaded content generally will remain of the original type and can be downloaded.

- Similar to consumer version, users will be able to remove their content, or copies of the content.
Appropriate Use
Owner responsibility

- Must regularly review sharing permissions to validate authorized parties have a valid need-to-know role and that content is appropriate for hosting under the Google Apps/G Suite for Education Agreement. Here is an article on how to limit sharing permissions of a specific file in your GoogleApps account: https://support.google.com/drive/answer/2494893.

- Must ensure compliance with MSU policies when granting access to documents and avoid improper disclosure of confidential data that could occur from sharing folders and individual documents.

- Neither MSU nor Google makes backups of information stored on these services. MSU users should assume individual responsibility for regularly backing up information and data they store with Google.
Appropriate Use

Space/quotas limitations

- **Google Drive** – *unlimited storage*, however a 50MB file size limit on presentations and documents uploaded to Drive and converted to Docs and Slides, and a 100MB limit on spreadsheets converted to Sheets.

- **Google Sites**: MSU as an institution is limited to a total of 100GB for all sites/users. A large number of pages on a Site will cause slowness.
Support for MSU Google Apps for Education

- MSU IT Service Desk (517-432-6200)

- Issues regarding the portal, log-in problems, usage.

- Google

  - Support for the specific use of the programs and services are provided by Google:
  - [http://drive.google.com/support/](http://drive.google.com/support/)
  - [http://docs.google.com/support/](http://docs.google.com/support/)
  - [https://www.google.com/support/calendar/](https://www.google.com/support/calendar/)
  - [https://www.google.com/support/sites/](https://www.google.com/support/sites/)
  - [https://support.google.com/edu/classroom](https://support.google.com/edu/classroom)
  - [https://support.google.com/hangouts](https://support.google.com/hangouts)

- Also, refer to [MSU Google Apps for Education: Known Issues](#) or to the [Google App Status Dashboard](#) if an outage is suspected.
Spartan 365

- http://spartan365.msu.edu/
- Spartan Mail
- Spartan Drive 1 TB

<table>
<thead>
<tr>
<th>Type of Data</th>
<th>ITS File Share</th>
<th>Spartan Drive</th>
<th>SharePoint 365</th>
<th>SharePoint Campus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Data</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>Confidential Data</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>FERPA Data</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>HIPPA Data</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>SSN</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
<tr>
<td>PCI &amp; SSN</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>
Spartan 365

- **Spartan365.msu.edu.**
  - Students have access to:
  - Spartan Drive
  - Delve
  - Office Tools
    - Work, Excel
Other Available Services

- **MediaSpace – Kaltura**
  - Create, store and share media
  - [https://mediaspace.msu.edu](https://mediaspace.msu.edu)

- **Qualtrics Survey Tool**
  - Qualtrics can be used for surveys as well as research data collection.
  - [https://qualtrics.msu.edu/](https://qualtrics.msu.edu/)
  - SMS Texting surveys available
Other Available Services

D2L - Learning Management System
https://D2L.msu.edu

- Zoom Web Conferencing
  - https://MSU.zoom.us
Cloud Services

- Enterprise agreements
- Department account provisioning
- Seamless single sign-on
- Discount pricing
- Google Cloud coming in 2018

https://tech.msu.edu/network/cloud-services/
Cloud Services

**Key Factors to Consider**

Before using any cloud provider or service give the following factors due consideration:

- Non-negotiated terms of use
- Control of user content
- Security and privacy
- Backups
- Non-negotiated changes to terms of use
- Non-negotiated changes to the service
- Non-negotiated changes to the business model
- Data formats
- Indemnity
Cloud Services

How do I ask for help?

- [ ] [https://tech.msu.edu/support/help/](https://tech.msu.edu/support/help/)
- [ ] Service Desk (517) 432-6200
- [ ] ithelp@msu.edu
Other Collaboration Tools & Services

- MSU Gitlab - https://gitlab.msu.edu
- Private code repository for all of MSU
- Comparable to github.com
- Collaborate with staff and students
MSU Infrastructure

Virtual Hosting
- Self-managed
- Managed

Colocation
- Colocation Service allows researchers to place their existing servers in the Computer Center Datacenter and have secure access.
Remote Access

Welcome to the
MSU SSL VPN Web Portal

Username
Password
Realm

SSL VPN Upgrade and Move to New Client

The MSU SSL VPN will be upgraded on August 6, 2015.
The upgrade allows us to improve and maintain this service, as well as add support for Windows 10.

Following this upgrade, Windows users will be prompted to install the new VPN client called Pulse Secure.
Support for the legacy Network Connect client will be deprecated.

Mac OS X users should follow the Quick Start Installation Guide to install the Pulse Secure desktop client.
This guide also provides installation instructions for Android, iOS, and Windows clients.

IT administrators can find instructions for downloading and testing the Pulse Secure client in the SSL VPN Guide for IT Administrators.

If you need assistance or more information, please contact the MSU IT Service Desk at (517) 432-6200 or itserv@msu.edu.

https://vpn.msu.edu/
How to get Support

- https://tech.msu.edu/support/help
- (517) 432-6200, ithelp@msu.edu
- Walk-ins in the Tech Store (students who can help file a ticket)
- Unit level contacts (e.g. ICER, CTSI, Libraries, Colleges)
- If your request is stalled, contact myself or Justin Booth and include your ticket ID #
Appendix B: NIST Control

FAMILY: ACCESS CONTROL

AC-1 ACCESS CONTROL POLICY AND PROCEDURES

Control: The organization:

a. Develops, documents, and disseminates to [Assignment: organization-defined personnel or roles]:
   1. An access control policy that addresses purpose, scope, roles, responsibilities, management commitment, coordination among organizational entities, and compliance; and
   2. Procedures to facilitate the implementation of the access control policy and associated access controls; and

b. Reviews and updates the current:
   1. Access control policy [Assignment: organization-defined frequency]; and
   2. Access control procedures [Assignment: organization-defined frequency].

Supplemental Guidance: This control addresses the establishment of policy and procedures for the effective implementation of selected security controls and control enhancements in the AC family. Policy and procedures reflect applicable federal laws, Executive Orders, directives, regulations, policies, standards, and guidance. Security program policies and procedures at the organization level may make the need for system-specific policies and procedures unnecessary. The policy can be included as part of the general information security policy for organizations or conversely, can be represented by multiple policies reflecting the complex nature of certain organizations. The procedures can be established for the security program in general and for particular information systems, if needed. The organizational risk management strategy is a key factor in establishing policy and procedures. Related control: PM-9.

Control Enhancements: None.

References: NIST Special Publications 800-12, 800-100.

Priority and Baseline Allocation:

| P1 | LOW AC-1 | MOD AC-1 | HIGH AC-1 |
NSF Data Management Plan

NSF Data Sharing Policy

- Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF Data Management Plan Requirements

- Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled “Data Management Plan”. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.
NSF Data Management Plan Contents

- **Products of the Research:** The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

- **Data Formats:** The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

- **Access to Data and Data Sharing Practices and Policies:** Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

- **Policies for Re-Use, Re-Distribution, and Production of Derivatives.**

- **Archiving of Data:** Plans for archiving data, samples, and other research products, and for preservation of access to them.
Help with data at MSU

- Michigan State University
  MSU Libraries: Lifecycle Data Management Planning
  http://www.lib.msu.edu/about/diginfo/ldmp.jsp

- VISIT THIS!
DATA SHARING

- Data sharing is the practice of making data used for scholarly research available to other investigators.

- Many funding agencies, institutions, and publication venues have policies regarding data sharing because transparency and openness are considered by many to be part of the scientific method.

What Data do you share?

- Large Data bases
- Microarray information
- Genetic Sequences
The NIH policy on data sharing applies:

- To the sharing of final research data for research purposes.
- To basic research, clinical studies, surveys, and other types of research supported by NIH. It applies to research that involves human subjects and laboratory research that does not involve human subjects. It is especially important to share unique data that cannot be readily replicated.
- To applicants seeking $500,000 or more in direct costs in any year of the proposed project period through grants, cooperative agreements, or contracts.
- To research applications submitted beginning October 1, 2003.
Journal Open-Data Policies

- [http://oad.simmons.edu/oadwiki/Journal_open-data_policies](http://oad.simmons.edu/oadwiki/Journal_open-data_policies)

This is a list of journals with data-sharing mandates for their published articles. For example:

- **African Journal of Ecology**
  - Data that is integral to the paper must be made available in such a way as to enable readers to replicate, verify and build upon the conclusions published in the paper. Any restriction on the availability of this data must be disclosed at the time of submission.

- **Evolution**
  - Require authors to submit DNA sequence data to GenBank and phylogenetic data to TreeBase.
Limitations & Restrictions on Sharing of Data

- Health Insurance Portability & Protection Act (HIPAA)
- Sensitive but Unclassified Data
- Classified & Restricted Research Activities
- Intellectual Property Protections
- Ethical considerations
A recent breach by a recipient investigator of the Data Use Certification led to the on-line publication by PNAS of a manuscript that should never even have been submitted (2). While both PNAS and the NIH will deal with this specific breach, it is the wider research community that must police itself and prevent inappropriate publication in the future.
...our editors became aware that Dr. Zhang had signed a Data Use Certification indicating his agreement to comply with the NIH Genome-Wide Association Studies Policy for Data Sharing, which applies to the Gene Environment Association (GENEVA) studies, of which the Study of Addiction, Genetics and Environment (SAGE) is a part. Under the policy, investigators agree not to submit findings of the SAGE dataset(s) for publication until September 23, 2009. The PNAS publication clearly violates the SAGE embargo, and the authors agreed to retract their work in PNAS on September 9, 2009.
Research/Scholarly Integrity Resources

- NAS, Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age

- NIH Data Sharing Policy and Implementation Guidance

- UM, Research Data Management and Publishing Support (NSF Plans)

- MN, Funding Agency and Data Management Guidelines
  - [http://www.lib.umn.edu/datamanagement/funding](http://www.lib.umn.edu/datamanagement/funding)

- MSU, Lifecycle Data Management Planning
  - [http://www.lib.msu.edu/about/diginfo/ldmp.jsp](http://www.lib.msu.edu/about/diginfo/ldmp.jsp)

- NCDDC.Noaa
Research Data Management Guidance

The Research Data Management Guidance service is a joint effort between the MSU Libraries and University Archives to respond to the emerging concern surrounding research data management. Effective research data management is an active and ongoing process during the entire research lifecycle, from project conception to long after it is finished.

Our purpose is:

1. To provide guidance to faculty in the development and execution of research data management plans.
2. To connect faculty to distributed research data support services throughout the university.
3. To educate students, faculty, and staff about the importance of data management.

Best practices for research data management can

- help you secure grant funding
- enhance the ability to share data internally during a project
- bring traditional library and archival expertise for the organization and care of digital information to your own personal research projects
- create a data output that can become part of the scholarly record, earning recognition and citations
- help you make efficient and effective use of research funding

Contact:

Researchdata@mail.lib.msu.edu

Team Members:

Anna Callis - Digital Creation Librarian
NEXT WORKSHOP

- Authorship
- October 4, 2018
  - [http://grad.msu.edu/rcr/integrity.aspx](http://grad.msu.edu/rcr/integrity.aspx)

- Workshop Series Syllabus
  - [http://grad.msu.edu/rcr/docs/syllabus.pdf](http://grad.msu.edu/rcr/docs/syllabus.pdf)