Data Management for RESEARCH
Learning Objectives

- Identify and categorize your research data based on its sensitivity
- Apply the university’s Data Classification and Data Protection Guides
- Define DUA and DMP and identify relevant GW processes
- Understand the expectation for managing data throughout the research data lifecycle (i.e. from the point of collection or acquisition to storage, transmission, archiving and/or destruction)
Data Management for Research

**Research Data Management** involves:

- the organization, documentation, storage, preservation, and sharing of data collected and used in a research project.
- everyday research-focused and functional activities with the data during the lifetime of a research project.
- Application of principles governing the storage, sharing and preservation of research data.
- Activities necessary for the preservation, access, and possible re-use of the data after the completion of the current research project.
Why talk about data management?
OPEN Government Data Act

H.R. 1770:
“A bill to expand the Government’s use and administration of data to facilitate transparency, effective governance, and innovation, and for other purposes.”
“Data sharing enables researchers to rigorously test the validity of research findings, strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery.”

(Draft NIH Policy for Data Management and Sharing)

Research Data Management Is Important

- Supports compliance with regulatory, institutional, and funder requirements
- Augments research impact by allowing for future re-use (when appropriate)
- Enhances data security and research integrity
- Minimizes the risk of data loss
- Increases research efficiency
- Saves time and resources in the long run
- Protects the University from reputational, financial & legal risk
Findable, Accessible, Interoperable, and Reusable

For more info: https://www.go-fair.org/fair-principles/
Privacy Principles throughout managing research data
Privacy principles applied throughout research data lifecycle
Data Classification helps identifying the level of privacy and security protection to be applied to University Data (and records) and the scope and means in which the data/records can be shared.

**PUBLIC**
- Low risk
- Little or no controls are required to protect the confidentiality of Public data, yet, some level of control is required to prevent unauthorized modification or destruction of it.
- Open Access

**Examples:**
- Census Data
- COVID19 data publicly available (no identifiers)

**RESTRICTED**
- Moderate risk
- Must be protected from unauthorized access, use or disclosure due to university policies, contract, or designation, or due to proprietary or privacy considerations.

**Example:**
- Research data (except: human subject research data with identifiers)

**REGULATED**
- High risk
- Must be protected from unauthorized access, use or disclosure due to university policies, contract, or designation, or due to proprietary or privacy considerations.

**Examples:**
- Student PII
- Protected Health Information;
- Personal Identifiable Information (e.g. SSN, DL, DoB)

**GW Guide for Data Classification**
### Data Protection

#### Information Category
- **Regulated**
- **Restricted**
- **Public**

#### Risk Category
- **High Risk**
- **Moderate Risk**
- **Low Risk**

## COMPUTING EQUIPMENT

<table>
<thead>
<tr>
<th>Laptops and Mobile Devices - GW-owned or approved (Desktop, laptop, phone, tablet)</th>
<th>Regulated Data should only be accessed and processed with GW-owned or approved computers or mobile devices (such devices are configured and managed by the university and must be encrypted). Devices used to access and process Regulated data should be registered and managed by GW IT.</th>
<th>Regulated Data should preferably be accessed and processed with GW-owned or approved computers or mobile devices (such devices are configured and managed by the university and must be encrypted). Devices used to access and process Restricted data should be registered and managed by GW IT.</th>
<th>No limitations</th>
</tr>
</thead>
</table>
| Personally-Owned Devices (Desktop, laptop, phone, tablet) | **Regulated or Restricted information may not be downloaded, stored or synchronized on personally-owned workstations or mobile devices.** Requirements for accessing Regulated Information from personally-owned workstations or mobile devices are:  
  - Full Disk Encryption (FDE)  
  - Use of VPN  
  - Must be password protected  
  - Anti-Virus / Anti-Spyware software must be active and maintained up to date (Symantec Endpoint Protection available from it.gwu.edu download page)  
  - Updates for all installed software, as well as firmware and driver updates should be installed within a reasonable period. | No limitations |

Contact your IT Department and/or GW Information Security for further instructions.
<table>
<thead>
<tr>
<th>Information Category</th>
<th>Regulated</th>
<th>Restricted</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Category</td>
<td>High Risk</td>
<td>Moderate Risk</td>
<td>Low Risk</td>
</tr>
</tbody>
</table>

**STORAGE**

<table>
<thead>
<tr>
<th>Storage</th>
<th>Regulated data in physical form (paper, media) should be locked at all times and access should be restricted only to authorized users, with a legitimate business need.</th>
<th>Restricted data in physical form (paper, media) should be secured at all times and access should be restricted only to authorized users, with a legitimate business need.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>May be used, stored, shared, or processed only on GW IT hosted or approved servers or services such as file sharing or collaboration services, cloud-based email services, cloud-based backup and recovery services, etc.)</td>
<td>May be stored on departmental, GW IT- hosted, or approved cloud-based systems. May be stored in the following GW systems:-GW G-Drive-GW SharePoint, GW Box and GW Docs.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>No limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Contact your IT Department and/or GW Information Security for further instructions.</td>
</tr>
</tbody>
</table>
## Data Protection

<table>
<thead>
<tr>
<th>Information Category</th>
<th>Regulated</th>
<th>Restricted</th>
<th>Public</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Category</td>
<td>High Risk</td>
<td>Moderate Risk</td>
<td>Low Risk</td>
</tr>
</tbody>
</table>

### TRANSMISSION

#### Emailing

**Avoid emailing regulated data.**
- If there is a legitimate business need to email regulated data, it must be *encrypted* during transmission and VPN must be used.
- Emailing Regulated Information to or from a personal email address is strictly prohibited.

**Avoid emailing restricted data.**
- If there is a legitimate business need to email restricted data, it must be *encrypted* during transmission and VPN must be used.
- Emailing Restricted Information to or from a personal email address is discouraged.

Contact your IT Department and/or GW Information Security for further instructions.

### PRINTING or Reproduction

#### Printing and Reproduction (copy)

**Avoid printing regulated data.**
- Regulated Data should never be printed on any public device.
- The minimum necessary prints / copies may be made only by permission of originator or his/her designates.
- Working copies (prints) containing regulated data should secured at all times and returned to office or permanently destroyed (shredded) when no longer needed.

**Avoid printing restricted data.**
- Restricted Data should never be printed on any public device.
- Only the minimum necessary copies (prints) should be made.
- Working copies containing restricted data in should secured at all times and returned to office or permanently destroyed (shredded) when no longer needed.

Contact Privacy@gwu.edu for further guidance.

No limitations
### Data Classification

Example of common Data Types accompanied by the corresponding Data Classification.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Description &amp; Examples</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Private Personal Information / Personally Identifiable Information (PII)</strong></td>
<td>This is a category of sensitive information that is associated with an individual person (eg: employee, student, etc.)</td>
<td>REGULATED</td>
</tr>
<tr>
<td>• Social Security number</td>
<td>□ Biometric information</td>
<td></td>
</tr>
<tr>
<td>• Individual Tax IDs</td>
<td>□ Biographic/demographic data (Date and location of birth, Country of citizenship, Citizenship status, Marital status, Military status)</td>
<td></td>
</tr>
<tr>
<td>• National ID number</td>
<td>□ Criminal record &amp; criminal background check information</td>
<td></td>
</tr>
<tr>
<td>• Passport number</td>
<td>□ Tax information (e.g., W-2, W-4, 1099)</td>
<td></td>
</tr>
<tr>
<td>• Visa permit number</td>
<td>□ Disciplinary information</td>
<td></td>
</tr>
<tr>
<td>• Driver’s license number</td>
<td>□ Leave-of-absence reason</td>
<td></td>
</tr>
<tr>
<td>• Bank and credit/debit card numbers</td>
<td>□ Payroll and benefits information security</td>
<td></td>
</tr>
<tr>
<td>• Disability information</td>
<td>□ Health information</td>
<td></td>
</tr>
<tr>
<td>• Ethnicity, Gender</td>
<td>□ Payroll and benefits information security</td>
<td></td>
</tr>
<tr>
<td>• Financial account number in combination with a code, access code or password that would permit access to the account</td>
<td>□ Health information</td>
<td></td>
</tr>
<tr>
<td>• Address (all geographic subdivisions smaller than including street county, zip code)</td>
<td>□ Payroll and benefits information security</td>
<td></td>
</tr>
<tr>
<td>• Internet protocol (IP)</td>
<td>□ Health information</td>
<td></td>
</tr>
</tbody>
</table>

| **Protected Health Information (PHI)**          | The following individually identifiable data elements, when combined with health information about that person, make such information protected health information (PHI):                                                                 | REGULATED       |
| • Name                                         | □ Social Security number                                                                                                                                                                                             |                 |
| • Telephone number                             | □ Medical record number                                                                                                                                                                                              |                 |
| • Fax number                                   | □ Health plan beneficiary number                                                                                                                                                                                     |                 |
| • Email address                                | □ Certificate/license number                                                                                                                                                                                           |                 |
| • Address (all geographic subdivisions smaller than including street county, zip code) | □ Biometric identifier, including finger and voice print                                                                                                                                                            |                 |
| • Internet protocol (IP)                       | □ Vehicle identifier and serial number, including license plate number                                                                                                                                               |                 |
| • Any other unique identifying state number, characteristic, address, city, code, or combination that precinct, or allows identification of an individual addresses | □ Any other unique identifying state number, characteristic, address, city, code, or combination that precinct, or allows identification of an individual addresses |                 |
## Data Classification

Example of common Data Types accompanied by the corresponding Data Classification.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Description &amp; Examples</th>
<th>Classification</th>
</tr>
</thead>
</table>
| **Personal Data from European Union (EU)**    | The EU’s General Data Protection Regulation (GDPR) defines personal data as any information that can identify a natural person, directly or indirectly, by reference to an identifier including  
• Name  
• Any identification number  
• Location data  
• An online identifier  
• One or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person  
Any personal data that is collected from individuals in European Economic Area (EEA) countries is subject to GDPR. | REGULATED       |
| **Student Education Records (FERPA)**         | Records that contain information directly related to a student and that are maintained by the University or by a person acting for the University. The Family Educational Rights and Privacy Act (FERPA) governs release of, and access to, student education records. “Directory information” about a student is not regulated by FERPA and can be released by the University without the student’s permission. Students can request non-disclosure from the Registrar’s Office. | REGULATED       |
Data Use Agreements
What is a Data Use Agreement (DUA)?

DUAs are legally binding contracts between GW and another party providing for the transfer of data from the provider organization to the recipient organization.

These agreements may also be called Data Transfer Agreements or Data Sharing Agreements.

In general, any sharing of “restricted use” data requires an agreement between the providing party (owner of the data) and the receiving party (data user).

Failure to follow DUA terms could result in significant liabilities including possible criminal sanctions as well as impacting other rights of the parties involved.
Data Use Agreements

All DUAs involving GW as an executing party must be reviewed and approved by ORI.

To begin the review process, please complete the [DUA Intake Form](GW NetID login required) and submit the proposed DUA.
Data Use Agreements Review

**Office of Research Integrity (ORI)**
ORI will review the DUA terms and conditions; propose or negotiate terms with the other party; and, if necessary, forward to partner units for consultation (e.g., Office of General Counsel, Office of Contracts and Insurance, GW Privacy, etc.).
ORI will work with the investigator to obtain additional information that may be required (e.g., documentation from the Office of Human Research or export controlled information).

**Data Services Librarian**
Consults on submitted DUAs to support researchers in creation of data management plans that comply with requirements such as DUA terms and data protection guidance. Available for consultation on other research data management topics.

**GW Privacy**
Reviews the data received or provided as part of the research project and assess regulatory requirements for confidentiality and protection (e.g. HIPAA regulated data; personal data of individuals in the EU – GDPR regulated, etc.)
Nature of Data

Type of other party
- Please select a value...

Is GW the provider or recipient of the data to be transferred under the proposed DUA?
- Provider
- Recipient

Is the data any of the following (check all that apply)?
- Animal Data
- Human Subjects Data

Is the data for use on a sponsored project?
- Yes
- No

Nature of the data
- Please select a value...

If other or unknown, please explain the type of data you are collecting

Export Controls and Intellectual Property
Data may be subject to export controls as Technical Data under the International Traffic in Arms Regulations, or the Export Administration Regulation, or otherwise.

DUA, DMP upload

Document Uploads

Data Use Agreement (DUA)
+ Add Attachment
Please upload a pdf or Microsoft Word version of the proposed DUA

Data Management Plan (DMP)
+ Add Attachment
A Data Management Plan (DMP) is required to document that research data will be appropriately managed and protected. Please upload a pdf or Microsoft Word version of the DMP to accompany the DUA (in either draft or final state). If you do not have a DMP, use https://dmp-tool.org to create one. If assistance is required in developing a DMP, contact GW's data services librarian at libdata@gwu.edu.

Other Supporting Documents
+ Add Attachment
If desired, attach other supporting documents:
DUA - Privacy Considerations

Human subject research:
– consider the sensitivity of the information collected / received
– protect the privacy and confidentiality of data

Privacy is about people.
• a right to be protected
• in the eye of the participant, not the researcher

Consider Privacy Regulations
e.g. HIPAA, FERPA, GDPR, GLBA, etc.

Confidentiality is about data.
• extension of privacy
• agreement about maintenance and who has access to identifiable data
• in regards to HIPAA, protects patients from inappropriate disclosures of "Protected Health Information" (PHI)

e.g. NIST data destruction requirements
Data Management Plan
Planning for Structure

- Resources
- Tools
- Expertise
The Research Data Management Lifecycle
Why Prepare a DMP? (1)

• Save time
  – Less reorganization later
  – Reduce delays

• Increase efficiency and consistency
  – Ensures you and others will be able to understand and use data in future

• Safeguard Data and Access to Data
Why Prepare a DMP? (2)

- Easier to preserve your data
- Prevents duplication of effort
- Can lead to new, unanticipated discoveries
- Increases visibility of research
- Makes research and data more relevant
- Funding agency requirement
A Tool for Creating Data Management Plans

https://dmptool.org

Sign in through “Your institution”
Components of a DMP

- Information about data and data format
- Metadata content and format
- Policies—access, security, sharing, reuse
- Long-term storage and preservation
- Budget
1. **Information about data and data format**

A. Description of data to be produced
B. How data will be acquired
C. How data will be processed
D. File format
E. Quality Assurance & Control
F. Existing data

- If existing data are used, what are their origins?
- Will your data be combined with existing data?
- What is the relationship between your data and existing data?

G. How data will be managed in short-term

- Version control
- Backing up
- Security & protection
- Who will be responsible?
2. Metadata content and format

A. Metadata defined:

• Documentation and reporting of data
• Contextual details: Critical information about the dataset
• Information important for using the data
• Descriptions of temporal and spatial details, instruments, parameters, units, files, etc.
2. Metadata content and format

B. What metadata are needed
   • Any details that make data meaningful

C. How metadata will be created and/or captured
   • Lab notebooks? GPS units?
   • Auto-saved on instrument?

D. What format will be used for the metadata
   • Standards for community
   • Justification for format chosen
3. Policies—access, security, sharing, reuse

A. Obligations for sharing
- Funding agency
- Institution
- Other organization
- Legal

B. Details of data sharing
- How long?
- When?
- How access can be gained?
- Data collector rights

C. Ethical/privacy issues with data sharing
4. **Long-term storage and preservation**

A. What data will be preserved
B. Where will it be archived
   - Most appropriate archive for data
   - Community standards
C. Data transformations/formats needed
   - Consider archive policies
D. Who will be responsible
   - Contact person for archive
Questions?