

The Life-cycle of Protected Data in DesignSafe

Maria Esteva
Publish Your Data Event
August 27, 2020



Protected data

- Data that should not be disclosed to unauthorized parties.
 - Data with Personal Identifiable Information.
 - Data under FERPA, HIPPA or other federal & state government restrictions (ex. security).
 - Data with very sensitive/confidential information.
- Protected data has different levels of risk.
 - Name and food preferences.
 - Name and bank account information.

Types of identifiers in data

- **DIRECT** Information that relates specifically to an individual:
 - names, postal address information other than town or city, state, and zip code; phone numbers; fax numbers; email addresses; social security numbers; medical record numbers; health plan beneficiary numbers; account numbers; certificate/license numbers; vehicle identifiers and serial numbers including license plate numbers; device identifiers and serial numbers; URLs; IP addresses; biometric identifiers; full face photographic images and any comparable images, and passport numbers.
- **INDIRECT** Information that combined can disclose the identity of an individual:
 - place of birth, race, religion, weight, activities, employment information, political affiliation, medical information, education information, sexual orientation, profession, and financial information.

Data and risks

- Risks of data being accessed by un-authorized people.
- Risks of data being tampered with.
- Risks of sensitive information being disclosed.
- Risk of data corruption/loss.
- Risks of identity theft.
- Risks of loose security practices within a team.

Since inception DesignSafe has not had a data security incident.

Managing data in DesignSafe

- Planning
 - Storing planning documents in the Data Depot/My project
 - Gathering data in the field
 - RAPApp
 - Data can be automatically transferred to the Data Depot
 - Storing data
 - In My Data (private to one user)
 - In My Projects (share with your team)
 - Curating data
 - Publishing data
 - Reusing data
- Private/team
- Public
- Private/team

Storage for protected data in DesignSafe

- Private to one user: My Data,
- Private to a team: My Projects
- Public: Published
- Private with extra security restrictions: TACC protected storage.
 - Complies with ISO and UT Austin security standards

Private storage in DesignSafe

- Prior to publication.
 - Raw data – not curated.
 - Assess its nature and evaluate risks.
 - NO HIPPA or FERPA data.
 - NO data with national security information.
 - NO data that contains extremely sensitive information.
 - If any of the above, consider using TACC's protected data storage.
 - PII lite data.
 - De-identified data

Guidelines Regarding the Storage and Publication of Protected Data in DesignSafe-CI

Researchers should always comply with the requirements, norms and procedures approved by the Institutional Review Board (IRB) or equivalent body, regarding human subjects' data storage and publication.

Protected data includes human subjects data with Personal Identifiable Information (PII), data that is protected under HIPAA, FERPA and FISMA regulations, as well as data that involves vulnerable populations and that contains sensitive information.

Storing Protected Data

DesignSafe My Data and My Projects are secure spaces to store raw protected data as long as it is not under HIPAA, FERPA or FISMA regulations. If data needs to comply with these regulations, researchers must contact DesignSafe through a help ticket to evaluate the case and use TACC's Protected Data Service. Researchers with doubts are welcome to send a ticket or join [curation office hours](#).

Publishing Protected Data

To publish protected data researchers should adhere to the following procedures:

1. Do not publish HIPAA, FERPA, FISMA, PII data or other sensitive information in DesignSafe.
2. To publish protected data and any related documentation (reports, planning documents, field notes, etc.) it must be properly anonymized. No direct identifiers and up to three indirect identifiers are allowed. Direct identifiers include items such as participant names, participant initials, facial photographs (unless expressly authorized by participants), home addresses, social security numbers and dates of birth. Indirect identifiers are identifiers that, taken together, could be used to deduce someone's identity. Examples of indirect identifiers include gender, household and family compositions, occupation, places of birth, or year of birth/age.
3. If a researcher needs to restrict public access to data because it includes HIPAA, FERPA, PII data or other sensitive information, consider publishing metadata and other documentation about the data.
4. Users of DesignSafe interested in the data will be directed to contact the project PI or designated point of contact through a published email address to request access to the data and to discuss the conditions for its reuse.
5. Please contact DesignSafe through a [help ticket](#) or join [curation office hours](#) prior to preparing this type of data publication.

[← Back](#)

[Finish](#)

Will you or other team members upload protected data to this project?

- No, we will upload only non-confidential and/or non-personal information:
- The data has been de-identified and/or there is no Personally Identifiable Information (PII), or you received approval to publish PII from the research subjects.
 - For an example of the type of data that fits category and has been published on DesignSafe, see: <https://doi.org/10.17603/e9wq-gz57>
- Yes, we will upload sensitive personal information
- Includes any of these types of confidential information that may pose a risk if disclosed to non-authorized individuals.
 - See examples of this type of data
- Yes, we will upload very sensitive confidential information
- Examples include any type of confidential information that would cause harm to individuals if not accessed only by authorized individuals. For example, medical diagnoses records, very sensitive financial records, criminal records, data involved in issues of national security, etc.

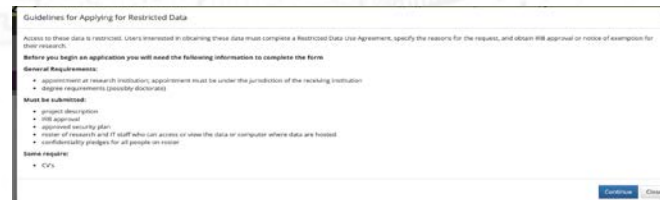
[← Back](#)

[Continue](#)

Publishing protected data

- Curated
- No HIPPA, FERPA or data under other federal constraints.
- No PII.
- Not more than three indirect identifiers that put together will not disclose identity.
- Use of keys should be explained in a data dictionary.
 - Evaluate in context with data size, geographic distribution, demographic distribution, etc.
- If the data turns not-comprehensible consider:
 - Will you share it on a person to person agreement?
 - Under what conditions?
 - Publish metadata and a file including conditions and contact information.?

From ICPSR



Guidelines for Applying for Restricted Data

Access to these data is restricted. Users interested in obtaining these data must complete a Restricted Data Use Agreement, specify the reasons for the request, and obtain IRB approval or notice of exemption for their research.

Before you begin an application you will need the following information to complete the form.

General Requirements:

- approval of research institution; approval must be under the jurisdiction of the requesting institution
- IRB requirements (provide dictionary)

Must be submitted:

- project description
- IRB approval
- approved security plan
- names of research and IT staff who have access to view the data or computer where data are hosted
- confidentiality pledges for all people on roster

Some require:

- CVs

Continue Close

Considerations during data management planning

- For storing and publishing protected data.
 - What are the most publication permissions that I can obtain from subjects considering the characteristics/ethical constraints of this research?
 - What permissions does your IRB support?
 - Considering interdisciplinary publications:
 - Can geographical location in the engineering collections disclose the identity of the subjects interviewed? Do we have permissions?

The Impact of Published Data

Maria Esteva
Publish Your Data Event
August 27, 2020



Publishing and sharing data

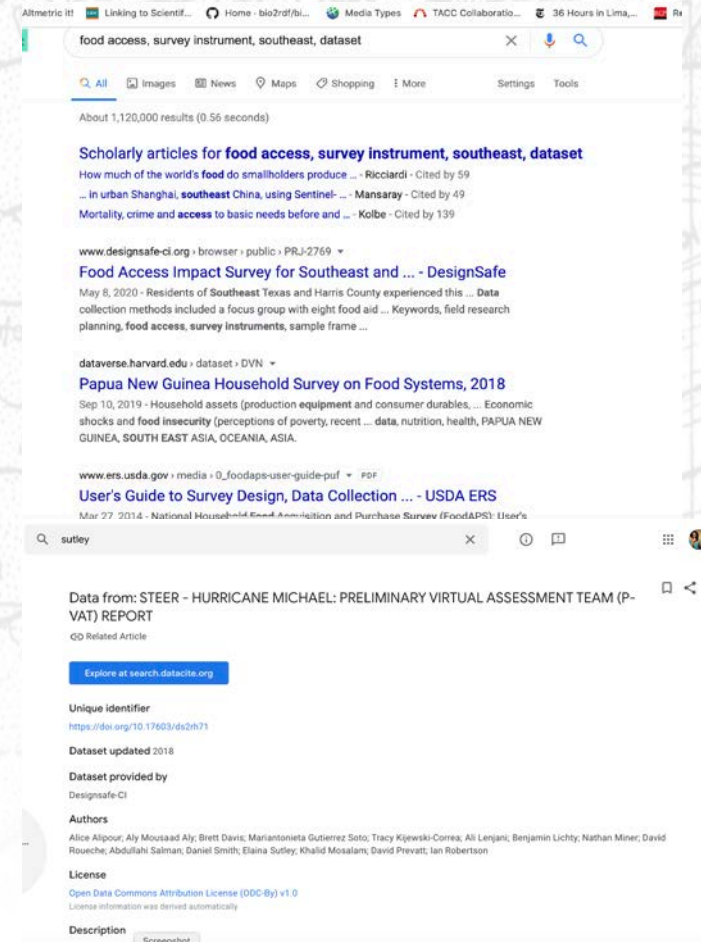
- Open data movement.
- For the social good.
- To contribute to the scientific record.
- For purposes of research validation and reliability.
- To get credit for your work.
- To promote your work.
- Consider a data citation as important as paper citations (and link them together)

Digital Object Identifiers (DOI)

- DesignSafe provides DOIs for data and documentation.
- A unique alphanumeric string that permanently resolves to the landing page URL the data is described and made available.
- Supported by technical and organizational efforts.
- PERMANENT

DOIs and data exposure

- DOIs are attached to metadata (descriptions) about your work.
- The metadata is exposed through web protocols.
- Academic aggregators and (hopefully) Google search/data/scholar indexes that metadata.
- This is how users find your data on the web using common browsers and Google Data.
- Good metadata improves the search.



Market your work

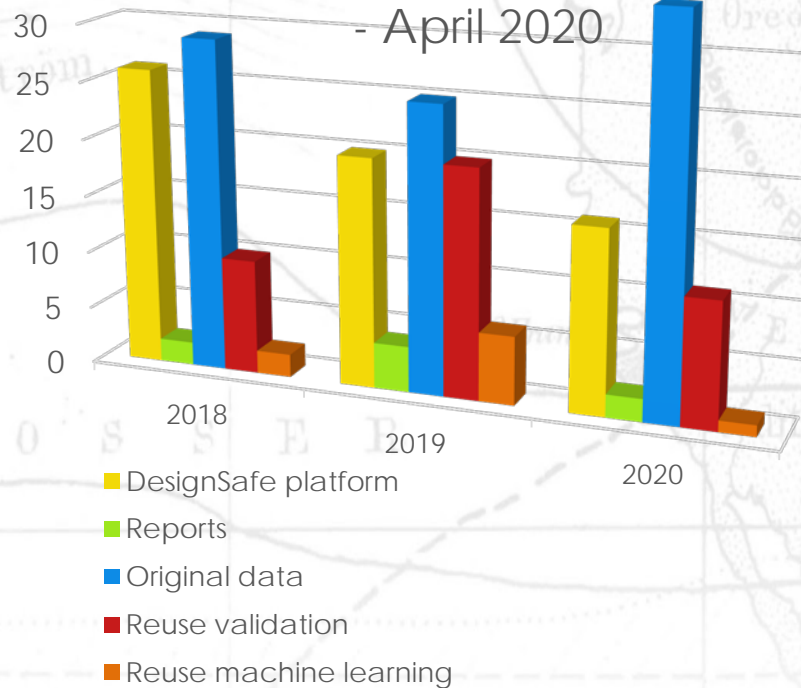
- Include the data citation in your papers in the reference section.
- Present about your data at conferences.
- Use social media to announce your data publication with the DOI.
- Use your site/page to include citation and news about your data including the DOI.
- Cross-referencing bumps your data ratings on the web searches.

Be ambassadors for all data

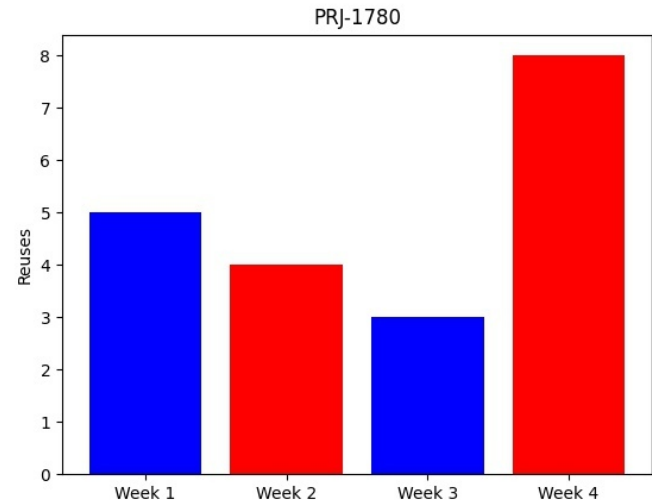
- Include the citation of data that you reuse in DesignSafe.
 - Related work
 - Related data
 - Reusing data in your projects (you already do a lot).
 - Beware of existing licenses and permissions.
 - Publish the synthetic data (derived from).
 - Reference the reused data in your publications in DesignSafe.
-

We are counting

Data Reuse - Annual Totals 2018



Citation counts per Google Scholar alert



Downloads, previews, copies and direct reuse in DesignSafe per project. (stats will be published in the future)

- Sign up for Publish your Data Event dedicated office hours.
 - <https://signup.com/go/fxHQnhr>
- Come to office hours when needed:
 - Tuesdays and Thursdays 1 to 2 pm Central

Zoom Link

<https://DesignSafe-ci.zoom.us/j/730745593>

- Use the Help ticket
- Email maria@tacc.utexas.edu