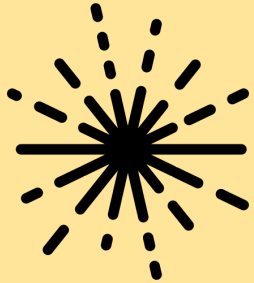


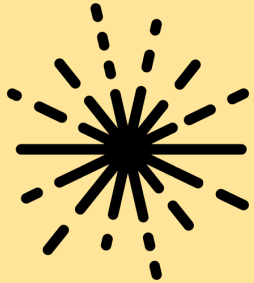
#OpenDP2020

# OpenDP Use Cases

@mercecrosas

# Uses cases envisioned for OpenDP

- Archival data repositories to offer academic researchers privacy-preserving access to sensitive data.
- Government agencies to safely share sensitive data with researchers, data-driven policy makers, and the broader public.
- Companies to share data on their users and customers with academic researchers or with institutions that bring together several such datasets.
- Collaborations between government, industry, and academia to provide greater access to and analysis of data that aids in understanding and combating the spread of COVID-19.



#OpenDP2020

# OpenDP Use Cases
# Data Repositories

@mercecrosas

# Three uses cases relevant to Data Repositories

An **OpenDP system** integrated with a data repository could be used to:

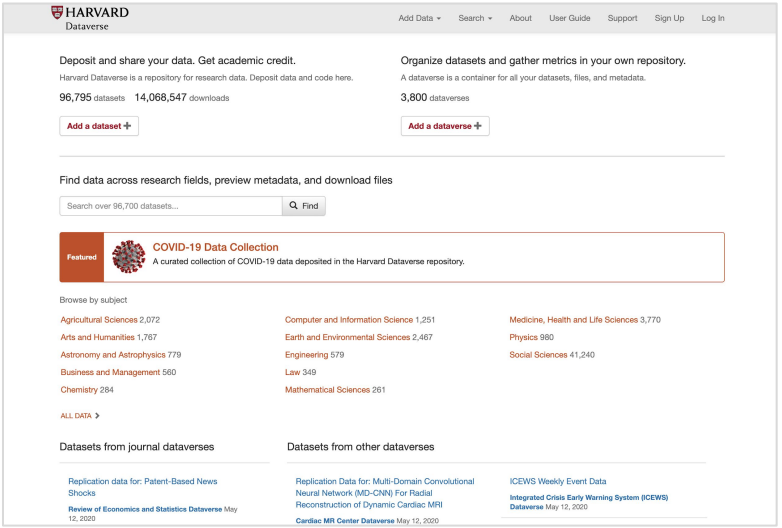
1. Enable variable search and exploration of sensitive datasets deposited in the repository
2. Facilitate reproducibility of research with sensitive datasets
3. Enable statistical analysis of sensitive datasets accessible through the repository

# Data Repository Example: Dataverse

## Dataverse Software


- Open-source software for data repositories
- 59 Dataverse repositories world-wide
- A community of users and contributors

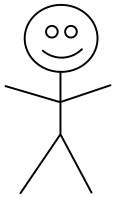
## Harvard Dataverse




- 96K searchable datasets
- 14M file downloads
- Open to all research fields

# Use Case 1: Variable search and exploration of sensitive datasets

Data Depositor



Deposit sensitive data file

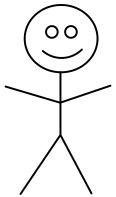
Dataverse

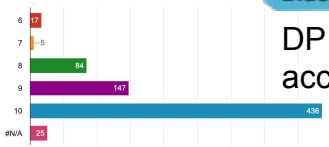
Generate DP Summary Statistics

OpenDP

1. Select privacy-loss parameter
2. Run DP query for Summary Statistics

Data User






Blue

DP summary statistics accessible

Search and explore variable metadata and summary statistics, with privacy-protection

Dataverse

OpenDP

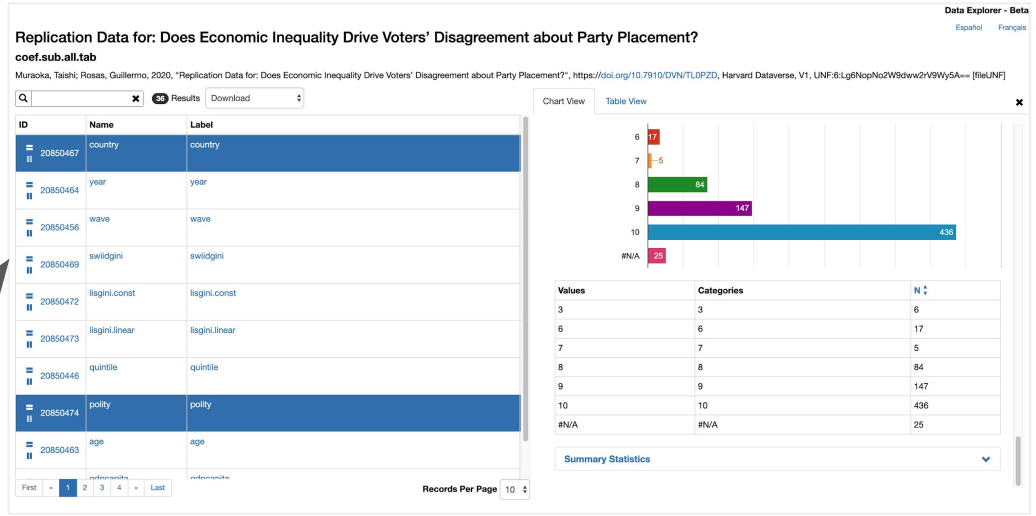
Orange

Data file **not** accessible

Return DP Summary Statistics metadata file

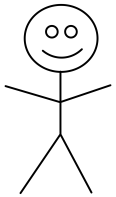
# Use Case 1: Variable search and exploration of sensitive datasets


Data Explorer uses DP summary statistics generated by OpenDP



# Use Case 2: Reproducibility of research with sensitive datasets

Data Author



Deposit a sensitive *replication* data file

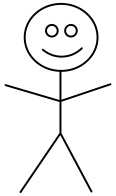
Dataverse

Generate DP release to be used to reproduce results

OpenDP

1. Select privacy-loss parameter
2. Select DP query to be released

Data Reviewer (Journal)



Use DP release to verify results in journal article

Blue

DP *replication* release accessible

Dataverse

Orange

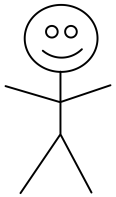
Data file **not** accessible

Return DP replication release

OpenDP

# Use Case 3: Statistical analysis of sensitive datasets

Data Depositor



Deposit a sensitive data file

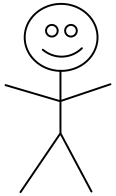
Dataverse

Enable DP statistics for this data file

OpenDP

1. Select privacy-loss parameter
2. Configure privacy-loss budget

Data User



OpenDP Interface

Run any statistics using OpenDP interface, limited by privacy-loss budget

Orange

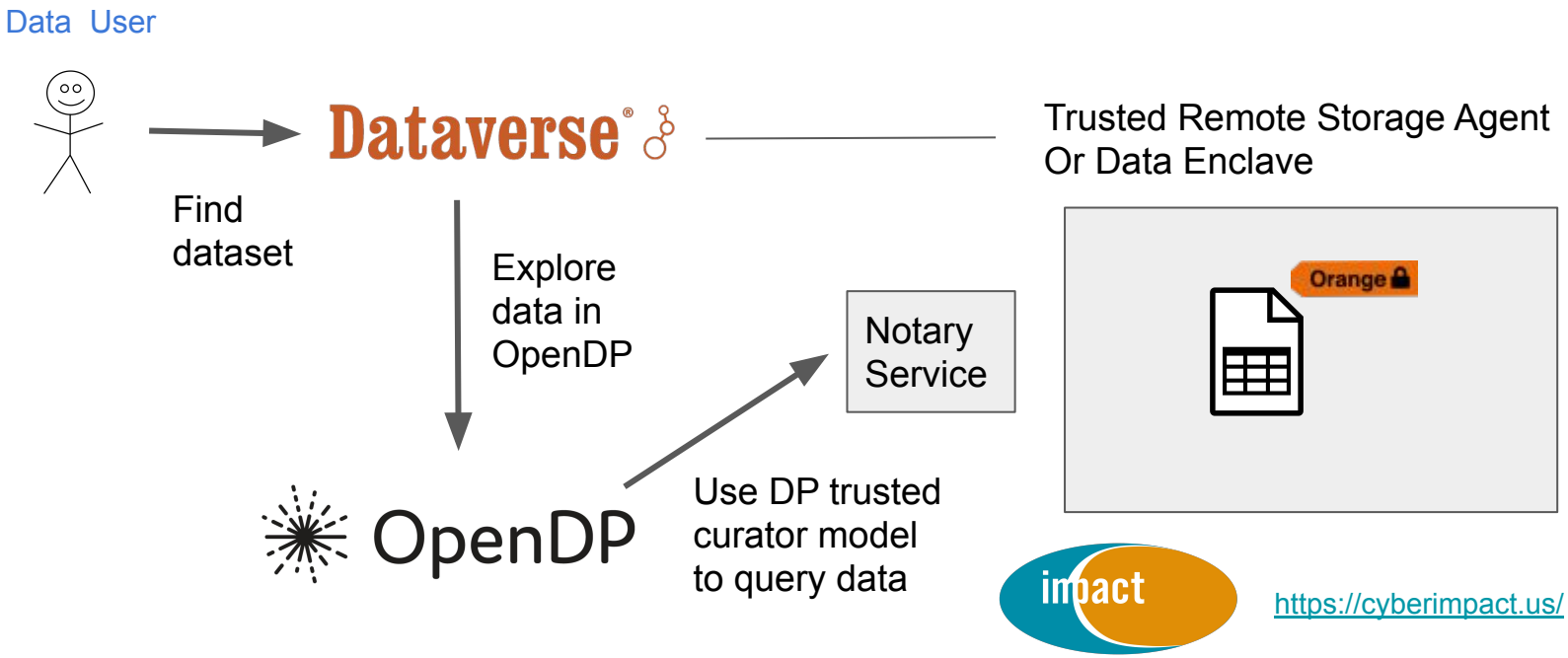
Data file **not** accessible

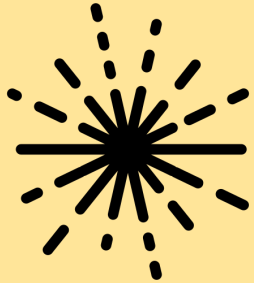
Dataverse

Turn on OpenDP query interface for end-user

OpenDP

# Dataverse and OpenDP integration?





#OpenDP2020

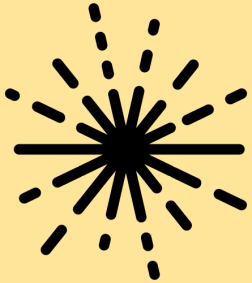
# OpenDP Use Cases Government Agencies

@mercecrosas

# Use Cases relevant to Government Agencies

An increasing number of datasets useful to policy makers and researchers are being collected by government agencies. When these datasets are sensitive, OpendDP could be used to:

- Produce rich statistical summaries of sensitive datasets to be shared widely without risk of revealing individual-level information
- Generate differentially private “synthetic data” that reflects many statistical properties of the original dataset
- Allow approved researchers to use the OpenDP query interface to run differentially private analysis of interest on the government agency data



#OpenDP2020

# Thanks

@mercecrosas