SHARING SENSITIVE DATA WITH CONFIDENCE:
THE DATATAGS SYSTEM

Mercè Crosas, Ph.D.
Chief Data Science and Technology Officer
IQSS
Harvard University

Michael Bar-Sinai
PhD candidate in Computer Science at the Ben-Gurion University of the Negev, Israel
Fellow at the Institute for Quantitative Social Science at Harvard University.

Latanya Sweeney
Professor of Government and Technology in Residence
Director of Data Privacy Lab
Harvard University
Data sharing: good for you and good for the world
Data sharing: good for you and good for the world

Researchers ➔ Get credit for their data
Data sharing: good for you and good for the world

Researchers → Get credit for their data

Publishers and Journals → Verify published work
Data sharing:
good for you and good for the world

- Researchers
  - Get credit for their data
- Publishers and Journals
  - Verify published work
- Federal funding agencies
  - Make public assets accessible
Data sharing: 
good for you and good for the world

- Researchers
  - Get credit for their data

- Publishers and Journals
  - Verify published work

- Federal funding agencies
  - Make public assets accessible

- Science
  - Validate, reuse and extend previous work
Open-source software developed at Harvard’s IQSS since 2006
Used to share, publish, cite and archive research data
Installed in 12 sites world wide
Serving 100s of universities and organizations
Harvard Dataverse: dataverse.harvard.edu

Started as a community repository for Social Science
Now open to all research fields and all researchers
More than 1300 dataverses
More than 59,000 datasets
More than 1,500,000 downloads
Data Repositories vs Repository Software

- **Domain-specific repositories**
  - GenBank
  - WW Protein Data Bank
  - SBGrid Data Bank
  - ...

- **General-purpose repositories**
  - Harvard Dataverse
  - DataDryad
  - Figshare
  - ...

- **Repository Software**
  - Dataverse Software
  - Dspace
  - Fedora
  - ...
Data Repositories vs Repository Software

Domain-specific repositories
- GenBank
- WW Protein Data Bank
- SBGrid Data Bank

General-purpose repositories
- Harvard Dataverse
- DataDryad
- Figshare

Repository Software
- Dataverse Software
- Dspace
- Fedora
- ...
- ...
- ...
- ...
Data Repositories vs Repository Software

Domain-specific repositories
- GenBank
- WW Protein Data Bank
- SBGrid Data Bank
- ...

General-purpose repositories
- Harvard Dataverse
- DataDryad
- Figshare
- ...

Repository Software
- Dataverse Software
- Dspace
- Fedora
- ...
Data Repositories vs Repository Software

Domain-specific repositories
- GenBank
- WW Protein Data Bank
- SBGrid Data Bank
- ...

General-purpose repositories
- Harvard Dataverse
- DataPryad
- Figshare
- ...

Repository Software
- Dataverse Software
- Dspace
- Fedora
- ...

...
But, existing community repositories do not support sensitive data
“User Uploads must be void of all identifiable information, such that re-identification of any subjects from the amalgamation of the information available from all of the materials (across datasets and dataverses) uploaded under any one author and/or user should not be possible.”
“Submitter represents and warrants that the Content does not contain any information (i) which identifies, or which can be used in conjunction with other publicly available information to personally identify, any individual;”
“If you are submitting human sequences to GenBank, do not include any data that could reveal the personal identity of the source. It is our assumption that you have received any necessary informed consent authorizations that your organizations require prior to submitting your sequences.”
HOW CAN WE MAXIMIZE SHARING SENSITIVE DATA WHILE BEING MINDFUL OF PRIVACY?
### Abstract

- **Tag Type**
- **Description**
- **Security Features**
- **Access Credentials**

<table>
<thead>
<tr>
<th>Tag Type</th>
<th>Description</th>
<th>Security Features</th>
<th>Access Credentials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>Public</td>
<td>Clear storage, Clear transmit</td>
<td>Open</td>
</tr>
<tr>
<td>Green</td>
<td>Controlled public</td>
<td>Clear storage, Clear transmit</td>
<td>Email- or OAuth Verified Registration</td>
</tr>
<tr>
<td>Yellow</td>
<td>Accountable</td>
<td>Clear storage, Encrypted transmit</td>
<td>Password, Registered, Approval, Click-through DUA</td>
</tr>
<tr>
<td>Orange</td>
<td>More accountable</td>
<td>Encrypted storage, Encrypted transmit</td>
<td>Password, Registered, Approval, Signed DUA</td>
</tr>
<tr>
<td>Red</td>
<td>Fully accountable</td>
<td>Encrypted storage, Encrypted transmit</td>
<td>Two-factor authentication, Approval, Signed DUA</td>
</tr>
<tr>
<td>Crimson</td>
<td>Maximally restricted</td>
<td>Multi-encrypted storage, Encrypted transmit</td>
<td>Two-factor authentication, Approval, Signed DUA</td>
</tr>
</tbody>
</table>

- We introduce datatags as a means of specifying security and access requirements for sensitive data.
- The datatags approach reduces the complexity of thousands of data-sharing regulations to a small number of tags.
- We show implementation details for medical and educational data and for research and corporate repositories.

*Definitions for each of six ordered Blue to Crimson sample datatags.*

---

A **datatag** is a set of security features and access requirements for file handling.
A **datatag** is a set of security features and access requirements for file handling.

A **datatags repository** is one that stores and shares data files in accordance with a standardized and ordered levels of security and access requirements.
A DataTags Repository must satisfy the following conditions:
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
   a. additional requirements cannot weaken the file security
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
   a. additional requirements cannot weaken the file security
   b. and cannot required the same or more security than a more restrictive datatag
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
   a. additional requirements cannot weaken the file security
   b. and cannot required the same or more security than a more restrictive datatag
3. A recipient of a file from the repository must:
A DataTags Repository must satisfy the following conditions:

1. Supports more than one data tag
2. Each file in the repository must have one and only one data tag
   a. additional requirements cannot weaken the file security
   b. and cannot required the same or more security than a more restrictive data tag
3. A recipient of a file from the repository must:
   a. satisfy file’s access requirements,
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
   a. additional requirements cannot weaken the file security
   b. and cannot required the same or more security than a more restrictive datatag
3. A recipient of a file from the repository must:
   a. satisfy file’s access requirements,
   b. produce sufficient credentials as requested,
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
   a. additional requirements cannot weaken the file security
   b. and cannot required the same or more security than a more restrictive datatag
3. A recipient of a file from the repository must:
   a. satisfy file’s access requirements,
   b. produce sufficient credentials as requested,
   c. and agree to any terms of use required to acquire the file.
A DataTags Repository must satisfy the following conditions:

1. Supports more than one datatag
2. Each file in the repository must have one and only one datatag
   a. additional requirements cannot weaken the file security
   b. and cannot required the same or more security than a more restrictive datatag
3. A recipient of a file from the repository must:
   a. satisfy file’s access requirements,
   b. produce sufficient credentials as requested,
   c. and agree to any terms of use required to acquire the file.
4. Provides technological guarantees for requirements 1, 2 and 3.
## Datatags Levels

<table>
<thead>
<tr>
<th>Tag Type</th>
<th>Description</th>
<th>Security Features</th>
<th>Access Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>Public</td>
<td>Clear storage, Clear transmission</td>
<td>Open</td>
</tr>
<tr>
<td>Green</td>
<td>Controlled public</td>
<td>Clear storage, Clear transmission</td>
<td>Email, OAuth verified registration</td>
</tr>
<tr>
<td>Yellow</td>
<td>Accountable</td>
<td>Clear storage, Encrypted transmit</td>
<td>Password, Registered, Approval, Click DUA</td>
</tr>
<tr>
<td>Orange</td>
<td>More accountable</td>
<td>Encrypted storage, Encrypted transmit</td>
<td>Password, Registered, Approval, Signed DUA</td>
</tr>
<tr>
<td>Red</td>
<td>Fully accountable</td>
<td>Encrypted storage, Encrypted transmit</td>
<td>Two-factor authentication, Approval, Signed DUA</td>
</tr>
<tr>
<td>Crimson</td>
<td>Maximally restricted</td>
<td>MultiEncrypt store, Encrypted transmit</td>
<td>Two-factor authentication, Approval, Signed DUA</td>
</tr>
</tbody>
</table>
DATATAGS WITH HARVARD DATAVERSE
DataTags vs Harvard Security Levels

- **Level 1:**
  - No sensitive data; open data
- **Level 1:**
  - De-identified data
- **Level 2:**
  - Confidential information by University standards; no material harm
- **Level 3:**
  - Confidential information that could cause material harm (non-level 4 FERPA)
- **Level 4:**
  - High risk confidential information (SSN)
- **Level 5:**
  - Information that would cause severe harm
Dataverses, Datasets, Data Files and DataTags

A Datatag is assigned to each Data File (not to the Dataset)
DataTags Workflow with Dataverse

http://datatags.org
http://privacytools.seas.harvard.edu
DataTags Workflow with Dataverse

Data File Ingestion

http://datatags.org
http://privacytools.seas.harvard.edu
DataTags Workflow with Dataverse

Data File Ingestion → Automatic Interview → Review Board Approval

Blue → Green → Yellow → Orange → Red → Crimson

http://datatags.org
http://privacytools.seas.harvard.edu
DataTags Workflow with Dataverse

Data File Ingestion

Automatic Interview

Review Board Approval

Blue
Green
Yellow
Orange
Red
Crimson

http://datatags.org
http://privacytools.seas.harvard.edu
DataTags Workflow with Dataverse

Data File Ingestion

Automatic Interview

Review Board Approval

Blue

Green

Yellow

Orange

Red

Crimson

Sensitive Dataset

http://datatags.org
http://privacytools.seas.harvard.edu
DataTags Workflow with Dataverse

Data File Ingestion → Automatic Interview → Review Board Approval → Direct Access

Sensitive Dataset → Authorized Signed DUA

http://datatags.org
http://privacytools.seas.harvard.edu
DataTags Workflow with Dataverse

Data File Ingestion → Review Board Approval → Sensitive Dataset → Direct Access → Authorized Signed DUA → Privacy Preserving Access

Automatic Interview

Blue
Green
Yellow
Orange
Red
Crimson

http://datatags.org
http://privacytools.seas.harvard.edu
A Curator Model for Privacy-Preserving Analysis

Differentially Private statistics (summaries, causal inference, regression, interactive queries)

Acknowledgement: Honaker, J. and Nissim, K., Data Privacy Tools Project
Credentials and Retrieval in Dataverse

- **Blue**
  - Data File not restricted
  - Guestbook – Email to access

- **Green**
  - Data File restricted; Dataverse/InCommon account; Request access; Click DUA

- **Yellow**
  - Data File restricted; Dataverse/InCommon account; Request access; Sign DUA

- **Orange**
  - Data File restricted; InCommon account; Request access; Sign DUA

- **Red**
  - Data File restricted; InCommon account; Request access; Two-Factor authentication

- **Crimson**
  - Sign DUA
OTHER TYPE OF DATATAGS REPOSITORIES
Betty: Sole Researcher

- Received consent from participants
- Repository for sharing highly sensitive data (not necessarily Harvard Dataverse)
<table>
<thead>
<tr>
<th>Ingestion and Decision-making Knowledge</th>
<th>IRB determination or an interview system.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Codification and Infrastructure</td>
<td>Blue, Green, Yellow, Orange, Red, Crimson.</td>
</tr>
<tr>
<td>Credentials and Retrieval</td>
<td>Different files may additionally require specific terms of use based on legal or regulatory requirements or adopted best practices.</td>
</tr>
</tbody>
</table>

(Same use case as Dataverse)
Adam: Large Medical Research Group

- Repository for sharing local data
- Repository for published data
- Repository for sharing with collaborators
Adam: Large Medical Research Group
Diane: Multinational Corporation

- Cloud contains data from all over the world, collected under a variety of terms, subject to different laws
- Repository that enforces requirements on employee access
Diane: Multinational Corporation
Charles: Institutional Review Board

- Document committee decisions
- Recommend handling based on prior decisions
Charles: Institutional Review Board
How technology impacts humans.
How technology impacts humans.
How technology impacts humans.

DATA

Direct Tagging

Direct Deposit

Editorial Board

Open.
Agree to cite.

Register email.
Agree to handling.

Confirm email.
Be approved.
Agree to handling.

Confirm email.
Be approved.
Sign agreement.

Confirm email, phone.
Be approved.
Sign agreement.

Confirm email, phone.
Be approved.
Sign agreement.
Facebook's Privacy Incident Response: a study of geolocation sharing on Facebook Messenger

Aran Khanna

- In 2012, a media outlet reported that Facebook Messenger shared personal geolocations by default.
- In 2015, my demonstration displayed Facebook's shared data on a map; it was downloaded over 85,000 times.
- After 9 days of news coverage, Facebook released an update that requires a user's permission to share geolocations.

News coverage by day

Facebook's Privacy Incident Response: a study of geolocation sharing on Facebook Messenger

Aran Khanna

News coverage by day

Days after extension release and blog post

0 1 2 3 4 5 6 7

News articles per day

3 10 28 43 15 17 12

Khanna, Aran, 2015, "Replication Data for: Facebook's Privacy Incident Response, a study of geolocation sharing on Facebook Messenger", http://dx.doi.org/10.7910/DVN/D2SNRI, Harvard Dataverse, V1

[UNF:8:hiXa2Odze5wPt9CL8yBGHDA==]

If you use these data, please add this citation to your scholarly resources. Learn about Data Citation Standards.

Description
This dataset was used for this paper published on 8/11/2015 on Technology Science http://techscience.org/a/2015081101/

Subject
Computer and Information Science

20 Files

bing_news_search_results_1.docx

bing_news_search_results_2.docx
De-anonymizing South Korean Resident Registration Numbers Shared in Prescription Data

Latanya Sweeney and Ji Su Yoo

<table>
<thead>
<tr>
<th>Letter</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
</tr>
<tr>
<td>c</td>
<td>3</td>
</tr>
<tr>
<td>d</td>
<td>4</td>
</tr>
<tr>
<td>e</td>
<td>5</td>
</tr>
<tr>
<td>f</td>
<td>6</td>
</tr>
<tr>
<td>g</td>
<td>7</td>
</tr>
<tr>
<td>h</td>
<td>8</td>
</tr>
<tr>
<td>i</td>
<td>9</td>
</tr>
<tr>
<td>j</td>
<td>0</td>
</tr>
</tbody>
</table>

Odd-digit

<table>
<thead>
<tr>
<th>Letter</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>0</td>
</tr>
<tr>
<td>g</td>
<td>9</td>
</tr>
<tr>
<td>h</td>
<td>8</td>
</tr>
<tr>
<td>i</td>
<td>7</td>
</tr>
<tr>
<td>j</td>
<td>6</td>
</tr>
<tr>
<td>k</td>
<td>5</td>
</tr>
<tr>
<td>l</td>
<td>4</td>
</tr>
<tr>
<td>m</td>
<td>3</td>
</tr>
<tr>
<td>n</td>
<td>2</td>
</tr>
<tr>
<td>o</td>
<td>1</td>
</tr>
</tbody>
</table>

Even-digit

- South Korea’s national identifier, the Resident Registration Number (RRN) includes encoded demographic information and a checksum with a publicly known pattern.
- We conducted two de-anonymization experiments on 23,163 encrypted RRNs from prescription data of South Koreans.
- We demonstrate the data’s vulnerability to de-anonymization by revealing all 23,163 unencrypted RRNs in both experiments.

Published 2015-09-29

DATATAGGING TOOLS
A Datatagging tool needs:

- **Formal description of a Datatag**
  - Capture the data handling policy of the tag
  - Capture the “stricter-than” ordering

- **Interview creation tool**
  - Support user-friendly interviews
  - Decide on the datatag based on the answers only
Formal Description of a Datatag

• Model data handling policies as a set of orthogonal aspects
  – Storage encryption, access requirements...

• Describe implementation options for each aspect; order implementations from lenient to strict
  – Clear < Encrypted < Multi Encrypt
Data Handling Policy Space

![Data Handling Policy Space Diagram]
Data Handling Policy Space

<table>
<thead>
<tr>
<th>Storage</th>
<th>None</th>
<th>Email/OAuth</th>
<th>Password</th>
<th>Two Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>MultiEncrypt</td>
<td>Green Tag's Compliance Subspace</td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
</tr>
<tr>
<td>Encrypted</td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
</tr>
<tr>
<td>Clear</td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
<td><img src="green_tag.png" alt="" /></td>
</tr>
</tbody>
</table>
Tags: TagsSpace file (.ts)

- Describe a tag space
- Convenience features: hierarchy, “slots” of different types, top-down design support, comments...

```
Authentication: some of
- None  [Available to anonymous individuals.],
- Email [Available to individuals with verified email address.],
- OAuth [Available to individuals with verified online identity or a mobile phone.],
- Password [Available to individuals having a password accounts on system.]

DataType: consists of
- Effort, Harm.

Effort: one of anonymous, deidentified, identifiable, identified.
Harm: one of noRisk, minimal, shame, civil, criminal, maxControl.
```
Comprehension Aid: Visualization
Comprehension Aid: Visualization
Comprehension Aid: Visualization
Finding the Right Tag – Decision Graph

• Directed, Acyclic Graph
• Node Types:
  – Ask
  – Set
  – Convenience: Call, End, Reject, Todo
Finding the Right Tag – Decision Graph

```xml
<*> Sample toy interview
<*><
[call: ensureLegality]
[ask:
  {text: Do the data contains personally identifiable information?}
  {terms: Personally identifiable information: Any information about an individual... }
  {answers:
    {yes: [set: Storage=encrypt; Transfer=encrypt]}
    {no: [set: Storage=clear; Transfer=clear]}}]
[todo: Test for additional ...] -- Issue #42 follows
[end]

> ensureLegality< ask:
  {text: Did you get parental consent?}
  {answers:
    {no: [reject: Must get parental consent before collecting data from subjects under 18.]]}
[end]
```

Screenshot from actual Atom package: Gal Maman, Matan Toledano, BGU
Interview Visualization

Interview credit: The Data Privacy Lab @ Harvard (Latanya Sweeney, Sean Hooley), Berkman Ctr. for Internet and Society (Alexandra Wood, David O’Brien, clinical students), IQSS (Mercè Crosas, Michael Bar-Sinai). Part of the Privacy tools for sharing research data project.
Interview Visualization

Interview credit: The Data Privacy Lab @ Harvard (Latanya Sweeney, Sean Hooley), Berkman Ctr. for Internet and Society (Alexandra Wood, David O’Brien, clinical students), IQSS (Mercè Crosas, Michael Bar-Sinai). Part of the Privacy tools for sharing research data project
Interview Visualization

Interview credit: The Data Privacy Lab @ Harvard (Latanya Sweeney, Sean Hooley), Berkman Ctr. for Internet and Society (Alexandra Wood, David O’Brien, clinical students), IQSS (Mercè Crosas, Michael Bar-Sinai). Part of the Privacy tools for sharing research data project.
Interview on the Web
Interview on the Web
Interview on the Web
Interview on the Web

Your dataset is tagged as Orange.

May include sensitive, identifiable personal information, shared with verified and/or approved recipients under agreement.

DataTags

Legal

MedicalRecords

HIPAA safeHarborDeidentified

EducationRecords

PPRA protectedDeidentified consent

ContractOrPolicy no

GovernmentRecords

DPPA highlyRestricted

Code orange

Assertions
Interview on the Web

DataTags

Dataset Can be Accepted

Your dataset is tagged as Orange

May include sensitive, identifiable personal information, shared with verified and/or approved recipients under agreement.

Legal

MedicalRecords

HIPAA safeHarborDeidentified

EducationRecords

PPRA protectedDeIdentified consent

ContractOrPolicy no

GovernmentRecords

DPPA highlyRestricted

Code orange

Interview available at datatags.org
Decision Graph Points
Decision Graph Points

• Familiar “interview with a specialist” metaphor
Decision Graph Points

- Familiar “interview with a specialist” metaphor
- Implicitly describe logic inference
Decision Graph Points

• Familiar “interview with a specialist” metaphor
• Implicitly describe logic inference
Decision Graph Points

• Analysis: Detection of Independent parts
Decision Graph Points

- Analysis: Detection of Independent parts
Decision Graph Points

• Analysis: Detection of Independent parts
• Queries, such as “what series of answers will create a datatags that allows clear storage?”
Decision Graph Points

• Optimizations

Example created by Eyal Ben-Simon, BGU
Decision Graph Points

- Optimizations

Example created by Eyal Ben-Simon, BGU
Decision Graph Points

• Optimizations

Example created by Eyal Ben-Simon, BGU
State of the Tags Tool

- Open-source project at GitHub
- Language getting more tools and feature
  - Project with BGU students
- Language Tools in progress
  - Inspectors, Visualizers, CLI development tool
- Tutorials and reference
  [datatagginglibrary.readthedocs.org](https://datatagginglibrary.readthedocs.org)
- Collaboration via, e.g. GitHub
State of the Tags Tool

- Open-source project at GitHub
- Language getting more tools and feature
  - Project with BGU students
- Language Tools in progress
  - Inspectors, Visualizers, CLI development tool
- Tutorials and reference
  datatagginglibrary.readthedocs.io
- Collaboration via, e.g. GitHub
State of the Tags Tool

• Open-source project at GitHub

• Language getting more tools and feature
  – Project with BGU students

• Language Tools in progress
  – Inspectors, Visualizers, CLI development tool

• Tutorials and reference
datatagginglibrary.readthedocs.org

• Collaboration via, e.g. GitHub
State of the Tags Tool

• Open-source project at GitHub

• Language getting more tools and feature
  – Project with BGU students

• Language Tools in progress
  – Inspectors, Visualizers, CLI development tool

• Tutorials and reference
datatagginglibrary.readthedocs.org

• Collaboration via, e.g. GitHub
State of the Tags Tool

• Open-source project at GitHub

• Language getting more tools and feature
  – Project with BGU students

• Language Tools in progress
  – Inspectors, Visualizers, CLI development tool

• Tutorials and reference
  datatagginglibrary.readthedocs.org

• Collaboration via, e.g.
  GitHub
State of the Tags Tool

• Open-source project at GitHub
• Language getting more tools and feature
  – Project with BGU students
• Language Tools in progress
  – Inspectors, Visualizers, CLI development tool
• Tutorials and reference
  datatagginglibrary.readthedocs.org
• Collaboration via, e.g. GitHub
State of the Tags Tool

• Open-source project at GitHub

• Language getting more tools and feature
  – Project with BGU students

• Language Tools in progress
  – Inspectors, Visualizers, CLI development tool

• Tutorials and reference
datatagginglibrary.readthedocs.org

• Collaboration via, e.g. GitHub
State of the Tags Tool

- Open-source project at GitHub
- Language getting more tools and feature
  - Project with BGU students
- Language Tools in progress
  - Inspectors, Visualizers, CLI development tool
- Tutorials and reference
datatagginglibrary.readthedocs.org
- Collaboration via, e.g.
  GitHub
State of the Tags Tool

- Open-source project at GitHub
- Language getting more tools and feature
  - Project with BGU students
- Language Tools in progress
  - Inspectors, Visualizers, CLI development tool
- Tutorials and reference
  [datatagginglibrary.readthedocs.org](datatagginglibrary.readthedocs.org)
- Collaboration via, e.g. GitHub
State of the Tags Tool

• Open-source project at GitHub

• Language getting more tools and feature
  – Project with BGU students

• Language Tools in progress
  – Inspectors, Visualizers, CLI development tool

• Tutorials and reference
  datatagginglibrary.readthedocs.org

• Collaboration via, e.g.
  GitHub
State of the Tags Tool

- Open-source project at GitHub

- Language tools and features:
  - Project with BGU students
  - Inspectors, Visualizers, CLI development tool
  - Tutorials and reference at datatagging.readthedocs.org

- Collaboration via, e.g. GitHub
State of the Tags Tool

- Open-source project at GitHub
- Language getting more tools and feature
  - Project with BGU students
- Language Tools in progress
  - Inspectors, Visualizers, CLI development tool
- Tutorials and reference: datatagginglibrary.readthedocs.org
- Collaboration via, e.g. GitHub
Future of the Tags Tool

• Update web interview application
  – Include upload and inspection features

• On-line collaboration environment
  – A-la Google docs?
Mercè Crosas, Michael Bar-Sinai, Latanya Sweeney

THANKS