

Lecture 4: Responsible Data Lifecycles

CMSC 25910

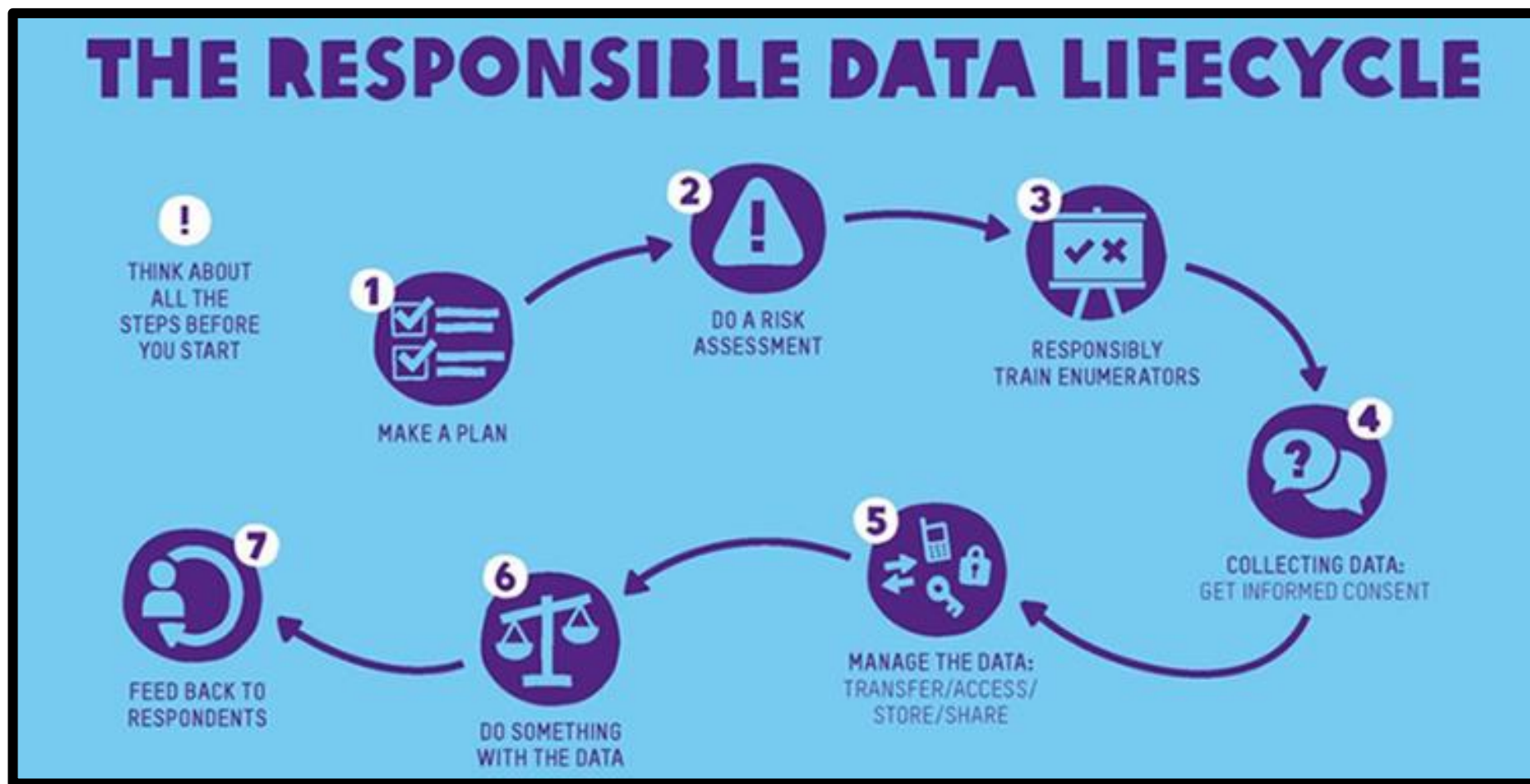
Spring 2023

The University of Chicago

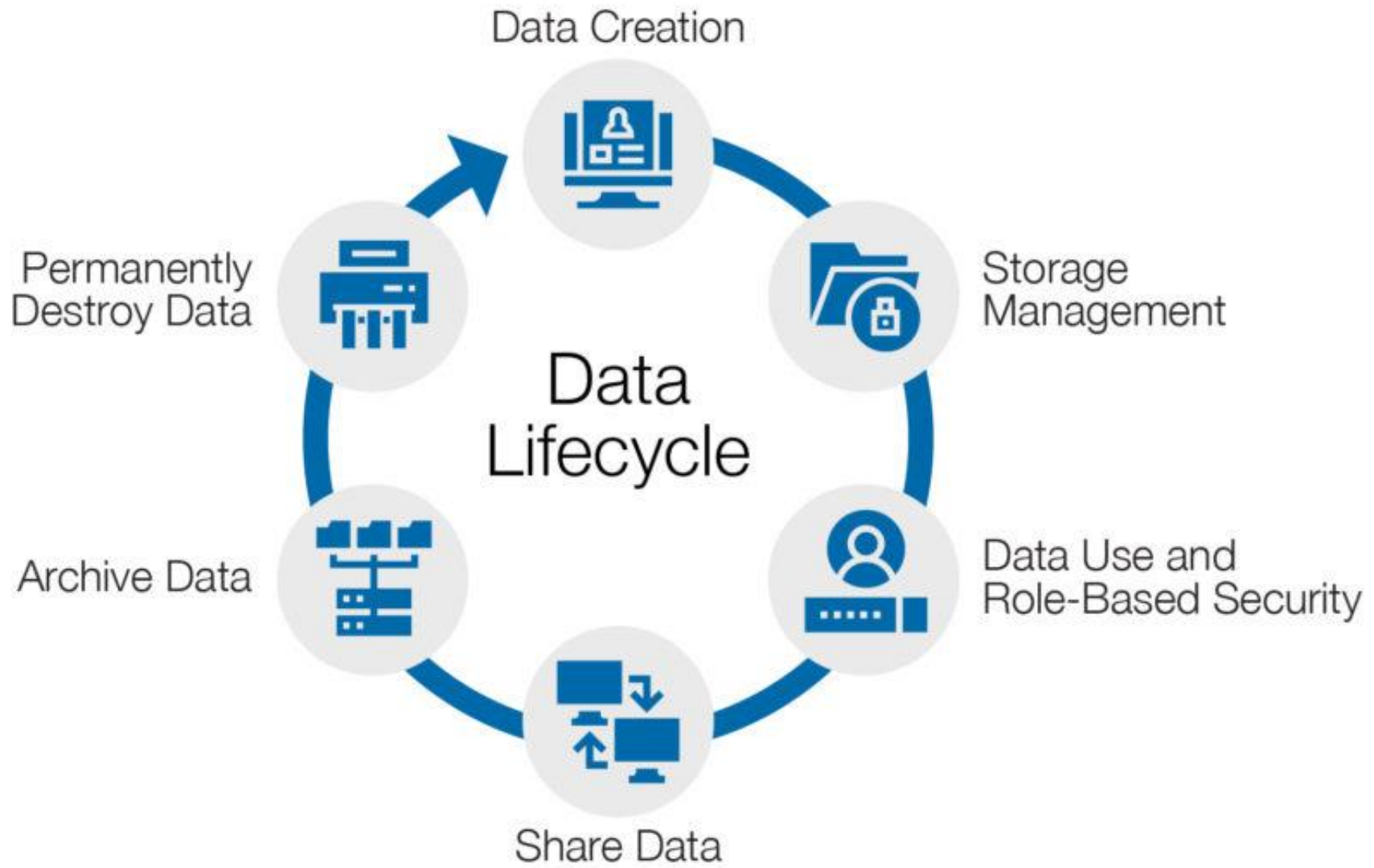


THE UNIVERSITY OF
CHICAGO

“90% of the data in the world today has been created in the last two years. Possibly the greatest challenge of the information age is how to manage data properly. As data is increasingly used for needs assessments, feedback, accountability and monitoring; management of data is a particularly crucial challenge in humanitarian response.”



Graphic and quote from Oxfam: <https://views-voices.oxfam.org.uk/2017/04/putting-responsible-data-into-practice/>



Graphic from <https://www.spirion.com/data-lifecycle-management/>

Data Repurposing

Repurposing Google data

- Google Buzz was a social network that replaced Google Wave
- Users automatically opted into Buzz, with weak privacy settings, based on Gmail data
- Automatically add as friends people you talk to frequently on Google Chat (Gchat)



Repurposing data

- Food delivery services have added restaurants to their site using their public menus (but without permission)

Banner ads



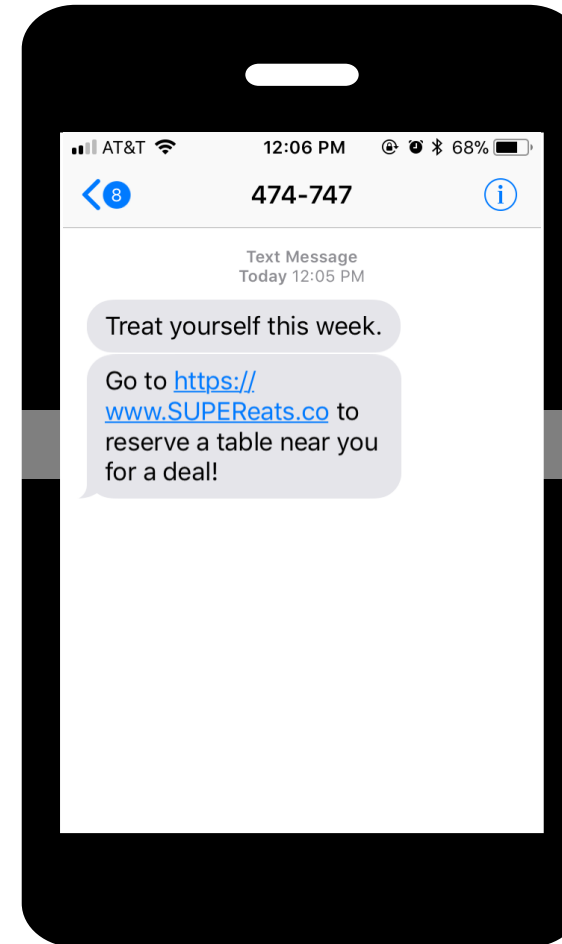
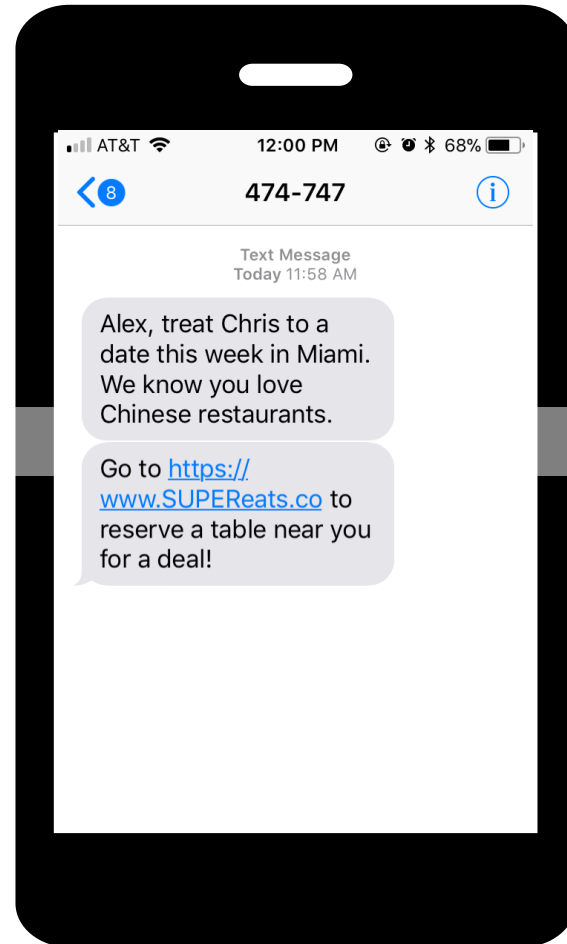
Personalized



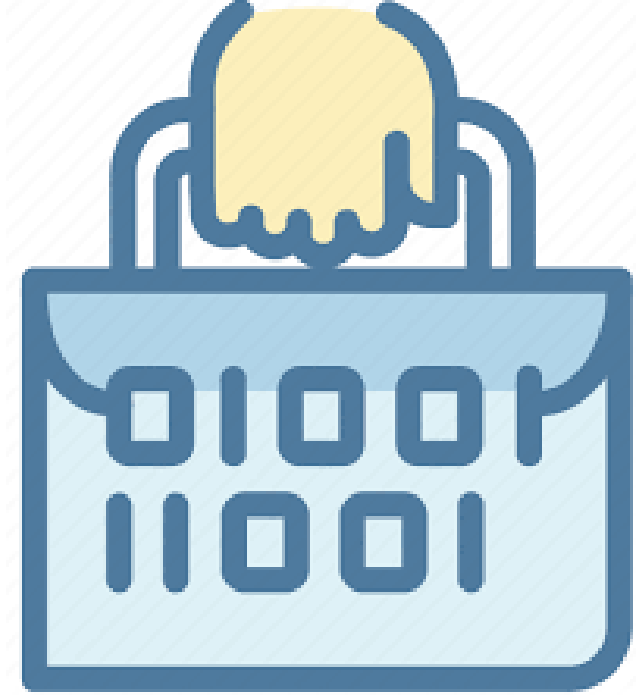
Generic

Robotext ads

Personalized



Generic



Data Portability

Art. 20 GDPR

Right to data portability

1. The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:

Data Access



Art. 15 GDPR

Right of access by the data subject

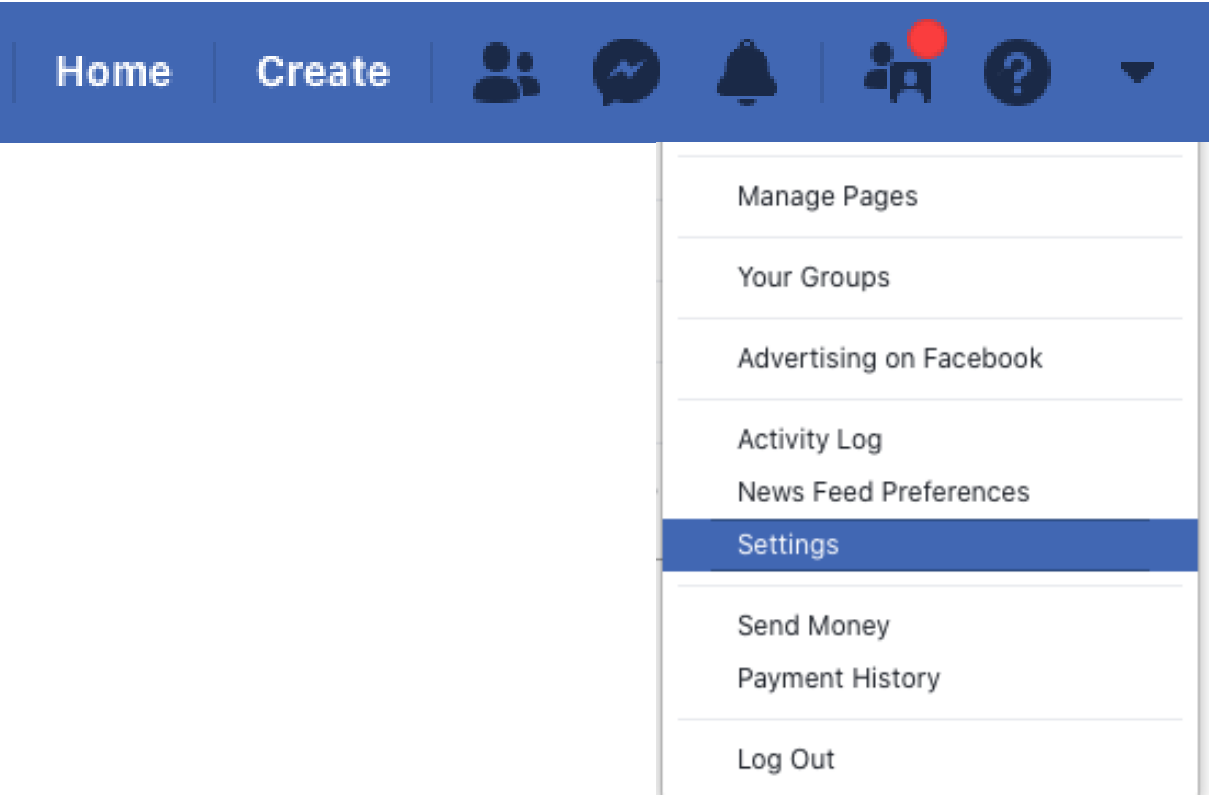
The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:

¹The controller shall take appropriate measures to provide any information referred to in Articles 13 and 14 and any communication under Articles 15 to 22 and 34 relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child. ²The information shall be provided in writing, or by other means, including, where appropriate, by electronic means. ³When requested by the data subject, the information may be provided orally, provided that the identity of the data subject is proven by other means.

14.3 How to exercise your rights

To exercise your rights, email data-privacy@buzzfeed.com. In order to protect your privacy and that of others, we will ask you to prove your identity before we take any steps in response to a request you have made. If a third party is making a request on your behalf, we will ask them to prove that they have your permission to act for you.







-  General
-  Security and Login

 Your Facebook Information

-  Privacy
-  Timeline and Tagging
-  Stories
-  Location
-  Blocking
-  Language and Region
-  Face Recognition

Access Your Information	View your information by category.	View
Download Your Information	Download a copy of your information to keep, or to transfer to another service.	View
Activity Log	View and manage your information and some settings.	View
Off-Facebook Activity	View or clear activity from businesses and organizations you visit off of Facebook.	View
Managing Your Information	Learn more about how you can manage your information.	View
Deactivation and Deletion	Temporarily deactivate or permanently delete your account.	View

Request Copy

Available Copies

Date Range:

All of my data ▼

Format:

JSON ▼

Media Quality:

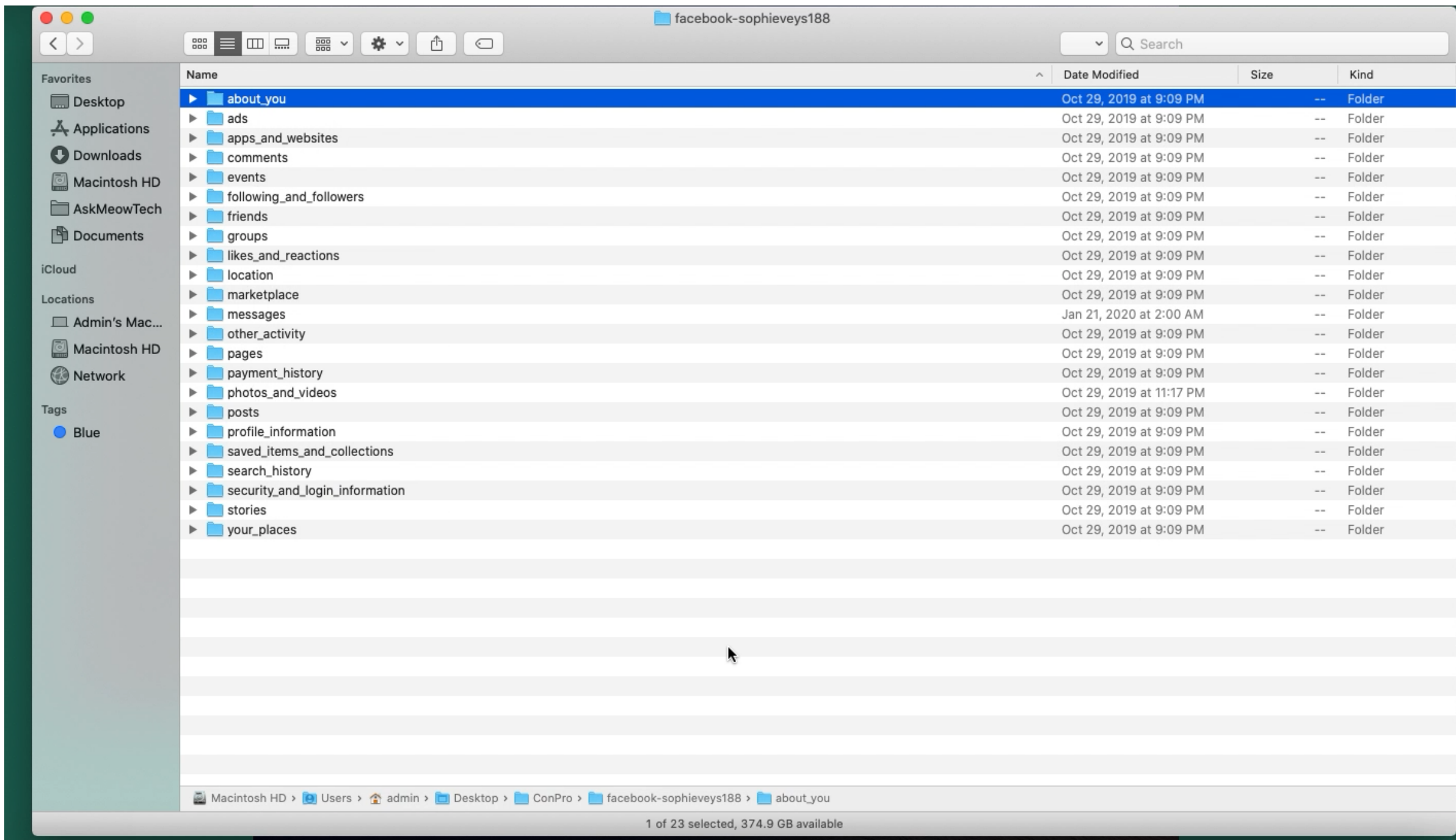
High ▼

Create File



A copy of your information is being created.

Your copy may contain more than one file, depending on how much information your request contains. We'll let you know when your copy is complete, so you can download it to your preferred device. You can [cancel this process](#) before the file is complete.



profile_information.json

```
31  },
32  "previous_names": [
33    {
34      "name": "Sophie Vays",
35      "timestamp": 1515289841
36    }
37  ],
38  "other_names": [
39
40  ],
41  "current_city": {
42    "name": "Chicago, Illinois",
43    "timestamp": 0
44  },
45  "hometown": {
46    "name": "Chattanooga, Tennessee",
47    "timestamp": 0
48  },
49  "education_experiences": [
50    {
51      "name": "The University of Chicago",
52      "start_timestamp": 1538413200,
53      "graduated": false,
54      "concentrations": [
55
56      ],
57      "school_type": "College"
58    }
59  ],
```

```
"advertiserInfo" : {  
  "advertiserName" : "Foundation Medicine",  
  "screenName" : "@FoundationATCG"  
},  
"matchedTargetingCriteria" : [ {  
  "targetingType" : "Keywords",  
  "targetingValue" : "#ASC019"  
}, {  
  "targetingType" : "Locations",  
  "targetingValue" : "United States"  
} ],  
"impressionTime" : "2019-06-01 08:21:18"
```



```
{  
  "sender_name": "Bob",  
  "timestamp_ms": 1569679976090,  
  "content": "Video call DND session?",  
  "type": "Generic"  
},
```



Data Portability ☒

Right to Access ☐

Challenges of data access

- Authenticating data subject access requests (SARs)
- Deciding what data should be included (e.g., derived data?)
- Finding all data that is included
- What should the format of the response be?
 - What definitions are needed?
- What should the user experience be?

Right to Erasure

GDPR Article 17

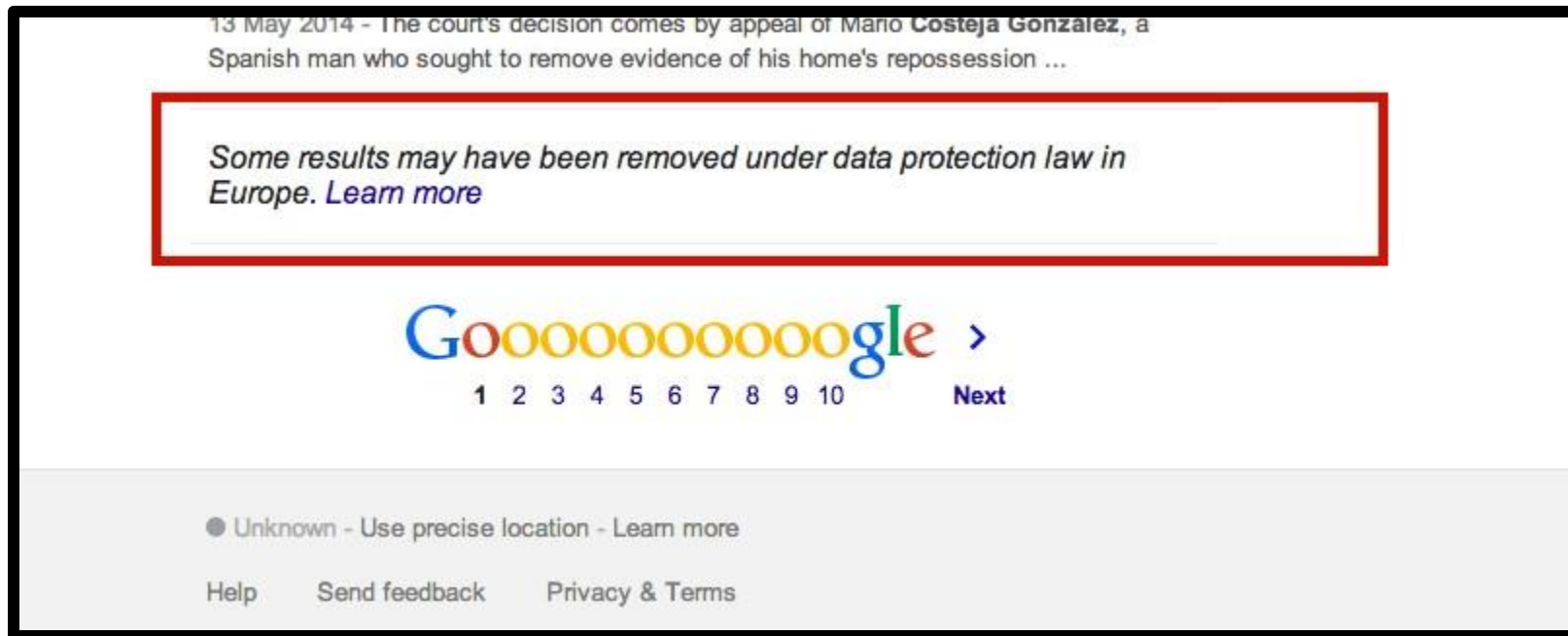
- **Right to erasure ('right to be forgotten'):**
- The data subject shall have the right to obtain from the controller the **erasure of personal data concerning him or her** without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:
 - the **personal data are no longer necessary in relation to the purposes for which they were collected** or otherwise processed;
 - the data subject **withdraws consent** on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;
 - the **data subject objects to the processing** pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);

CCPA (California Civic Code 1798.105)

- (a) A consumer shall have the **right to request that a business delete any personal information about the consumer which the business has collected from the consumer...**
- (c) A business that receives a verifiable consumer request from a consumer to delete the consumer's personal information pursuant to subdivision (a) of this section shall delete the consumer's personal information from its records and direct any service providers to delete the consumer's personal information from their records.
- (d) A business or a service provider shall not be required to comply with a consumer's request to delete the consumer's personal information if it is necessary for the business or service provider to maintain the consumer's personal information in order to:
 - (1) **Complete the transaction** for which the personal information was collected, fulfill the terms of a written warranty or product recall conducted in accordance with federal law, provide a good or service requested by the consumer, or reasonably anticipated within the context of a business' ongoing business relationship with the consumer, or otherwise perform a contract between the business and the consumer.
 - (2) **Detect security incidents, protect against malicious, deceptive, fraudulent, or illegal activity**; or prosecute those responsible for that activity.
 - (3) **Debug to identify and repair errors** that impair existing intended functionality...
 - (6) Engage in public or peer-reviewed scientific, historical, or statistical research in the public interest that adheres to all other applicable ethics and privacy laws, when the business' deletion of the information is likely to render impossible or seriously impair the achievement of such research, if the consumer has provided informed consent.
 - (7) To enable solely internal uses that are **reasonably aligned with the expectations of the consumer** based on the consumer's relationship with the business

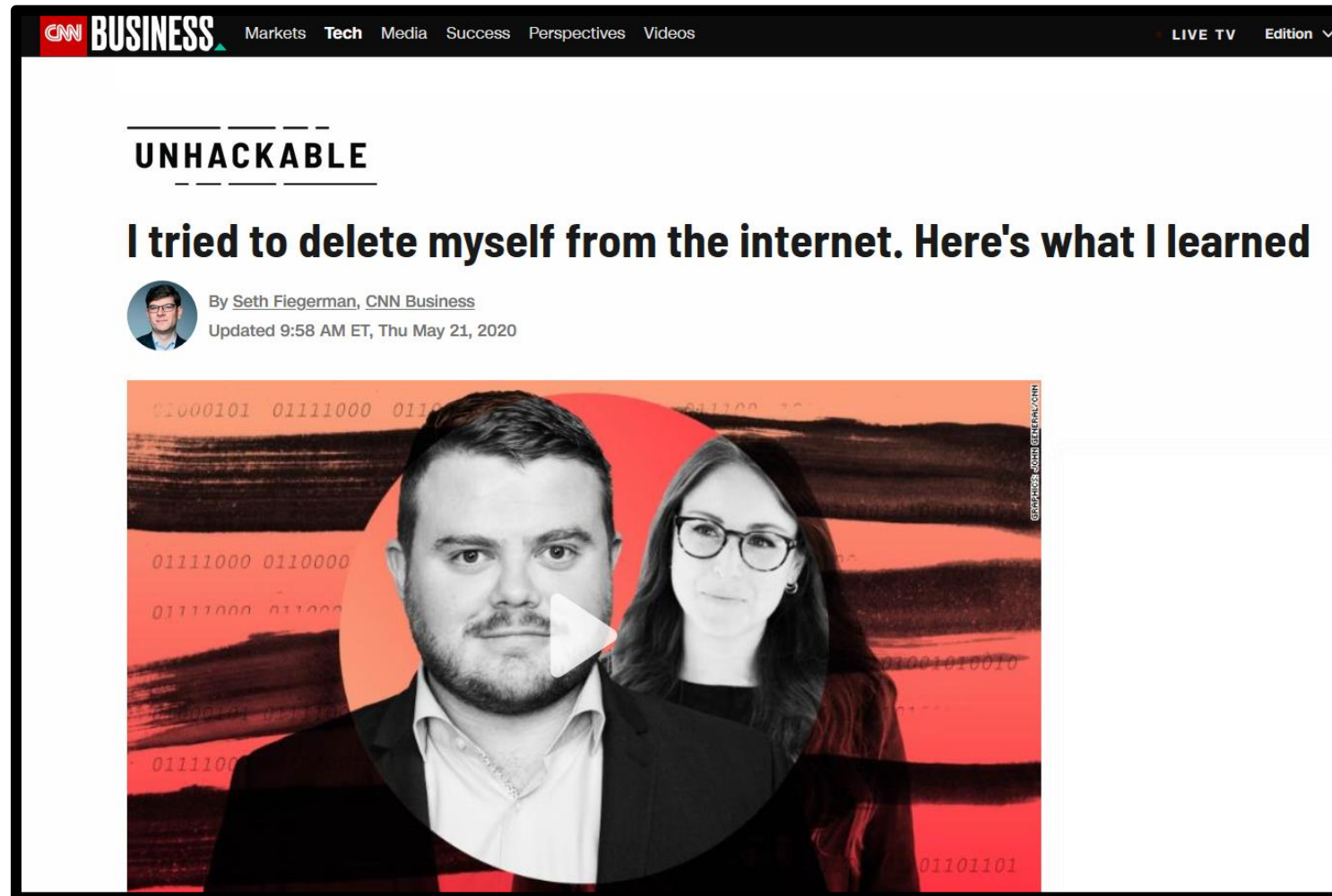
Right to Erasure

- Bound to jurisdictions
- Not applicable to public figures / issues of public interest



The Difficulty of Data Deletion

Making Others Delete Your Data is Hard



<https://www.cnn.com/2020/05/21/tech/deleting-personal-data-online/index.html>

Learning Who Has Your Data is Hard



<https://www.nytimes.com/2019/11/04/business/secret-consumer-score-access.html>

Backup Copies Are Fun



Shared Responsibility is Also Fun



<http://www.gdprtoons.com/2017/11/gdpr-addresses-joint-controllers-with.html>

Challenges for data deletion

- Backup copies and data duplication
- Log files and other subtle records
- Removing the data that should be erased from aggregated data
 - How does this impact an ML model?
- Legal, contractual, or policy considerations for data *retention*
- Provenance tracking
 - How do you store metadata about the subject of particular data? How does this vary based on the data structure / data type?
 - What if the data is sold?
 - What if the entity holding the data is sold?

Design exercise: data deletion

- You have a policy that all personal data must be deleted after 90 days
 - Data that has been pseudonymized can be kept forever
- You decide that you will also respond within 72 hours to erasure requests

Who is the Data Subject?



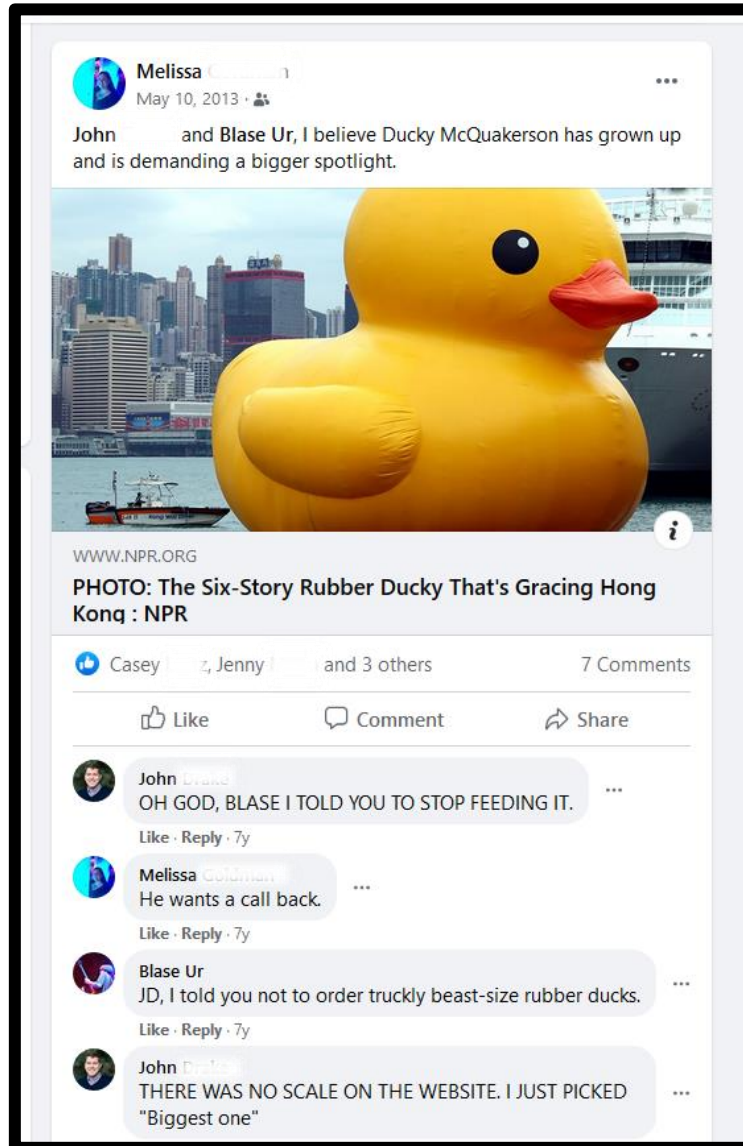
Image from <https://freesvg.org/group-of-people>

Who is the Data Subject?

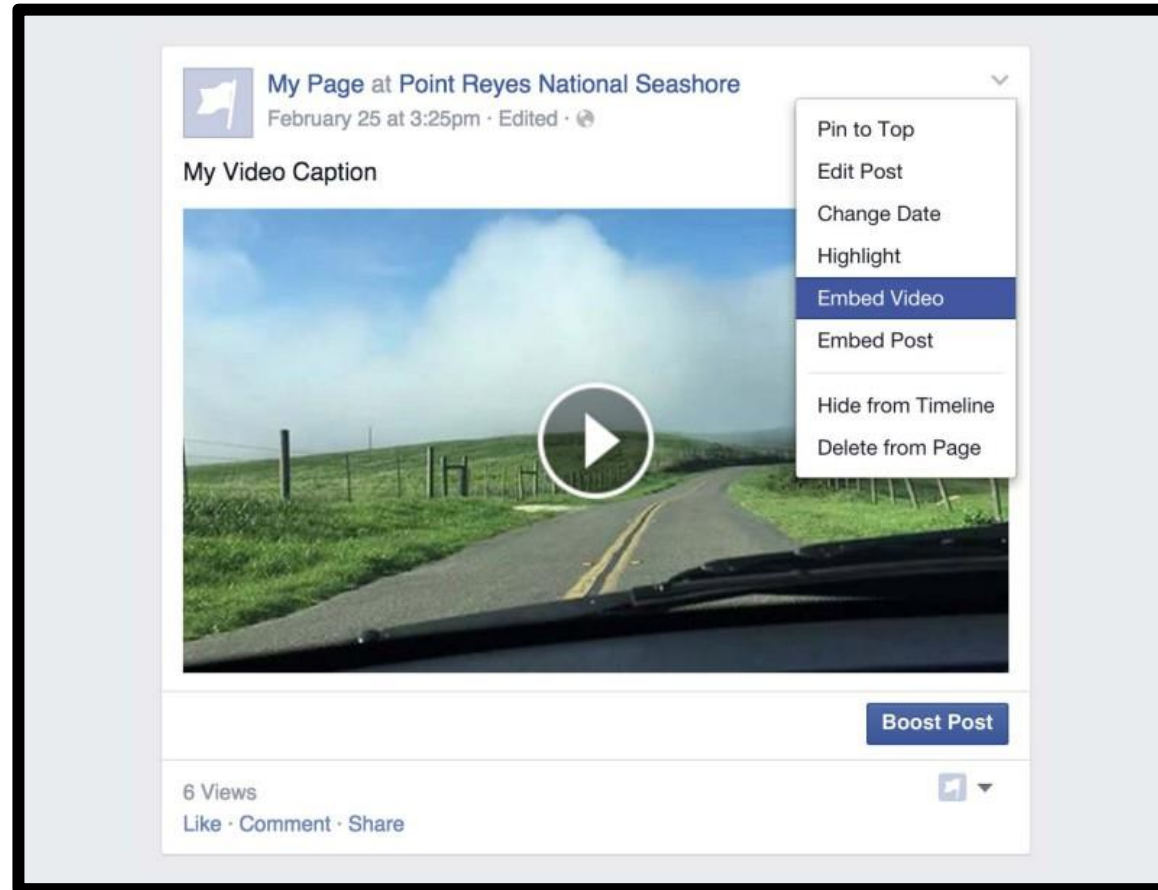
A screenshot of a HuffPost article page. The page has a dark header with the HuffPost logo in the center. To the right of the logo are links for 'Log In' and 'Join HuffPost Plus'. Below the header is a navigation bar with links for 'NEWS', 'CORONAVIRUS', 'POLITICS' (highlighted in green), '2020 ELECTIONS', 'ENTERTAINMENT', 'LIFE', 'PERSONAL', 'VIDEO', and 'SHOPPING'. On the far right of the navigation bar is a link for 'U.S. Edition'. The main content area is white. At the top right of this area is an 'AdChoices' icon. Below this is the article's category 'POLITICS' and its timestamp '05/26/2020 09:46 am ET | Updated 1 hour ago'. The article title is 'Widower Pleads With Twitter To Delete Trump's Conspiratorial Tweets About Late Wife' in large, bold black font. Below the title is a summary: 'Timothy Klausutis slammed Trump's baseless conspiracy theory about MSNBC's Joe Scarborough killing his intern in a powerful letter to Twitter.' Below the summary is the author's name 'By Lee Moran' next to a small circular icon with a green bar. At the bottom of the article preview is a short paragraph: 'Timothy Klausutis, the widowed husband of the woman who [Donald Trump](#) has [baselessly suggested](#) was killed by [MSNBC's Joe Scarborough](#), has written a powerful letter to Twitter CEO Jack Dorsey calling for the deletion of the president's tweets'.

https://www.huffpost.com/entry/donald-trump-joe-scarborough-letter-husband-twitter_n_5ecd036ac5b6c992ff263b7b

The Nuances of Deletion



The Nuances of Deletion



The Nuances of Deletion



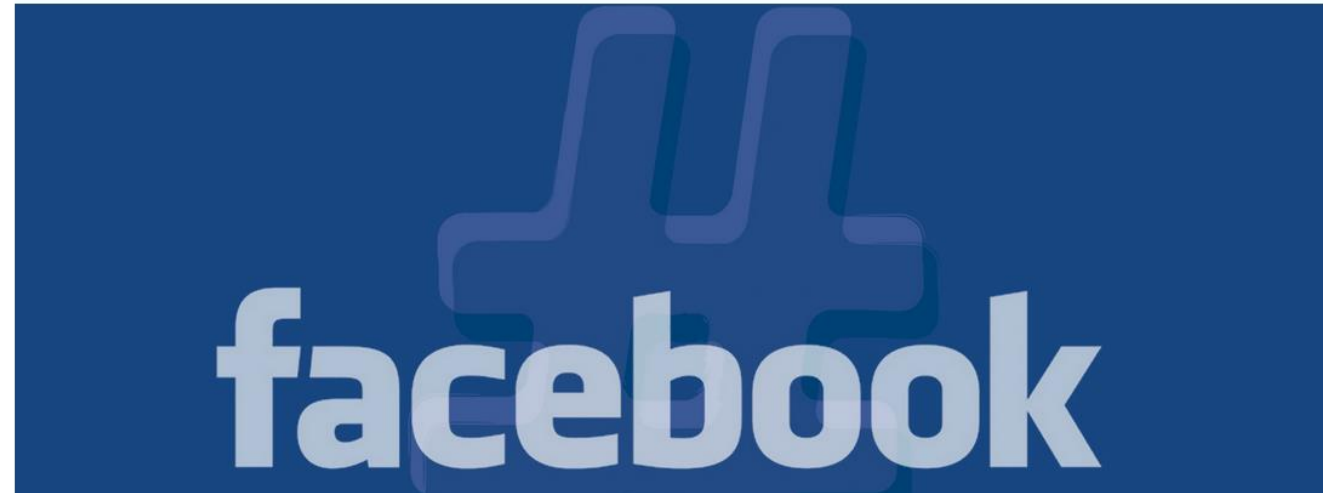
<https://www.facebook.com/notes/bob-baldwin/new-videos-in-comments/10156919502775277/>

The Nuances of Deletion

Facebook Launches New Video Embeds & Comment Syncing From Site To Page

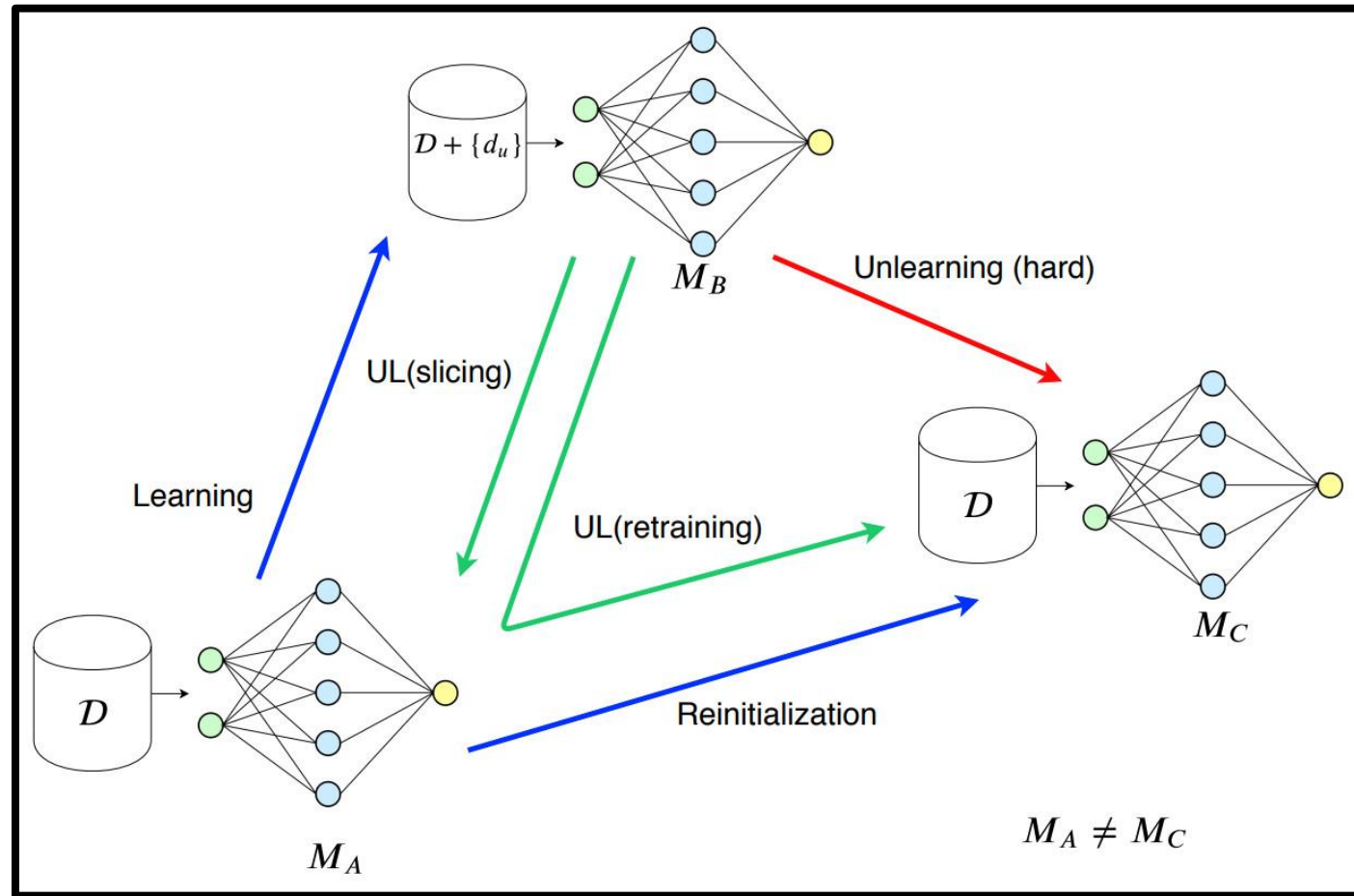
Facebook video now can be embedded independent of Facebook posts; conversation threads in the Facebook comment plugin will have the option to be mirrored on Facebook Pages.

Martin Beck on March 25, 2015 at 3:25 pm



Machine Unlearning

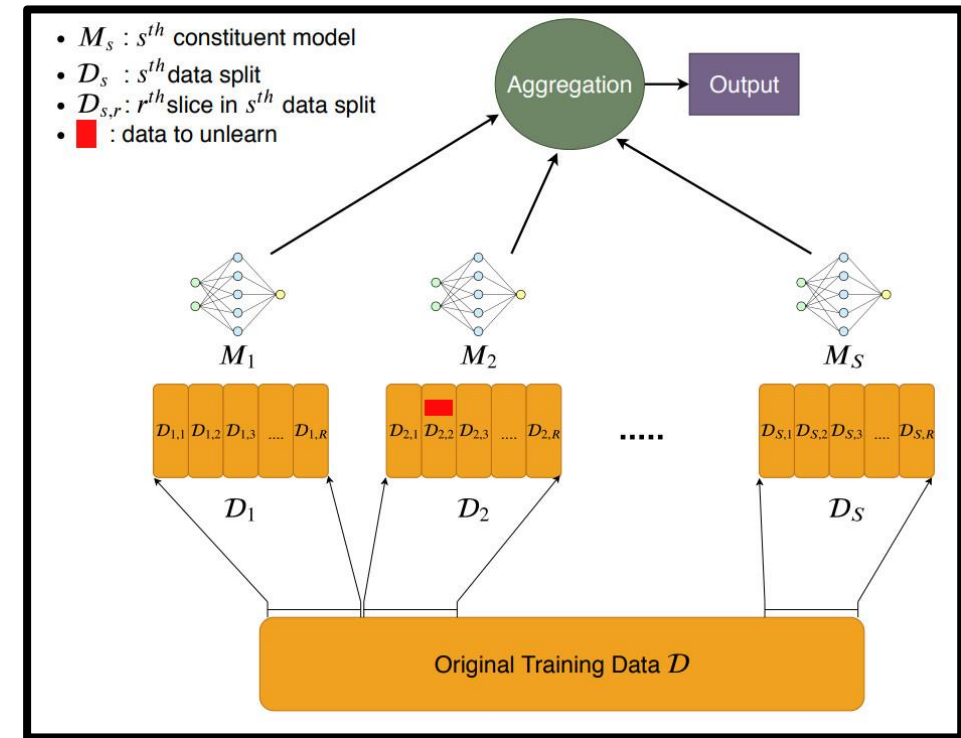
Machine Unlearning



Images taken from <https://arxiv.org/abs/1912.03817>

The work is summarized on <http://www.cleverhans.io/2020/07/20/unlearning.html>

Machine Unlearning



- Shard data: Each subject's data is contained in only one shard
 - You only need to retrain 1 of N shards
- Slice data within a shard: Train for Slice 1, Slices 1-2, Slices 1-3, and maintain state
 - Expectation that you can start retraining halfway through the slices

Images taken from <https://arxiv.org/abs/1912.03817>

The work is summarized on <http://www.cleverhans.io/2020/07/20/unlearning.html>

Genetic Data

The Ownership of Biological Data

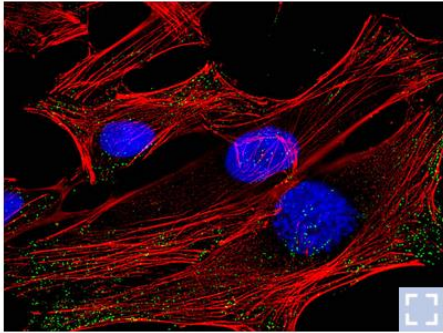
'Henrietta Lacks': A Donor's Immortal Legacy

February 2, 2010 · 12:00 PM ET
Heard on **FRESH AIR**

 **37-Minute Listen**    

In 1951, an African-American woman named Henrietta Lacks was diagnosed with terminal cervical cancer. She was treated at Johns Hopkins University, where a doctor named George Gey snipped cells from her cervix without telling her. Gey discovered that Lacks' cells could not only be kept alive, but would also grow indefinitely.

For the past 60 years Lacks' cells have been cultured and used in experiments ranging from determining the long-term effects of radiation to

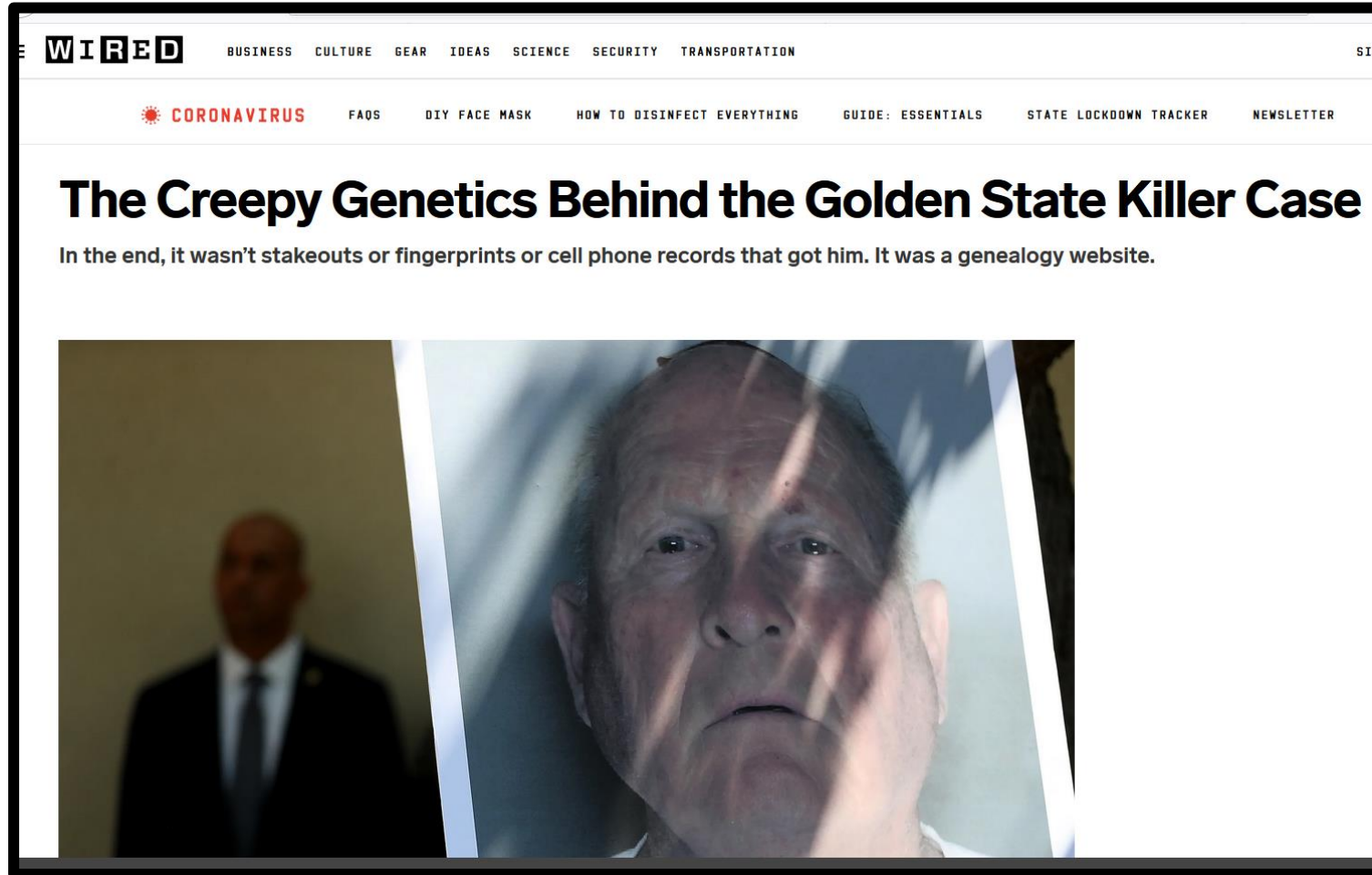


A fluorescence micrograph of HeLa cells, derived from cervical cancer cells taken from Henrietta Lacks and named in her honor

Tomasz Szul/Visuals Unlimited, Inc./Getty Images

<https://www.npr.org/2010/02/02/123232331/henrietta-lacks-a-donors-immortal-legacy>

Wins (?) for DNA Data



<https://www.wired.com/story/detectives-cracked-the-golden-state-killer-case-using-genetics/>

Losses (?) for DNA Data



<https://www.theguardian.com/lifeandstyle/2018/sep/18/your-fathers-not-your-father-when-dna-tests-reveal-more-than-you-bargained-for>