

Exercise 1: Case Study of a Technical Researcher

Goal:	The goal of this exercise is to identify resource requirements, process guidelines and ideal ways to support technical researchers who are doing work that involves large amounts of potentially sensitive data.
--------------	--

Materials:	Scratch paper, sticky notes, pens/markers
-------------------	---

Use Case:	<p>Alice is a current PhD student in Computer Science working at the intersection between computer vision and machine learning. For her dissertation, she needs a large corpus of images for a training set. Let's imagine that her advisor (an expert in image detection) has never worked with human subjects or scraped data before.</p>
------------------	---

Alice intends to get funding from the National Science Foundation to help support her research. In order to apply for a grant, she needs to write a Data Management Plan. (She will also need to file a human subjects research protocol with the IRB.)

She's also in conversations with a company (let's call it ImageCompany) about possibly giving her large corpuses of image data to work with. She also intends to use Amazon's Mechanical Turk to have people classify the images she's encountering.

Datasets she'll be working with:

- 1) Scraped data from the internet
- 2) Results from Amazon Mechanical Turk hits
- 3) ImageCompany's image data

Exercise Questions:	<p>Below are a series of questions that can drive your conversation. Feel free to also identify other issues in the research process that haven't been raised below.</p>
----------------------------	--

- How will Alice know how to write a data management plan for her

work? Where will she see what others have written?

- Who will teach her about how to navigate the IRB? How will she learn how to create an appropriate protocol?
- How will Alice store and secure her data? Are there services on campus that she can turn to that will support her? What will she be expected to manage?
- How will she know which private services to use in her work? (e.g., Dropbox, AWS, Azure, etc.) Who helps her figure out the security and terms of service issues that might arise?
- Who will help her negotiate a contract with ImageCompany?
- Who will help her learn how to navigate Amazon Mechanical Turk's protocols and processes?
- Who can help her deal with best (technical and social) practices in scraping data or navigating data collection on Amazon Mechanical Turk?
- Who will help her understand how to assess the limitations and biases in her data sets?
- How can Alice get help with technical issues she's encountered? (Consider statistics questions, ML questions, etc. Imagine what happens when her advisor is too busy, lacks the expertise, or she's too afraid to ask.)
- Should her processes change depending on the social implications of her project? What if she's consulting for ImageCompany as part of getting access to the data?

What are the questions Alice hasn't even thought of yet?

Part 1:

Help prepare Alice to be able to do her work well by identifying the resources and tools she can use. There are many resources at your organization to help Alice, both formal and informal. The goal here is to identify what exists and what should exist.

On the provided worksheet, identify your organization's resources that might be useful to Alice as she tries to do her research. Make notes of any barriers that might exist with these resources. Do not fill out the other columns in this

stage.

When you have finished listing resources, pass the worksheet to a neighbor and have them check off if they have heard of the resource or used the resource in the past. When everyone is done, share with your small group.

Part 2

Once everyone in the group has shared what they wrote down, transcribe resources your organization currently offers to one color of sticky notes and group together based on type. *Ex: colleagues, friends, experts, databases, institutions, websites, the library and other places, etc.*

Next, think of services that you know of at other organizations or can imagine that might be helpful to Alice. Transcribe these ideas to sticky notes in a new color and place in relationship with the sticky notes from the previous exercise. Take notes for later discussion.

Additional Discussion:

Imagine that something about her process changes mid-way. Imagine that ImageCompany goes bankrupt and closes (but she still has the data); part of what comes up is that ImageCompany gets into trouble for sharing their data in violation of user desires. Where will she turn to for advice on how to proceed? What if she realizes that her use of the data violates the Terms of Service? What if she's already had a breakthrough because of the use of the data?
