

Community Engagement Framework for Machine Learning Datasets

April 2024

Key Elements of Community Engagement in Dataset Development:

- Identification of key communities involved in and/or affected by the development of improved datasets; including, impacted stakeholders, data providers, and potential data users (i.e., those who can be engaged to support solution generation).
- Participation and consent of impacted communities in information gathering and shaping solution generation.
- Engagement of support communities in information gathering and solution generation.
- Preparation for future full ownership of the datasets by impacted communities

IDENTIFICATION OF IMPACTED COMMUNITIES

In this initial step, the research team should work to identify those communities that are impacted by the problem at hand, as well as those communities that may be part of generating and developing solutions to address that problem. Some questions to consider:

- Which communities have been most negatively impacted by this issue?
- What kinds of burdens beyond health impacts are created by this issue?
- Which communities have the ability to avoid negative impacts from this issue? Can these communities utilize resources to support solution generation that can reduce negative health outcomes and other impacts on the most burdened communities?
- Would it be helpful to map these different communities?
- Are there historical issues (economic, political, cultural) that might hinder any efforts at solution generation for this issue? Are there differing opinions/positions about this issue in the community that might affect any efforts to address it?

PARTICIPATION AND CONSENT OF IMPACTED COMMUNITIES

- How does this effort propose to engage the impacted communities identified above in gathering information that might improve datasets related to this issue?

- Is it possible to engage both the leaders of institutions that will provide data (e.g., hospital administrators) and the staff who are collecting the data (e.g., the individuals who are taking X-rays or other images)?
- Are there steps you can take to build trust across stakeholders from different sectors or disciplines (e.g., healthcare, government, patients, etc.)? Can inviting input on and establishing agreements about how data will be used/not used and how privacy will be protected help to build trust?
- Will you meet with stakeholders one-on-one or as a group (or both)?
- How will you reach impacted communities when they are geographically remote or distributed?
- How many meetings will you conduct and at what stages of the project?
- What is the agenda for the meeting(s)?
- Where will the meetings take place and will transportation be provided? Will **child** care be provided?
- What time will the meetings be conducted and will affected stakeholders be available at this time(s)?
- How else will information be provided to affected communities? Via a website or written materials?
- How will impacted communities provide input on:
 - What data will be collected?
 - How data will be managed and privacy protected?
 - How data will be used?
- Do the impacted communities understand the information gathering process this effort has proposed? Have they expressed consent (either individually or as a group) with the information gathering process?
- If consent is not possible because data is from past patients, what steps should be taken to protect privacy? (See **Data** Anonymization Guide)
- What benefits can be offered to individuals/community who are providing data? Information to help treat or manage their health issues? Benefits to future patients, for example through a screening or diagnostic tool? Ask researchers who use data to commit to something that will provide community benefit, such as volunteering to do rounds in the impacted community.

ENGAGEMENT OF SUPPORT COMMUNITIES

- How might the following communities support information gathering and solution generation:
 - Data science community
 - Clinical research community
 - Medical community (to include clinical practitioners utilizing the improved datasets)

- Government policy community
- Are there any historical/cultural factors that should be considered when deciding whether/how to involve some of the groups above?
- Are there other communities that could support information gathering and solution generation?
- How does the effort propose that the support communities will work with the impacted communities in the information gathering and solution generation phases?
- Does this interaction between the impacted and support communities require capacity building so that the groups can understand each other? What resources will be necessary to build that capacity?

DATA STANDARDS

- Are common data standards needed to facilitate data sharing and use?

Plan for Engagement of Impacted Communities

DRAFT TEMPLATE

Meeting Date, Time, Location				
Invitees				
Topics to Cover (e.g. data needs, proposal development, data collection, data access and use, progress)				
Number of Participants				