S9 Data Structure Overview

The data structure consists of 8 collections, 6 of which were implemented. We define them below along with their key attributes. Collections and attributes denoted with an * were not implemented because some data was not collected and other features were more meaningful for replicability purposes and not for the initial deployment.

Project Collection: Stores contextual information about the community engagement process

- `user_id*` (string): user that created the project
- `project_name` (string): name of project
- `date_created*` (int): timestamp of when project was created
- `description` (string): information about the project at a high level
- `phases` (json): list of JSON objects describing the different phases. Each item has the following fields
    - `phase_index` (int)
    - `status` (enum)
    - `title` (string): e.g. create, refine
    - `description` (string): goal
    - `start_date` (int)
    - `end_date` (int)
    - `who` (string): e.g. planners, community
    - `output` (string): link to other collections
- `data*` (json): existing conditions data like population and KPIs

Events Collection: Captures diverse feedback-gathering mechanisms ranging from surveys to community workshops and tabling events, including descriptive information for each engagement activity.

- `event_name` (string): name of event
- `type` (string): workshop activity, survey, emails, interactive map etc.
- `description` (string): description of event
- `summary` (string): event summary (e.g. here is what we heard during the zoom meeting)
- `date` (string): date or date range of event
- `phase_index` (int): which phase the event was conducted in
- `neighborhood_id*` (string): which neighborhood the event took place in
- `location*` (tuple): lat / lon of where event was hosted
- `involved_stakeholder_ids*` (list): list of stakeholder ids
- `is_public_event` (boolean): whether feedback from this event can be shared publicly
- `instructions*` (string): any instructions for event activities
- `prompts` (list): actual questions asked to the community
- `img_links*` (list): list of links to images of the event
- `resource_links*` (list): list of links to any resources used in the event (e.g. presentations)

Voices Collection: Records community input with associated metadata connecting voices to events, topics, and outputs, forming the central repository of community feedback that drives subsequent analysis.

- `feedback` (string): the feedback text if the prompt is open-ended or number if the prompt is closed-ended
- `is_numeric*` (boolean): whether the feedback is quantitative or qualitative
- `is_verbatim` (boolean): whether the feedback was transcribed verbatim
- `note*` (string): any notes to provide context around the piece of feedback
- `prompt` (string): the question that prompted the feedback if applicable
- `phase_index` (int): which phase the feedback was conducted in
- `collection_date` (int): the date that the feedback was collected
- `upload_date*` (int): the date that the feedback was uploaded to the platform
- `participant_id*` (string): the person who contributed that feedback
- `stakeholder_id*` (string): applicable if the participant explicitly mentions speaking from a particular stakeholder hat for that comment
- `event_id` (string): the event id which describes the source of the feedback
- `themes` (list): list of themes that the feedback is tagged with
- `structural_codes*` (list): list of structural codes that the feedback is tagged with
- `feedback_status` (enum): public, private, flagged
- `location` (list): any lat / lon information embedded within the feedback
- `neighborhood_ids` (list): which neighborhoods the feedback mentions
- `output_ids` (list): list of output ids that the feedback is connected with

Sub-Geographies Collection: Contains information about distinct areas within the broader study region, particularly relevant for larger planning initiatives that span multiple neighborhoods.

- `sub_geography_name` (string): name of sub-geography / neighborhood
- `description*` (string): description of sub-geography / neighborhood
- `geo_data` (string): link to file with geo json data with neighborhood boundaries

Topics Collection: Maintains information about overarching themes used to organize and make sense of community input. These topics may be predefined or emerge organically through the sensemaking process.

- `main_topics` (list): list of JSON objects describing the different topics. Each item has the following fields
    - `name` (string)
    - `definition` (string)
    - `date_created*` (int)
    - `last_edited*` (int)
- `other_tags` (list): list of JSON objects describing the different rationales for why a feedback is uncited. Each item has the following fields
    - `name` (string)

    - `rationale` (string)
    - `date_created*` (int)
    - `last_edited*` (int)
- `sub_topics*` (list): list of JSON objects describing the different sub-topics. Each item has the following fields
    - `name` (string)
    - `definition` (string)
    - `main_topic` (string)
    - `AI_suggested` (boolean)
    - `date_created` (int)
    - `last_edited` (int)
- `structural_codes*` (list): list of structural codes that are applied to the project (e.g. Hope, Concern, Idea, Experience, Opinion, Position Claim, Question)
- `partner_agencies*` (list): list of partner agencies to consider (e.g. department of transportation)

Outputs Collection: Stores the synthesized products of the engagement process, categorized as insights (key findings), goals (high-level visions), and recommendations (specific strategies).

- `type` (enum): whether the output is a insight, goal, or recommendation
- `phase_index` (int): project phase when the output was created
- `date_created*` (int)
- `date_published*` (int)
- `is_public` (boolean)
- `nickname` (string): short title
- `title` (string): full title
- `description` (string): description
- `summary` (string): summary to justify output
- `notes` (string): internal notes about the output
- `topics` (list): list of topics that an output is tagged with
- `partner_agencies*` (list): list of relevant partner agencies
- `cited_voices` (list): list of feedback ids that relate to the output
- `relevant_participants*` (list): ids of participants whose feedback is cited
- `relevant_stakeholders*` (list): ids of stakeholders that might be impacted by the output
- `neighborhood_ids*` (list): link to neighborhood(s) that the output describes or would impact
- `sparked_by` (list): list of ids that informed the output
- `next_steps` (list): list of ids that followed the output
- `quantitative_data*` (list): text and/or images
- `matrix*` (json): some way of evaluating recommendations
- `connection_to_external_projects*` (list): list of JSON objects describing relevant ongoing efforts. Each item has the following fields
    - `name` (string)

    - description (string)
    - home_organization (string)
    - link (string)

Stakeholders Collection*: Captures the different groups that could impact or be impacted by the project

- name (string)
- description (string): information about the stakeholder
- neighborhood_id (string): sub-geography associated with stakeholder if applicable
- summary (string): bullet point / summary of what stakeholder currently thinks about the project
- project_ids (list): list of projects that stakeholder was involved in
- relationship (json):
    - holder (string): user who holds the relationship
    - date_initiated (int)
    - contact_history (list): history of contacts
    - last_contact (int): date last contacted
    - notes (string): notes on relationship

Participants Collection*: Captures information about the individuals who gave feedback

- alias (string): name / pseudonym of participant
- stakeholder_ids (list): list of stakeholder ids that the participant is part of
- feedback_ids (list): list of feedback ids that the participant contributed
- project_ids (list): list of projects that stakeholder was involved in
- demo_public (boolean): whether to share demographic information publicly
- demo (json): meta-data associated with participant. Could include fields like
    - gender (string)
    - race (string)
    - age (string)