

# Nouveau Navis Health Systems

## Vignette

The following is a vignette designed to prompt reflection and discussion about issues at the intersection of datasets, AI, documentation, and transparency as you go through activities in the Data Cards Playbook.

Our vignettes are fictional scenarios that are carefully crafted to capture the complexity in dataset documentation and transparency, as well as to help Playbook participants analyze and translate the dynamics of the relationship between various stakeholders. They form running examples that connect the frameworks and concepts introduced throughout the Playbook so participants can apply the Playbook to their own datasets.

This vignette is entirely fictitious and is not representative of any known product development efforts at Google or otherwise. Any resemblance to places, products, organizations, or individuals is purely coincidental.

## Background

In a contract with The Government of Ceres's Division of Reproductive Health (DRH), Nouveau Navis Health Systems has been tasked to create a Maternal, Child and Fetal Health program for rural counties across the country. One of the deliverables include a new AI-powered resource planning platform, project codename Caretelligent, for low-cost clinics in these counties.

## Collecting the Data / Identify Producers

Nouveau Navis has partnered with Forestide Data Services to create a dataset which will power Caretelligent. Under this partnership:

- **Nouveau Navis** will provide up to 5 domain experts to help the Caretelligent team and Forestide Data Services follow the best practices when creating this dataset.
- **Forestide Data Services** will:
  - Contract healthcare workers to collect pregnancy data in Rural Ceres.
  - Work with the Caretelligent team to turn this raw data into a usable format.
  - Hand off data to the Caretelligent team after all quality control checks have been passed.
- After the hand-off, the Caretelligent team will be responsible for this dataset.

Over the next 4 weeks, 2500 healthcare workers will visit and survey on an average ~25 expecting mothers a day, collecting ~1 million data points that describe:

- Demographic information
- Socio-economic information
- Nutrition & Dietary practices
- Past pregnancies & health history
- Current healthcare practices

Taxonomy of Stakeholders	In this scenario: Source Producer → Expecting mothers Core Producer → Forestide Data Services Adjacent Producers → Healthcare workers, Domain Experts Impacted Producer → The Caretelligent Team
--------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## Assessing Product Use / Identify Agents

Nouveau Navis health systems prides itself on being multi-disciplinary and transparent. The team that will build Caretelligent is split into three workstreams:

- The **product workstream** is responsible for how Caretelligent looks, behaves, and is used. They will also act on feedback from low-cost clinics in rural Ceres who have signed up for the pilot.
- The **infrastructure workstream** is responsible for tools and platforms needed to deploy Caretelligent.
- The **ML/AI workstream** will build the models needed in Caretelligent – they will use data from Forestide, along with existing data from Nouveau Navis to train these models.

Like all AI projects at Nouveau Navis, Caretelligent will be subjected to frequent and thorough internal reviews by a Research Ethics Board, Nouveaus Navis' Human Health Risk and Assessment (HHRA) Board, and Nouveau Navis' AI Solutions Commision (AISC).

Before it can be used in Caretelligent, stakeholders at Nouveau Navis must review and document different aspects of the dataset from Forestide:

- The **product workstream**, who are non-experts, must assess product suitability and determine how the data will be used in Caretelligent.
- The **infrastructure workstream**, who are experts in data but not necessarily in the content of this data, must identify how to integrate Forestide's data into their existing pipeline.
- The **ML/AI workstream**, who are experts in both the data and its domain, must figure out how to use the data in their models, and how to optimize models for Caretelligent.
- The **review boards** must assess if the data has been ethically collected, will be used in an ethical way, and identify any risks to individuals and groups.
- The **pilot clinics**, who are representative of users of Caretelligent, will assess if the data is representative of the patients they see, and has been used in Caretelligent in a way that is credible and reliable.

Taxonomy of Stakeholders	In this scenario: Core Agents → Caretelligent team (all workstreams) Adjacent Agents → Review Boards Impacted Agents → Pilot Clinics
--------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------

## Making Trade-Offs / Identify Users

As part of the pilot trials, Caretelligent has launched successfully in five clinics that represent different zones in rural Ceres. Four of these clinics have separate obstetric and neonatal wards, whereas the fifth has a joint maternity ward. In these clinics, Caretelligent is used by administrative staff, doctors, nurses, physician's assistants, and midwives.

Over the course of a quarter of testing, the product workstream found that Caretelligent works better where the wards are separate. They also find that Caretelligent's forecasts are significantly more accurate for returning mothers than first-time parents. The Caretelligent team investigates this and finds that Forestide collected data primarily from expecting women who had already conceived, based on existing records from clinics. As such, first-time mothers, and first-time parents from non-traditional situations, such as IVF/IUI and surrogacy were underrepresented.

After much deliberation, Caretelligent decided that at this point, they are unable to establish a data collection protocol for first-time parents from non-traditional circumstances, but are able to collect data from first-time mothers. As a response, Caretelligent requests clinics to start contributing data from first-time mothers to a dataset to add to the data collected by Forestide.

<b>Taxonomy of Stakeholders</b>	<p>In this scenario:</p> <p>Impacted Users → Clinics that have been unable to plan resources due to erroneous predictions; first-time parents</p> <p>Typical Users → Maternity Ward in Clinics / Staff at clinics</p> <p>Contributing Users → First-time mothers</p>
---------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



The [Data Cards Playbook](#) <sup>↗</sup> by Google Research is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. You are free to share and adapt this work under the [appropriate license terms](#) <sup>↗</sup>.