VF Agent - Comprehensive Question Reference

This document consolidates all questions the VF Agent is designed to answer, organized by category.

MoSCoW Definitions

MoSCoW is an industry standard for defining feature requirements. Here is what they mean for this project for each question.
- **Must Have:** the agent must reliably handle this case.
- **Should Have:** the agent should be able to handle this case, but we need to build prototypes to confirm this is possible.
- **Could Have:** the agent might be able to handle this case; major prototyping is required to determine feasibility.
- **Won't Have:** the agent will not handle this case because development effort outweighs the value of the question.

Agent Component Glossary

- **Supervisor Agent**: Simple router that recognizes intent and delegates user queries to the appropriate sub-agents.
- **Genie Chat**: Databricks text to SQL agent that supports converting plaintext english questions into SQL queries.
- **Geospatial Calculation**: non-standard geospatial calculations e.g. geodesic distance.
- **Medical Reasoning Agent**: a reasoning agent with medical expertise. It will add context, modify user queries, or perform reasoning on results returned by other agents.
- **External Data**: data that is not currently in the Foundational Data Refresh (FDR). This data will be added to the Virtue Foundation workspace or queried in real-time.
- **Vector Search with Filtering**: semantic lookup on plaintext fields in addition to metadata-based filtering.
- **Unknown**: the ideal architecture for this question has unknowns and needs a research spike.

---

1. Basic Queries & Lookups

These are fundamental queries for finding and counting facilities.

#	Question	MoSCoW	Architecture
1.1	How many hospitals have cardiology?	Must Have ▾	Genie Chat

#	Question	MoSCoW	Architecture
1.2	How many hospitals in [country/region] have the ability to perform [specific procedure]?	Must Have ▾	Genie Chat
1.3	What services does [Facility Name] offer?	Must Have ▾	Vector Search with Filtering
1.4	Are there any clinics in [Area] that do [Service]?	Must Have ▾	Vector Search with Filtering
1.5	Which region has the most [Type] hospitals?	Must Have ▾	Genie Chat

2. Geospatial Queries

Queries involving location, distance, and geographic analysis.

#	Question	MoSCoW	Architecture
2.1	How many hospitals treating [condition] are within [X] km of [landmark/location]?	Must Have ▾	Genie Chat + Geospatial Calculation
2.2	What areas have known disease prevalence for [condition] but no facilities treating it within [X distance]?	Could Have ▾	Genie Chat + Geospatial Calculation + Medical Reasoning Agent + External Data
2.3	Where are the largest geographic "cold spots" where a critical procedure is absent within X km / X hours travel time?	Must Have ▾	Genie Chat + Geospatial Calculation
2.4	What is the service provision gap between urban areas (>[X] population) and rural areas (<[Y] population) for [subspecialty]?	Could Have ▾	Genie Chat + Geospatial Calculation + Medical Reasoning Agent + External Data

3. Validation & Verification

Questions to verify facility claims and detect inconsistencies.

#	Question	MoSCoW	Architecture
3.1	Which facilities claim to offer [specific subspecialty service] but lack the basic equipment required to perform it?	Should H… ▾	Genie Chat + Completeness Assumption
3.2	Which facilities have equipment listed that appears to be temporary or brought in periodically rather than permanently installed?	Could Have ▾	VS Index Point Lookup
3.3	What percentage of facilities claiming [specific capability] show evidence of permanent vs. traveling services and equipment?	Could Have ▾	Genie Agent + **Unknown**
3.4	For each procedure, what % of facilities that claim its presence and also list the minimum required equipment (e.g., cataract surgery ↔ operating microscope)?	Should H… ▾	Medical Reasoning Agent + VS Index Point Lookup + Genie Agent + **Unknown**
3.5	Which procedures/equipment claims are most often corroborated by multiple independent websites?	Should H… ▾	Genie Agent + **Unknown**

---

4. Misrepresentation & Anomaly Detection

Questions to identify facilities with suspicious or inconsistent claims.

#	Question	MoSCoW	Architecture
4.1	What correlation exists between website quality indicators and actual facility capabilities?	Should Have ▾	Medical Reasoning Agent + Genie Chat + **Unknown**
4.2	Which facilities have high bed-to-operating-room ratios or other	Should Have ▾	Genie Chat

#	Question	MoSCoW	Architecture
	metrics indicative of misrepresentation on the internet and social media?		
4.3	Which facilities show other abnormal patterns where expected correlated features don't match?	Should Have ▾	Medical Reasoning Agent + Genie Chat
4.4	Which facilities claim an unrealistic number of procedures relative to their size?	Must Have ▾	Medical Reasoning Agent + Genie Chat
4.5	What physical facility features correlate with genuine advanced capabilities? What about with legitimate subspecialty depth?	Should Have ▾	Medical Reasoning Agent + Genie Chat + **Unknown**
4.6	What facilities show mismatches between claimed subspecialties and supporting infrastructure (e.g., claiming advanced subspecialties without appropriate facility size)?	Should Have ▾	Medical Reasoning Agent + Genie Chat + **Unknown**
4.7	What correlations exist between facility characteristics that move together (e.g., number of operating rooms, depth of subspecialties, equipment sophistication)?	Must Have ▾	Genie Chat
4.8	Which facilities have an unusually high breadth of claimed procedures relative to their stated/observed infrastructure signals (e.g., "200 procedures" + minimal equipment list)?	Must Have ▾	Medical Reasoning Agent + Genie Chat
4.9	Where do we see "things that shouldn't move together" (e.g., very large bed count but minimal surgical equipment; highly specialized claims with no supporting signals)?	Must Have ▾	Medical Reasoning Agent + Genie Chat
4.10	Do spelling/website quality markers correlate with reliability of claims	Won't Have ▾	Medical Reasoning Agent + Genie Chat + External Data

#	Question	MoSCoW	Architecture
	(measured by internal consistency + cross-source confirmation)?		

---

5. Service Classification & Inference

Questions about classifying and understanding the nature of services.

#	Question	MoSCoW	Architecture
5.1	Which procedures appear to be delivered primarily via itinerant outreach (language like "visiting surgeon," "camp," "twice a year") vs permanently staffed service lines?	Could Have ▾	Medical Reasoning Agent + VS Index Point Lookup
5.2	Which facilities' language suggests they refer patients for a procedure ("we can arrange," "we collaborate," "we send to…") rather than actually perform it?	Could Have ▾	Medical Reasoning Agent + VS Index Point Lookup
5.3	How often do strong clinical claims appear alongside weak operational capability signals (limited hours, unclear contact, no appointment workflow, etc.)?	Could Have ▾	Medical Reasoning Agent + Genie Chat
5.4	What combinations of procedures/equipment co-occur as stable "service bundles" (e.g., glaucoma bundle, cornea bundle), and how do these bundles vary by region/income tier?	Could Have ▾	Medical Reasoning Agent + Genie Chat
5.5	Can we derive a "service maturity" index from procedure depth + equipment sophistication + operational capability signals (hours, OR details, staffing mentions)?	Won't Have ▾	NA

6. Workforce Distribution

Questions about healthcare workforce availability and distribution.

#	Question	MoSCoW	Architecture
6.1	Where is the workforce for [subspecialty] actually practicing in [country/region]?	Must Have ▾	Medical Reasoning Agent + Genie Chat
6.2	What is the ratio of certified [specialty] practitioners to population in [region], and how does that compare to expected need?	Won't Have ▾	Medical Reasoning Agent + Genie Chat + External Data
6.3	Which regions have specialists but unclear information about where they actually practice?	Could Have ▾	Medical Reasoning Agent + Genie Chat + External Data + **Unknown**
6.4	How many facilities in [region] have evidence of visiting specialists vs. permanent staff?	Should Have ▾	Medical Reasoning Agent + Genie Chat
6.5	What areas show evidence of surgical camps or temporary medical missions rather than standing services?	Should Have ▾	Medical Reasoning Agent + VS Index Point Lookup
6.6	Where do signals indicate services are tied to individuals rather than institutions (named surgeons, "visiting consultant"), implying fragility in continuity?	Should Have ▾	Medical Reasoning Agent + VS Index Point Lookup

7. Resource Distribution & Gaps

Questions about equipment, infrastructure, and resource availability.

#	Question	MoSCoW	Architecture
7.1	What is the problem type by region: lack of equipment, lack of training, or lack of practitioners?	Could Have ⌄	Medical Reasoning Agent + Genie Chat
7.2	What areas have high practitioner numbers but insufficient equipment to practice?	Could Have ⌄	Medical Reasoning Agent + Genie Chat
7.3	Are there facilities in resource-poor but population-rich areas that lack web presence but show other evidence of high patient volume?	Could Have ⌄	Medical Reasoning Agent + Genie Chat + External Data
7.4	Which regions show the highest prevalence of older/legacy equipment vs newer modalities (based on named devices/models, if captured)?	Could Have ⌄	Medical Reasoning Agent + VS Index Point Lookup + External data
7.5	In each region, which procedures depend on very few facilities (e.g., only 1–2 facilities claim corneal transplant capacity)?	Must Have ⌄	Medical Reasoning Agent + Genie Chat
7.6	Where is there oversupply concentration (many facilities claim the same low-complexity procedure) vs scarcity of high-complexity procedures?	Must Have ⌄	Medical Reasoning Agent + Genie Chat

---

8. NGO & International Organization Analysis

Questions about NGO presence and its impact on healthcare delivery.

#	Question	MoSCoW	Architecture
8.1	Which regions have multiple NGOs or international organizations providing overlapping services? How many of these are permanent vs. temporary?	Should Have ⌄	Medical Reasoning Agent + Genie Chat

#	Question	MoSCoW	Architecture
8.2	What facilities or regions show evidence of coordination vs. competition between different medical mission organizations?	Could Have ⌄	Medical Reasoning Agent + Genie Chat + **Unknown**
8.3	Where are there gaps in the international development map where no organizations are currently working despite evident need?	Must Have ⌄	Medical Reasoning Agent + Genie Chat + Geospatial Calculation + External data
8.4	Where do we see evidence that periodic NGO activity substitutes for permanent capacity (e.g., outreach events)?	Could Have ⌄	Medical Reasoning Agent + VS Index Point Lookup + External data

9. Unmet Needs & Demand Analysis

Questions about identifying underserved populations and unmet healthcare needs.

#	Question	MoSCoW	Architecture
9.1	Which regions have a population size and demographic profile that would demand [X number] of [specific surgeries] per year, but lack evidence of adequate services?	Could Have ⌄	Medical Reasoning Agent + Genie Chat + External Data
9.2	Which population centers between [X and Y quantity of people] at GDP per capita of [range] show highest likelihood of unmet surgical need?	Could Have ⌄	Medical Reasoning Agent + Genie Chat + External Data
9.3	What is an age bracket that identifies an underserved area by specialty?	Could Have ⌄	Medical Reasoning Agent + Genie Chat + External Data
9.4	Which regions have age demographics indicating high need for [age-related	Could Have ⌄	Medical Reasoning Agent + Genie Chat + External Data

#	Question	MoSCoW	Architecture
	procedure] but insufficient service evidence?		
9.5	Which facility regions serve population areas too large for their stated capabilities?	Could Have ‧	Medical Reasoning Agent + Genie Chat + External Data
9.6	Using population + demographics (if joined later), where do predicted high-demand procedures (e.g., cataracts) have low extracted supply signals?	Could Have ‧	Medical Reasoning Agent + Genie Chat + External Data

10. Benchmarking & Comparative Analysis

Questions for comparing regions and facilities against standards.

#	Question	MoSCoW	Architecture
10.1	How does the ratio of [specialists] per [X] population in [region] compare to WHO guidelines or developed country averages?	Could Have ‧	Medical Reasoning Agent + Genie Chat + External Data
10.2	What facilities fall into the "sweet spot" cluster (high population, some infrastructure, but currently ignored by development efforts)?	Should Have ‧	Medical Reasoning Agent + Genie Chat + External Data
10.3	Which regions between population sizes of [X-Y] at [GDP range] show [Z]%+ probability of being high-impact intervention sites?	Should Have ‧	Medical Reasoning Agent + Genie Chat
10.4	For each specialty, what population/GDP/urbanicity bands correlate with highest likelihood of unmet need (high demand + low verified capability)?	Could Have ‧	Medical Reasoning Agent + Genie Chat + External Data

11. Data Quality & Freshness

Questions about the reliability and currency of data.

#	Question	MoSCoW	Architecture
11.1	How quickly do procedure/equipment mentions go stale by region (e.g., last-updated language, outdated announcements, old campaign posts), and where should re-scraping be prioritized?	Won't Have ˅	Medical Reasoning Agent + Genie Chat + External Data
11.2	Which procedures/equipment claims are most often corroborated by multiple independent sources (website text + images + social/other pages) vs single-source only?	Won't Have ˅	Medical Reasoning Agent + Genie Chat + External Data

Summary by Query Type

Category	Count
Basic Queries & Lookups	6
Geospatial Queries	4
Validation & Verification	5
Misrepresentation & Anomaly Detection	11
Service Classification & Inference	5
Workforce Distribution	6
Resource Distribution & Gaps	6
NGO & International Organization Analysis	4

Category	Count
Unmet Needs & Demand Analysis	6
Benchmarking & Comparative Analysis	4
Data Quality & Freshness	2
Total	**59**