

# Schema Documentation: All Columns and Definitions

## Organization Extraction

- **ngos** (list[str]): NGO names present in the text. An NGO is any non-profit organization that delivers tangible, on-the-ground healthcare services in low- or lower-middle-income settings.
- **facilities** (list[str]): Healthcare facility names present in the text. A publishable healthcare facility is any physical site that is currently operating and delivers in-person medical diagnosis or treatment to patients.
- **other\_organizations** (list[str]): Named entities that don't meet facility or NGO classifications.

## Base Organization Fields (Shared)

### Contact Information

- **name** (str): Official name of the organization. Use complete, unabbreviated form with proper capitalization, excluding business suffixes like "Ltd", "LLC", "Inc".
- **phone\_numbers** (list[str]): The organization's phone numbers in E164 format (e.g. '+233392022664').
- **officialPhone** (str): Official phone number associated with the organization in E164 format. Must be the organization's primary official contact number.
- **email** (str): The organization's primary email address.

### Web Presence

- **websites** (list[str]): Websites associated with the organization.
- **officialWebsite** (str): Official website associated with the organization. Only include the domain name, not the full URL, and must correspond to the organization's official website.
- **facebookLink** (str): URL to the organization's Facebook page.
- **twitterLink** (str): URL to the organization's Twitter profile.
- **linkedinLink** (str): URL to the organization's LinkedIn page.
- **instagramLink** (str): Instagram account URL.
- **logo** (str): URL linking to the organization's logo image.

### General

- **yearEstablished** (int): The year in which the organization was established.

- **acceptsVolunteers** (bool): Indicates whether the organization accepts clinical volunteers.

## Address

- **address\_line1** (str): Street address only (building number, street name). Do NOT include city, state, or country here.
- **address\_line2** (str): Additional street address information (apartment, suite, building name).
- **address\_line3** (str): Third line of street address if needed.
- **address\_city** (str): City or town name of the organization. Parse from comma-separated location strings if needed.
- **address\_stateOrRegion** (str): State, region, or province of the organization. Parse from comma-separated location strings if needed.
- **address\_zipOrPostcode** (str): ZIP or postal code of the organization.
- **address\_country** (str): Full country name of the organization. Always extract if country or country code information is present, using contextual clues from URL domain, phone numbers, or website content if not explicitly stated.
- **address\_countryCode** (str): ISO alpha-2 country code of the organization. Derive from country name if needed - this field is REQUIRED when country is known.

## Facility-Specific Fields

- **facilityTypeId** (enum): Type of facility. Levels: `hospital`, `pharmacy`, `doctor`, `clinic`, `dentist`.
- **operatorTypeId** (enum): Indicates if the facility is privately or publicly operated. Levels: `public`, `private`.
- **affiliationTypeIds** (list[enum]): Indicates facility affiliations (one or more). Levels: `faith-tradition`, `philanthropy-legacy`, `community`, `academic`, `government`.
- **description** (str): A brief paragraph describing the facility's services and/or history.
- **area** (int): Total floor area of the facility in square meters.
- **numberDoctors** (int): Total number of medical doctors working at the facility.
- **capacity** (int): Overall inpatient bed capacity of the facility.

## NGO-Specific Fields

- **countries** (list[str]): Countries where the NGO operates (array of ISO alpha-2 codes).
- **missionStatement** (str): The NGO's formal mission statement.
- **missionStatementLink** (str): A URL to the NGO's published mission statement.
- **organizationDescription** (str): A neutral, factual description derived from the mission statement. Removes explicitly religious or subjective language.

## Medical Specialties

- **specialties** (list[str]): The medical specialties associated with the organization. Must use exact case-sensitive matches from the specialty hierarchy (e.g., `internalMedicine`, `familyMedicine`, `pediatrics`, `cardiology`, `generalSurgery`, `emergencyMedicine`, `gynecologyAndObstetrics`, `orthopedicSurgery`, `dentistry`, `ophthalmology`). Choose the most specific appropriate specialty and only predict specialties clearly mentioned or strongly implied.

## Facility Facts (Free-form)

- **procedure** (list[str]): Specific clinical services performed at the facility—medical/surgical interventions and diagnostic procedures and screenings (e.g., operations, endoscopy, imaging- or lab-based tests). Each fact should be a clear, declarative statement including specific quantities when available (e.g., "Offers hemodialysis treatment 3 times weekly").
- **equipment** (list[str]): Physical medical devices and infrastructure—imaging machines (MRI/CT/X-ray), surgical/OR technologies, monitors, laboratory analyzers, and critical utilities (e.g., piped oxygen/oxygen plants, backup power). Include specific models when available (e.g., "Has Siemens SOMATOM Force dual-source CT scanner").
- **capability** (list[str]): Medical capabilities defining what level and types of clinical care the facility can deliver—trauma/emergency care levels (e.g., "Level II trauma center"), specialized units (ICU/NICU/burn unit), clinical programs (stroke care, IVF), diagnostic capabilities, clinical accreditations (e.g., "Joint Commission accredited"), care setting (inpatient/outpatient), staffing levels, and patient capacity. Excludes addresses, contact info, business hours, and pricing.