Revised Design — Based on Actual Product Data Analysis

Key Design Adjustments

Domain Shift

- **Original assumption**: Large-scale industrial refrigeration (screw compressors, evaporative condensers, chillers)

- **Actual domain**: Laboratory, pharmacy, and vaccine storage refrigerators and freezers

- **Impact**: Different spec taxonomy, different compliance frameworks (NSF/ANSI 456 vs ASHRAE 15), different user personas (lab managers, pharmacists, hospital procurement vs. facility engineers)

Product Families Identified (from samples + inference)

Family	Example Models	Key Differentiators
premier_lab_refrigerator	ABT-HC-26S, ABT-HC-49S	Solid/glass door, capacity steps, Premier controller
chromatography_refrigerator	ABT-HC-CS-26, ABT-HC-CS-47	Glass doors, chromatography-optimized interior
standard_lab_refrigerator	ABT-HC-30R	Natural refrigerants, newer variable-speed compressor
pharmacy_vaccine_refrigerator	PH-ABT-NSF-UCFS-0504	NSF/ANSI 456 certified, probe data, undercounter
lab_freezer	(not in samples)	Sub-zero ranges, manual/auto defrost
ultra_low_freezer	(not in samples)	-40°C to -86°C range
flammable_storage	(visible in images 2,5,6)	Explosion-proof, flammable material warnings
undercounter	PH-ABT-NSF-UCFS-0504	Compact form factor, freestanding or built-in
blood_bank	(not in samples)	Specific temperature range, FDA compliance
dual_temp	(not in samples)	Combination refrigerator/freezer

Revised Canonical Product Schema

```sql
sql
```

```sql
CREATE TABLE products (
    id                  UUID PRIMARY KEY DEFAULT gen_random_uuid(),
    model_number        TEXT NOT NULL,          -- 'ABT-HC-26S'
    family              TEXT NOT NULL,          -- 'premier_lab_refrigerator'
    subfamily           TEXT,                   -- 'solid_door', 'glass_door'
    product_line        TEXT,                   -- 'Premier', 'Standard', 'Pharmacy/Vaccine'
    brand               TEXT NOT NULL,          -- 'ABS' or 'LABRepCo'
    status              TEXT NOT NULL DEFAULT 'active',
    product_type        TEXT NOT NULL,          -- 'refrigerator' or 'freezer'
    controller_tier     TEXT,                   -- 'standard', 'ultra_touch', 'precision'

    -- ===== CAPACITY & STORAGE =====
    storage_capacity_cuft    NUMERIC,       -- 26, 30, 49, 5.2
    interior_volume_liters   NUMERIC,       -- auto-calculated: cuft * 28.3168

    -- ===== TEMPERATURE =====
    temp_range_min_c         NUMERIC,       -- 1
    temp_range_max_c         NUMERIC,       -- 10
    temp_setpoint_range_notes TEXT,         -- 'Minimum temperature limited to avoid freezing'

    -- ===== DOOR CONFIGURATION =====
    door_count          INTEGER,        -- 1, 2
    door_type           TEXT,           -- 'solid', 'glass', 'glass_sliding'
    door_hinge          TEXT,           -- 'right', 'left', 'right_and_left'
    door_features       TEXT[],         -- ['self_closing', 'magnetic_gasket', 'keyed_lock']
    interior_door       BOOLEAN DEFAULT false,

    -- ===== SHELVING =====
    shelf_count         INTEGER,        -- 4, 5, 8
    shelf_type          TEXT,           -- 'adjustable', 'fixed', 'mixed'
    shelf_adjustment_increment TEXT,    -- '1/2 inch'
    shelf_notes         TEXT,           -- 'guard rail on back'

    -- ===== DIMENSIONS (stored in inches as source, plus metric) =====
    exterior_width_in       NUMERIC,    -- 28.375
    exterior_depth_in       NUMERIC,    -- 36.75
    exterior_height_in      NUMERIC,    -- 81.75
    interior_width_in       NUMERIC,    -- 23.75
    interior_depth_in       NUMERIC,    -- 28
    interior_height_in      NUMERIC,    -- 52.25
    door_swing_in           NUMERIC,    -- 26.375
    total_open_depth_in     NUMERIC,    -- 63.125
    -- Metric equivalents (auto-calculated triggers or application layer)
```

```sql
    exterior_width_mm           NUMERIC GENERATED ALWAYS AS (exterior_width_in * 25.4) STORED,
    exterior_depth_mm           NUMERIC GENERATED ALWAYS AS (exterior_depth_in * 25.4) STORED,
    exterior_height_mm          NUMERIC GENERATED ALWAYS AS (exterior_height_in * 25.4) STORED,


    -- ===== WEIGHT =====
    product_weight_lbs          NUMERIC,        -- 235, 330, 396
    shipping_weight_lbs         NUMERIC,        -- 275, 360, 446
    product_weight_kg           NUMERIC GENERATED ALWAYS AS (product_weight_lbs * 0.453592) STORED,


    -- ===== ELECTRICAL =====
    voltage_v                   INTEGER,        -- 115 or 110-120
    voltage_min_v               INTEGER,        -- 110 (for range specs)
    voltage_max_v               INTEGER,        -- 120
    frequency_hz                INTEGER,        -- 60
    phase                       INTEGER,        -- 1
    amperage                    NUMERIC,        -- 3, 3.1, 4.5
    horsepower                  TEXT,           -- '1/5', '1/4'
    breaker_amps                INTEGER,        -- 15
    plug_type                   TEXT,           -- 'NEMA_5-15P'
    cord_length_ft              NUMERIC,        -- 6, 8-10


    -- ===== REFRIGERATION SYSTEM =====
    compressor_type             TEXT,           -- 'hermetic', 'hermetic_variable_speed'
    refrigerant                 TEXT,           -- 'R290', 'R600a'
    refrigerant_description     TEXT,           -- 'Hydrocarbon, natural refrigerant'
    condenser_type              TEXT,           -- 'static_exterior_walls', 'tube_and_grid_fanless'
    evaporator_type             TEXT,           -- 'fin_and_tube', 'plate_wall'
    defrost_type                TEXT,           -- 'cycle', 'off_cycle_no_heat'


    -- ===== PERFORMANCE (not all products have this) =====
    uniformity_c                NUMERIC,        -- 1.4 (±)
    stability_c                 NUMERIC,        -- 1.3 (±)
    max_temp_variation_c        NUMERIC,        -- 3.6
    energy_consumption_kwh_day  NUMERIC,        -- 1.15, 1.5
    heat_rejection_btu_hr       NUMERIC,        -- 237, 500
    noise_dba                   INTEGER,        -- 41
    pulldown_time_min           INTEGER,        -- 35, 42
    recovery_notes              TEXT,           -- 'All probes recover to under 8°C within 6 min'


    -- ===== FREEZER-SPECIFIC =====
    freezer_compartment_count   INTEGER,        -- 7 inner doors for manual defrost freezers
    inner_door_count            INTEGER,        -- same as above, alternate naming
    defrost_disclaimer          TEXT,           -- 'Auto defrost freezers incorporate an electric heater...'
    compressor_speed_min_rpm    INTEGER,        -- 2000 (for VSC units)
```

```sql
    compressor_speed_max_rpm    INTEGER,         -- 4500


    -- ===== FLAMMABLE STORAGE SPECIFIC =====
    nfpa_compliance             TEXT[],          -- ['NFPA_45', 'NFPA_30']
    atex_rated_interior         BOOLEAN DEFAULT false,
    intrinsically_safe          BOOLEAN DEFAULT false,
    flammable_disclaimer        TEXT,            -- 'NOT designed for use in volatile/explosive environments'


    -- ===== CONTROLLER & MONITORING =====
    controller_type             TEXT,            -- 'microprocessor', 'parametric_microprocessor', 'touchscreen_microprocessor'
    display_type                TEXT,            -- 'digital_temperature', 'led_0.1c_resolution', 'touchscreen', 'color_touchscreen_8in'
    display_resolution          TEXT,            -- '0.1°C'
    display_units_switchable    BOOLEAN DEFAULT false,  -- C/F switchable
    digital_communication       TEXT,            -- 'RS-485_MODBUS', null
    data_transfer               TEXT,            -- 'USB_csv_pdf', 'non_applicable'
    data_logging                BOOLEAN DEFAULT false,
    data_logging_intervals      TEXT,            -- '5, 10, or 15 minutes'
    chart_recorder              TEXT,            -- 'digital_24hr', 'non_applicable'
    battery_backup              TEXT,            -- null, 'optional_accessory', '12V_high_capacity'
    password_protection         TEXT,            -- null, 'single', 'multi_level_user_supervisor_admin'
    usb_port                    BOOLEAN DEFAULT false,
    four_twenty_ma_output       BOOLEAN DEFAULT false,  -- 4-20mA output


    -- ===== ALARMS =====
    alarms                      TEXT[],          -- ['high_low_temp', 'remote_contacts', 'sensor_error',
                                                 -- 'power_failure', 'min_max_history', 'door_ajar',
                                                 -- 'alarm_validation', 'alarm_mute_ringback']
    external_alarm_connection   TEXT,            -- 'remote_alarm_contacts', 'state_switching_remote'


    -- ===== CONSTRUCTION =====
    mounting_type               TEXT,            -- 'swivel_casters', 'leveling_legs'
    interior_lighting           TEXT,            -- 'led_shielded_switched', 'led_full_coverage_balanced'
    airflow_type                TEXT,            -- 'forced_draft', 'forced_air_patent_pending'
    probe_access                TEXT,            -- '3/4_inch_rear', '3/8_inch_rear'
    insulation_type             TEXT,            -- 'epa_urethane_foam'
    exterior_material           TEXT,            -- 'white_powder_coated_steel', 'powder_coated_steel'
    interior_material           TEXT,            -- 'white_powder_coated_steel'
    access_control              TEXT,            -- 'keyed_door_lock', 'pyxis_omnicell_acudose_compatible'


    -- ===== PHARMACY / VACCINE SPECIFIC =====
    cdc_tmd_compliant           BOOLEAN DEFAULT false,  -- Temperature Monitor Device per CDC guidelines
    nist_calibration            BOOLEAN DEFAULT false,  -- NIST traceable calibration certificate
    nist_calibration_years      INTEGER,                -- 3 years certification
    probe_count                 INTEGER,                -- 2 (1 air + 1 sample bottle)
```

```sql
    probe_configuration        TEXT,           -- '1 in air and 1 in sample bottle'
    vaccine_toolkit_included   BOOLEAN DEFAULT false,
    pyxis_omnicell_compatible  BOOLEAN DEFAULT false,


    -- ===== CERTIFICATIONS =====
    certifications             TEXT[],         -- ['ETL', 'C-ETL', 'UL471', 'Energy_Star',
                                               --  'NSF_ANSI_456', 'UL_60335-1', 'CSA_C22.2_No120',
                                               --  'EPA_SNAP']
    certification_standards    TEXT[],         -- specific standard references
    epa_snap_compliant         BOOLEAN DEFAULT false,
    energy_star_certified      BOOLEAN DEFAULT false,
    nsf_ansi_456_certified     BOOLEAN DEFAULT false,  -- vaccine storage


    -- ===== WARRANTY =====
    general_warranty_years     INTEGER,        -- 2
    compressor_warranty_years  INTEGER,        -- 5
    warranty_notes             TEXT,           -- 'excluding display probe calibration'


    -- ===== ACCESSORIES & OPTIONS =====
    included_accessories       TEXT[],         -- ['pharmacy_toolkit', 'temperature_logs']
    compatible_accessories     TEXT[],         -- ['wire_basket']
    options                    JSONB DEFAULT '{}',


    -- ===== INSTALLATION REQUIREMENTS =====
    ventilation_clearance_in   INTEGER,        -- 4 (inches on all sides)
    installation_notes         TEXT,           -- 'improper installation will void warranty...'
    operational_environment    TEXT,           -- 'Indoor use only. +18°C to +26°C, <70% RH'

    -- ===== PROBE TEMPERATURE DATA (vaccine units) =====
    -- Stored as JSONB for flexible probe count
    probe_temperature_data     JSONB,
    -- Example: [{"probe": 1, "avg": 3.9, "min": 2.7, "max": 5.2}, ...]


    -- ===== LIFECYCLE =====
    effective_from             DATE DEFAULT CURRENT_DATE,
    effective_to               DATE,
    replaced_by                UUID[],
    version                    INTEGER DEFAULT 1,
    approval_status            TEXT DEFAULT 'approved',
    created_at                 TIMESTAMPTZ DEFAULT now(),
    updated_at                 TIMESTAMPTZ DEFAULT now(),
    updated_by                 TEXT,


    -- ===== DOCUMENT REFERENCE =====
```

```sql
    data_sheet_doc_id        UUID,
    cut_sheet_doc_id         UUID,
    revision                 TEXT,          -- 'Rev_03.18.25', 'Rev_07232025'

    UNIQUE(model_number, version)
);
```

Revised Spec Taxonomy

```sql
    data_sheet_doc_id        UUID,
    cut_sheet_doc_id         UUID,
    revision                 TEXT,          -- 'Rev_03.18.25', 'Rev_07232025'
```

```sql
-- Based on actual field names found in sample documents
INSERT INTO spec_taxonomy (canonical_name, display_name, canonical_unit, data_type, synonyms, unit_aliases, family_s

-- Capacity
('storage_capacity_cuft', 'Storage Capacity', 'cu.ft.', 'numeric',
 ARRAY['capacity', 'volume', 'cu ft', 'cubic feet', 'gross volume'],
 '{"liters": 28.3168, "L": 28.3168}'::jsonb,
 ARRAY['all']),

-- Temperature
('temp_range_min_c', 'Min Temperature', '°C', 'numeric',
 ARRAY['minimum temperature', 'low temp', 'setpoint range low'],
 '{"F": "convert_f_to_c"}'::jsonb,
 ARRAY['all']),

('temp_range_max_c', 'Max Temperature', '°C', 'numeric',
 ARRAY['maximum temperature', 'high temp', 'setpoint range high'],
 '{"F": "convert_f_to_c"}'::jsonb,
 ARRAY['all']),

-- Door
('door_type', 'Door Type', '', 'enum',
 ARRAY['door', 'door style', 'door configuration'],
 '{}'::jsonb,
 ARRAY['all']),

('door_count', 'Number of Doors', '', 'numeric',
 ARRAY['doors', 'door quantity'],
 '{}'::jsonb,
 ARRAY['all']),

-- Dimensions (inches as canonical since all US-market)
('exterior_width_in', 'Width (Exterior)', 'in', 'numeric',
 ARRAY['width', 'w', 'external width'],
 '{"mm": 0.03937, "cm": 0.3937}'::jsonb,
 ARRAY['all']),

('exterior_depth_in', 'Depth (Exterior)', 'in', 'numeric',
 ARRAY['depth', 'd', 'external depth'],
 '{"mm": 0.03937, "cm": 0.3937}'::jsonb,
 ARRAY['all']),

('exterior_height_in', 'Height (Exterior)', 'in', 'numeric',
```

```sql
  ARRAY['height', 'h', 'external height'],
  '{"mm": 0.03937, "cm": 0.3937}'::jsonb,
  ARRAY['all']),

-- Electrical
('voltage_v', 'Voltage', 'V', 'numeric',
 ARRAY['voltage', 'volts', 'VAC', 'supply voltage'],
 '{}'::jsonb,
 ARRAY['all']),

('amperage', 'Rated Amperage', 'A', 'numeric',
 ARRAY['amps', 'rated amps', 'current draw', 'rated amperage'],
 '{}'::jsonb,
 ARRAY['all']),

('plug_type', 'Power Plug', '', 'enum',
 ARRAY['plug', 'power cord', 'NEMA plug', 'power plug/power cord'],
 '{}'::jsonb,
 ARRAY['all']),

-- Refrigeration
('refrigerant', 'Refrigerant', '', 'enum',
 ARRAY['refrigerant type', 'gas type', 'hydrocarbon'],
 '{}'::jsonb,
 ARRAY['all']),

('compressor_type', 'Compressor', '', 'enum',
 ARRAY['compressor', 'compressor technology'],
 '{}'::jsonb,
 ARRAY['all']),

('defrost_type', 'Defrost Method', '', 'enum',
 ARRAY['defrost', 'defrost system', 'defrost type'],
 '{}'::jsonb,
 ARRAY['all']),

-- Performance
('uniformity_c', 'Temperature Uniformity', '°C (±)', 'numeric',
 ARRAY['uniformity', 'cabinet air uniformity', 'temperature uniformity'],
 '{}'::jsonb,
 ARRAY['standard_lab_refrigerator', 'pharmacy_vaccine_refrigerator']),

('stability_c', 'Temperature Stability', '°C (±)', 'numeric',
 ARRAY['stability', 'cabinet air stability', 'temperature stability'],
```

```sql
 '{}'::jsonb,
 ARRAY['standard_lab_refrigerator', 'pharmacy_vaccine_refrigerator']),


('energy_consumption_kwh_day', 'Energy Consumption', 'kWh/day', 'numeric',
 ARRAY['energy consumption', 'daily energy', 'power consumption'],
 '{"kWh/yr": 0.00274}'::jsonb,
 ARRAY['standard_lab_refrigerator', 'pharmacy_vaccine_refrigerator']),


('noise_dba', 'Noise Level', 'dBA', 'numeric',
 ARRAY['noise', 'sound level', 'noise pressure level', 'sound pressure'],
 '{}'::jsonb,
 ARRAY['pharmacy_vaccine_refrigerator']),


('pulldown_time_min', 'Pull Down Time', 'min', 'numeric',
 ARRAY['pull down time', 'pulldown', 'cool down time'],
 '{}'::jsonb,
 ARRAY['standard_lab_refrigerator', 'pharmacy_vaccine_refrigerator']),


-- Weight
('product_weight_lbs', 'Product Weight', 'lbs', 'numeric',
 ARRAY['weight', 'unit weight', 'net weight'],
 '{"kg": 2.20462}'::jsonb,
 ARRAY['all']),


-- Certifications (as searchable attributes)
('certifications', 'Certifications', '', 'list',
 ARRAY['agency listing', 'agency listing and certification', 'listings', 'certified'],
 '{}'::jsonb,
 ARRAY['all']),


-- Controller
('controller_type', 'Controller Technology', '', 'enum',
 ARRAY['controller', 'controller technology', 'temperature controller'],
 '{}'::jsonb,
 ARRAY['all']),


('digital_communication', 'Digital Communication', '', 'enum',
 ARRAY['communication', 'data interface', 'MODBUS', 'RS-485'],
 '{}'::jsonb,
 ARRAY['premier_lab_refrigerator']);
```

Model Number Pattern Analysis

```python
MODEL_PATTERNS = {
    # Premier Chromatography Series: ABT-HC-CS-{capacity}
    r'^ABT-HC-CS-(\d+)'
}
```

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	cut_sheet	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	product_data_sheet	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	product_image	Image-only file, no structured data
Product photo with dimensional annotations	dimensional_drawing	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	performance_data_sheet	NSF/ANSI 456 probe data, temperature charts

---

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:

```python

```

```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',       # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',

    'W"': 'interior_width_in',       # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
```

```python
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']

    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)

```sql
sql

INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

: {
'brand': 'ABS',
'family': 'chromatography_refrigerator',
'product_line': 'Premier',
'product_type': 'refrigerator',
'capacity_field': 'group_1',
},

```
# Premier Lab Series: ABT-HC-{capacity}{door_type}
r'^ABT-HC-(\d+)(S|G)
```

}

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	`cut_sheet`	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	`product_data_sheet`	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	`product_image`	Image-only file, no structured data
Product photo with dimensional annotations	`dimensional_drawing`	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	`performance_data_sheet`	NSF/ANSI 456 probe data, temperature charts

---

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:
```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',
```

```python
    'W"': 'interior_width_in',       # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']
```

```
    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)
```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

```
: {
    'brand': 'ABS',
    'family': 'premier_lab_refrigerator',
    'product_line': 'Premier',
    'product_type': 'refrigerator',
    'capacity_field': 'group_1',
    'door_type_map': {'S': 'solid', 'G': 'glass'},
},

# Standard Lab Series: ABT-HC-{capacity}R
r'^ABT-HC-(\d+)R

}
```

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	cut_sheet	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	product_data_sheet	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	product_image	Image-only file, no structured data
Product photo with dimensional annotations	dimensional_drawing	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	performance_data_sheet	NSF/ANSI 456 probe data, temperature charts

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:

```python
```

```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',       # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',

    'W"': 'interior_width_in',       # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
```

```python
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}


# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']

    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)

```sql
sql

INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

: {
'brand': 'ABS',
'family': 'standard_lab_refrigerator',
'product_line': 'Standard',
'product_type': 'refrigerator',
'capacity_field': 'group_1',
},

```
# ABS Pharmacy Premier: PH-ABT-HC-{capacity}{door_type}
r'^PH-ABT-HC-(\d+)(S|G)
```

}

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	`cut_sheet`	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	`product_data_sheet`	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	`product_image`	Image-only file, no structured data
Product photo with dimensional annotations	`dimensional_drawing`	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	`performance_data_sheet`	NSF/ANSI 456 probe data, temperature charts

---

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:
```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',
```

```python
    'W"': 'interior_width_in',      # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']
```

```python
    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)
```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

```
    : {
        'brand': 'ABS',
        'family': 'pharmacy_vaccine_refrigerator',
        'product_line': 'Pharmacy Premier',
        'product_type': 'refrigerator',
        'capacity_field': 'group_1',
        'door_type_map': {'S': 'solid', 'G': 'glass'},
    },

    # ABS Pharmacy NSF Undercounter: PH-ABT-NSF-UCFS-{code}
    r'^PH-ABT-NSF-UCFS-(\d+)
}
```

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	cut_sheet	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	product_data_sheet	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	product_image	Image-only file, no structured data
Product photo with dimensional annotations	dimensional_drawing	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	performance_data_sheet	NSF/ANSI 456 probe data, temperature charts

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:

```
python
```

```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',

    'W"': 'interior_width_in',      # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
```

```python
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']

    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)

```sql
sql

INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

: {
'brand': 'ABS',
'family': 'pharmacy_vaccine_nsf',
'product_line': 'Pharmacy NSF',
'product_type': 'refrigerator',
'nsf_ansi_456': True,
},

```
# LABRepCo Ultra Touch Manual Defrost Freezer: LHT-{capacity}-FMP
r'^LHT-(\d+)-FMP
```

}

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	`cut_sheet`	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	`product_data_sheet`	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	`product_image`	Image-only file, no structured data
Product photo with dimensional annotations	`dimensional_drawing`	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	`performance_data_sheet`	NSF/ANSI 456 probe data, temperature charts

---

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:
```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',
```

```python
    'W"': 'interior_width_in',        # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',    # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']
```

```
  'shelf_config_raw': 'parse_shelf_config',
  # "Four adjustable shelves (adjustable in ½" increments)"
  # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

  'temp_range_raw': 'parse_temp_range',
  # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

  'refrigerant_raw': 'parse_refrigerant',
  # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
  # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

  'electrical_raw': 'parse_electrical',
  # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
  # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

  'certifications_raw': 'parse_certifications',
  # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
  # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

  'dimension_fraction': 'parse_fraction',
  # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)
```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

```
    : {
        'brand': 'LABRepCo',
        'family': 'manual_defrost_freezer',
        'product_line': 'Ultra Touch',
        'product_type': 'freezer',
        'controller_tier': 'ultra_touch',
        'capacity_field': 'group_1',
    },

    # LABRepCo FUTURA Auto Defrost Freezer: LHT-{capacity}-FASS
    r'^LHT-(\d+)-FASS

}
```

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	cut_sheet	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	product_data_sheet	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	product_image	Image-only file, no structured data
Product photo with dimensional annotations	dimensional_drawing	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	performance_data_sheet	NSF/ANSI 456 probe data, temperature charts

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:

```python
```

```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',       # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',

    'W"': 'interior_width_in',       # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
```

```python
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']

    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)

```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

: {
'brand': 'LABRepCo',
'family': 'auto_defrost_freezer',
'product_line': 'Ultra Touch FUTURA',
'product_type': 'freezer',
'controller_tier': 'ultra_touch',
'capacity_field': 'group_1',
},

```
# LABRepCo FUTURA Manual Defrost Freezer: LHT-{capacity}-FM
r'^LHT-(\d+)-FM
```

}

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	`cut_sheet`	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	`product_data_sheet`	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	`product_image`	Image-only file, no structured data
Product photo with dimensional annotations	`dimensional_drawing`	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	`performance_data_sheet`	NSF/ANSI 456 probe data, temperature charts

---

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:
```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',
```

```python
    'W"': 'interior_width_in',      # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']
```

```
  'shelf_config_raw': 'parse_shelf_config',
  # "Four adjustable shelves (adjustable in ½" increments)"
  # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

  'temp_range_raw': 'parse_temp_range',
  # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

  'refrigerant_raw': 'parse_refrigerant',
  # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
  # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

  'electrical_raw': 'parse_electrical',
  # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
  # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

  'certifications_raw': 'parse_certifications',
  # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
  # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

  'dimension_fraction': 'parse_fraction',
  # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)
```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

```
: {
    'brand': 'LABRepCo',
    'family': 'manual_defrost_freezer',
    'product_line': 'FUTURA',
    'product_type': 'freezer',
    'controller_tier': 'ultra_touch',
    'capacity_field': 'group_1',
},

# LABRepCo Ultra Touch Flammable Refrigerator: LHT-{capacity}-RFP
r'^LHT-(\d+)-RFP

}
```

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	cut_sheet	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	product_data_sheet	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	product_image	Image-only file, no structured data
Product photo with dimensional annotations	dimensional_drawing	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	performance_data_sheet	NSF/ANSI 456 probe data, temperature charts

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:

```python
```

```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',

    'W"': 'interior_width_in',      # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
```

```python
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']

    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)

```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

: {
'brand': 'LABRepCo',
'family': 'flammable_storage_refrigerator',
'product_line': 'Ultra Touch',
'product_type': 'refrigerator',
'controller_tier': 'ultra_touch',
'capacity_field': 'group_1',
},

```
# LABRepCo Precision Freezer: LPVT-{capacity}-FA
r'^LPVT-(\d+)-FA
```

}

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	`cut_sheet`	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	`product_data_sheet`	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	`product_image`	Image-only file, no structured data
Product photo with dimensional annotations	`dimensional_drawing`	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	`performance_data_sheet`	NSF/ANSI 456 probe data, temperature charts

---

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:
```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',
```

```python
    'W"': 'interior_width_in',       # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',    # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']
```

```
    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)
```sql
INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```

```
: {
    'brand': 'LABRepCo',
    'family': 'precision_freezer',
    'product_line': 'Precision',
    'product_type': 'freezer',
    'controller_tier': 'precision',
    'capacity_field': 'group_1',
  },
}
```

Document Type Classification Rules

Based on the sample files:

Pattern	Doc Type	Key Indicators
Has "CUTSHEET" header, single-page, compact table	cut_sheet	"CUTSHEET" text, dimensional drawings, single spec table
Has "Product Data Sheet" header, multi-page, structured sections	product_data_sheet	"Product Data Sheet", section headers like "Refrigeration System", "Dimensions"
Product photo with no text overlay	product_image	Image-only file, no structured data
Product photo with dimensional annotations	dimensional_drawing	Image with measurement callouts
Has "Performance" section with probe data, uniformity, stability	performance_data_sheet	NSF/ANSI 456 probe data, temperature charts

Extraction Field Mapping (Document → Schema)

Fields appear differently across document types. The extraction pipeline needs this mapping:

```python
```

```python
FIELD_MAPPINGS = {
    # Source field name → canonical column name
    'Storage capacity (cu. ft)': 'storage_capacity_cuft',
    'Storage capacity': 'storage_capacity_cuft',
    'Cu. Ft': 'storage_capacity_cuft',

    'Door': 'door_config_raw',      # needs parsing: "One swing solid door, self-closing, right hinged"
    'Int Door': 'interior_door',

    'Shelves': 'shelf_config_raw',   # needs parsing: "Four adjustable shelves (adjustable in ½" increments)"

    'Adjustable Temperature Range': 'temp_range_raw',  # needs parsing: "1°C to 10°C"
    'Temperature setpoint range': 'temp_range_raw',

    'Refrigerant': 'refrigerant_raw',  # "Hydrocarbon, natural refrigerant (R290)" → 'R290'

    'Compressor': 'compressor_type',

    'Defrost': 'defrost_type',

    'W"': 'interior_width_in',      # cut sheet format
    'D"': 'interior_depth_in',
    'H"': 'interior_height_in',

    'Width (in.)': 'width_in_raw',   # context needed: exterior vs interior
    'Depth (in.)': 'depth_in_raw',
    'Height (in.)': 'height_in_raw',

    'Rated Amperage': 'amperage',
    'Amps': 'amperage',

    'H.P.': 'horsepower',

    'Weight': 'product_weight_lbs',
    'Product Weight (lbs)': 'product_weight_lbs',
    'Shipping Weight (lbs)': 'shipping_weight_lbs',

    'Power Plug/Power Cord': 'plug_type_raw',
    'Facility Electrical Requirement': 'electrical_raw',

    'Agency Listing and Certification': 'certifications_raw',

    'Controller technology': 'controller_type',
```

```python
    'Display technology': 'display_type',
    'Digital Communication': 'digital_communication',

    'External alarm connection': 'external_alarm_connection',

    'Uniformity (Cabinet air)': 'uniformity_c',
    'Stability (Cabinet air)': 'stability_c',
    'Maximum temperature variation': 'max_temp_variation_c',
    'Energy Consumption (KWh/day)': 'energy_consumption_kwh_day',
    'Average Heat Rejection (BTU/hr)': 'heat_rejection_btu_hr',
    'Noise pressure level (dBA)': 'noise_dba',
    'Pull down time to nominal operating temp': 'pulldown_time_min',
}

# Parsers for complex fields
FIELD_PARSERS = {
    'door_config_raw': 'parse_door_config',
    # "One swing solid door, self-closing, right hinged"
    # → door_count=1, door_type='solid', door_hinge='right', door_features=['self_closing']

    'shelf_config_raw': 'parse_shelf_config',
    # "Four adjustable shelves (adjustable in ½" increments)"
    # → shelf_count=4, shelf_type='adjustable', shelf_adjustment_increment='1/2 inch'

    'temp_range_raw': 'parse_temp_range',
    # "1°C to 10°C" → temp_range_min_c=1, temp_range_max_c=10

    'refrigerant_raw': 'parse_refrigerant',
    # "Hydrocarbon, natural refrigerant (R290)" → refrigerant='R290'
    # "EPA SNAP compliant R600a Isobutane" → refrigerant='R600a'

    'electrical_raw': 'parse_electrical',
    # "115V, 60 Hz, 3 Amps, 1/5 HP" → voltage_v=115, frequency_hz=60, amperage=3, hp='1/5'
    # "110 - 120V AC, 15A breaker, NEMA 5-15 receptacle" → voltage_min=110, voltage_max=120, breaker=15

    'certifications_raw': 'parse_certifications',
    # "ETL, C-ETL listed and certified to UL471 standard, Energy Star Certified"
    # → ['ETL', 'C-ETL', 'UL471', 'Energy_Star']

    'dimension_fraction': 'parse_fraction',
    # "23 ¾" → 23.75, "48 ⅝" → 48.625
}
```

Equivalence Rules (Revised for Lab Refrigeration)

```sql
sql

INSERT INTO equivalence_rules (family, rule_name, required_match, tolerance_map, priority_specs) VALUES

('premier_lab_refrigerator', 'capacity_match',
 ARRAY['door_type', 'refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.20, "product_weight_lbs": 0.30}'::jsonb,
 ARRAY['storage_capacity_cuft', 'energy_consumption_kwh_day', 'shelf_count']),

('pharmacy_vaccine_refrigerator', 'vaccine_storage_match',
 ARRAY['nsf_ansi_456_certified', 'voltage_v', 'door_type'],
 '{"storage_capacity_cuft": 0.20, "uniformity_c": 0.15}'::jsonb,
 ARRAY['uniformity_c', 'stability_c', 'storage_capacity_cuft']),

('standard_lab_refrigerator', 'standard_match',
 ARRAY['refrigerant', 'voltage_v'],
 '{"storage_capacity_cuft": 0.15, "amperage": 0.25}'::jsonb,
 ARRAY['energy_consumption_kwh_day', 'storage_capacity_cuft', 'uniformity_c']);
```