

FAO Database Exploration Guide

Table of Contents

1. [Database Overview](#)
 2. [Database Structure](#)
 3. [Reference Tables Guide](#)
 4. [Dataset Tables by Category](#)
 5. [Exploration Strategy](#)
 6. [Sample Queries by Topic](#)
 7. [Analysis Tips](#)
-

Database Overview

The Food and Agriculture Organization (FAO) database is a comprehensive repository of global agricultural, food security, and environmental data. This database contains **84 tables** with data spanning multiple decades, covering everything from crop production to greenhouse gas emissions, trade flows, and food security indicators.

Key Features:

- **Global Coverage:** Data from virtually every country and territory
- **Time Series:** Most tables contain annual data, some with monthly granularity
- **Multi-dimensional:** Each data point is characterized by location, commodity, measurement type, and time
- **Quality Indicators:** Built-in data quality flags for transparency

Database Statistics:

- **Total Tables:** 84 (14 reference + 70 dataset tables)
 - **Primary Domains:** Production, Trade, Food Security, Environment, Investment, Prices
 - **Typical Time Range:** 1961-2023 (varies by dataset)
 - **Geographic Coverage:** 200+ countries and territories
-

Database Structure

Data Model Overview

The database follows a star schema pattern where:

- **Reference tables** serve as dimension tables (prefixed with lookups like area_codes, item_codes)
- **Dataset tables** serve as fact tables containing measurements
- **Foreign keys** link dataset tables to reference tables using `_id` suffix

Common Table Structure Pattern

Most dataset tables follow this structure:

- `id`: Primary key
 - `area_code_id`: Foreign key to `area_codes` (country/region)
 - `item_code_id`: Foreign key to `item_codes` (commodity)
 - `element_code_id`: Foreign key to `elements` (measurement type)
 - `flag_id`: Foreign key to `flags` (data quality)
 - `year`: The year of observation
 - `value`: The actual measurement
 - `unit`: Unit of measurement
 - Various other specific columns
-

Reference Tables Guide

1. Geographic References

`area_codes` (Primary geographic reference)

- Contains country codes, names, and M49 codes
- Links to almost every dataset table
- Includes aggregated regions (e.g., "World", "Africa", "European Union")

`geographic_levels`

- Defines geographic granularity (national, subnational, etc.)
- Used primarily in survey data tables

2. Commodity References

`item_codes` (Primary commodity reference)

- Comprehensive list of agricultural products
- Includes crops, livestock, processed products
- Contains cross-references to other classification systems (CPC, FBS, SDG)

food_groups

- Higher-level categorization of food items
- Used in dietary and nutrition analysis

3. Measurement References

elements (Types of measurements)

- Defines what is being measured (Production, Yield, Area Harvested, Import Quantity, etc.)
- Critical for understanding the meaning of values in dataset tables

indicators

- Specific indicators used in analytical datasets
- Common in food security and employment tables

4. Data Quality References

flags

- Single-character codes indicating data source/quality
- Examples: "A" (Aggregate), "E" (Estimated), "M" (Missing), "O" (Official)

5. Other Important References

- **currencies:** ISO currency codes
- **donors:** Development assistance sources
- **purposes:** Categories for development aid
- **sources:** Data sources
- **surveys:** Household survey identifiers

Dataset Tables by Category

Production & Agricultural Output (11 tables)

Core Production Data:

- **production_crops_livestock:** Primary production statistics for all agricultural commodities
 - Coverage: Quantities produced, areas harvested, yields, livestock numbers
 - Time span: Usually 1961-present

- Key uses: Track agricultural output trends, productivity analysis
- `production_indices`: Index numbers showing production trends
 - Base period comparison (typically 2014-2016 = 100)
 - Useful for comparing growth rates across countries
- `value_of_production`: Monetary value of agricultural production
 - Constant and current prices
 - Essential for economic analysis

Supply Utilization:

- `sua_crops_livestock`: Detailed supply and utilization accounts
- `commodity_balances_*`: Non-food uses of agricultural products

Trade & Market Access (8 tables)

Trade Flows:

- `trade_crops_livestock`: Bilateral trade data
 - Import/export quantities and values
 - Partner country details
- `trade_detailed_trade_matrix`: Who trades what with whom
 - Reporter and partner country pairs
 - Detailed commodity breakdown
- `fertilizers_detailed_trade_matrix`: Specialized fertilizer trade flows

Trade Analytics:

- `trade_indices`: Trade performance indicators
- `trade_crops_livestock_indicators`: Derived trade metrics

Food Security & Nutrition (12 tables)

Food Balance:

- `food_balance_sheets`: Comprehensive food supply/utilization
 - Per capita food availability
 - Dietary energy supply
 - Critical for hunger analysis

Food Security Indicators:

- `food_security_data`: Suite of food security indicators
 - Prevalence of undernourishment
 - Food insecurity experience scale data

Nutrition & Diet:

- `cost_affordability_healthy_diet_co_ahd`: Diet affordability metrics
- Various household survey tables: Detailed consumption patterns

Environmental & Climate (19 tables)

Emissions Data: Eight specialized emissions tables covering:

- `emissions_crops`: Crop-related GHG emissions
- `emissions_livestock`: Animal-related emissions
- `emissions_land_use_forests`: Forestry emissions/removals
- `emissions_totals`: Aggregated emissions data

Resource Use:

- `inputs_land_use`: Agricultural land utilization
- `environment_land_cover`: Land cover changes over time
- `aquastat`: Water resources and irrigation

Environmental Indicators:

- `environment_temperature_change`: Climate change impacts
- `environment_emissions_intensities`: Efficiency metrics

Economic & Financial (13 tables)

Prices:

- `prices`: Producer and retail prices
 - Monthly and annual data
 - Local currency units
- `consumer_price_indices`: Food inflation tracking
- `exchange_rate`: Currency conversion data

Investment:

- Multiple `investment_*` tables covering:
 - Government expenditure
 - Foreign direct investment
 - Capital stock
 - Agricultural credit

Development Aid:

- `development_assistance_to_agriculture`: Aid flows to agriculture

Agricultural Inputs (8 tables)

Fertilizers:

- `inputs_fertilizers_nutrient`: Nutrient-based fertilizer data
- `inputs_fertilizers_product`: Product-based fertilizer data

Pesticides:

- `inputs_pesticides_use`: Pesticide application data
- `inputs_pesticides_trade`: Pesticide trade flows

Specialized Datasets (11 tables)

Forestry:

- `forestry`: Forest product statistics
- `forestry_trade_flows`: Timber trade

Population & Labor:

- `population`: Rural/urban population data
- `employment_indicators_*`: Agricultural employment

Exploration Strategy

Phase 1: Foundation (Week 1)

1. Map the Reference Tables

- Query each reference table to understand available values

- Create lookup dictionaries for common codes
- Identify data quality patterns from flags

2. Assess Data Coverage

- Determine year ranges for each major dataset
- Identify countries with most complete data
- Note any major data gaps

Phase 2: Core Agricultural Data (Week 2-3)

1. Production Analysis

- Start with major commodities (cereals, meat, dairy)
- Compare production trends across regions
- Calculate productivity changes over time

2. Trade Patterns

- Identify major importers/exporters
- Analyze trade balance changes
- Track commodity price relationships

Phase 3: Food Security Deep Dive (Week 4)

1. Food Balance Analysis

- Calculate per capita food availability
- Compare dietary patterns across countries
- Identify food security hotspots

2. Nutrition Indicators

- Analyze undernourishment trends
- Explore diet affordability data
- Link to production/trade patterns

Phase 4: Environmental Analysis (Week 5)

1. Emissions Tracking

- Calculate agricultural emissions by source
- Compare emission intensities
- Identify mitigation opportunities

2. Resource Use

- Analyze land use changes
- Track water resource utilization
- Assess input use efficiency

Phase 5: Economic Integration (Week 6)

1. Price Analysis

- Track commodity price volatility
- Analyze price transmission
- Compare local vs. international prices

2. Investment Flows

- Assess agricultural investment trends
 - Compare public vs. private investment
 - Link investment to productivity changes
-

Sample Queries by Topic

Understanding Data Coverage

```
sql

-- What years and countries have the most complete data?
WITH data_coverage AS (
    SELECT
        ac.area,
        pcl.year,
        COUNT(DISTINCT ic.item_code) AS commodities_covered,
        COUNT(DISTINCT ec.element_code) AS elements_covered,
        COUNT(*) AS total_records
    FROM production_crops_livestock pcl
    JOIN area_codes ac ON ac.id = pcl.area_code_id
    JOIN item_codes ic ON ic.id = pcl.item_code_id
    JOIN elements ec ON ec.id = pcl.element_code_id
    GROUP BY ac.area, pcl.year
)
SELECT
    area,
    MIN(year) AS earliest_year,
    MAX(year) AS latest_year,
    AVG(commodities_covered) AS avg_commodities,
    AVG(total_records) AS avg_records_per_year
FROM data_coverage
GROUP BY area
HAVING COUNT(DISTINCT year) > 20
ORDER BY avg_records_per_year DESC
LIMIT 20;
```

Production Analysis

```
sql
-- Top 10 wheat producers in latest available year
SELECT
    ac.area,
    pcl.year,
    pcl.value as production_tonnes,
    pcl.unit
FROM production_crops_livestock pcl
JOIN area_codes ac ON ac.id = pcl.area_code_id
JOIN item_codes ic ON ic.id = pcl.item_code_id
JOIN elements ec ON ec.id = pcl.element_code_id
WHERE ic.item = 'Wheat'
    AND ec.element = 'Production'
    AND pcl.year = (SELECT MAX(year) FROM production_crops_livestock WHERE item_code_id = pcl.item_code_id)
ORDER BY pcl.value DESC
LIMIT 10;
```

Food Security Analysis

```
sql
-- Countries with highest undernourishment
SELECT
    ac.area,
    ic.item as indicator,
    fsd.year,
    fsd.value,
    fsd.unit
FROM food_security_data fsd
JOIN area_codes ac ON ac.id = fsd.area_code_id
JOIN item_codes ic ON ic.id = fsd.item_code_id
WHERE ic.item LIKE '%undernourishment%'
    AND fsd.year >= 2020
    AND fsd.value IS NOT NULL
ORDER BY fsd.value DESC
LIMIT 20;
```

Trade Network Analysis

sql

```
-- Major wheat trade relationships
SELECT
    reporter.area as exporter,
    partner.area as importer,
    ic.item,
    dtm.year,
    dtm.value as trade_volume,
    dtm.unit
FROM trade_detailed_trade_matrix dtm
JOIN item_codes ic ON ic.id = dtm.item_code_id
JOIN elements ec ON ec.id = dtm.element_code_id
JOIN (SELECT DISTINCT area_code, area FROM area_codes) reporter
    ON reporter.area_code = dtm.reporter_country_code
JOIN (SELECT DISTINCT area_code, area FROM area_codes) partner
    ON partner.area_code = dtm.partner_country_code
WHERE ic.item = 'Wheat'
    AND ec.element = 'Export Quantity'
    AND dtm.year = 2022
    AND dtm.value > 100000
ORDER BY dtm.value DESC
LIMIT 20;
```

Environmental Impact

```

sql
-- Agricultural emissions by country
SELECT
    ac.area,
    SUM(CASE WHEN ic.item LIKE '%CH4%' THEN et.value ELSE 0 END) as methane_emissions,
    SUM(CASE WHEN ic.item LIKE '%N2O%' THEN et.value ELSE 0 END) as nitrous_oxide_emissions,
    SUM(CASE WHEN ic.item LIKE '%CO2%' THEN et.value ELSE 0 END) as co2_emissions,
    et.year,
    et.unit
FROM emissions_totals et
JOIN area_codes ac ON ac.id = et.area_code_id
JOIN item_codes ic ON ic.id = et.item_code_id
WHERE et.year = 2021
    AND ac.area_code NOT IN (SELECT area_code FROM area_codes WHERE area LIKE '%World%')
GROUP BY ac.area, et.year, et.unit
ORDER BY (methane_emissions + nitrous_oxide_emissions + co2_emissions) DESC
LIMIT 20;

```

Analysis Tips

1. Data Quality Considerations

- Always check the `flag` field to understand data reliability
- Official data (flag 'O') is generally most reliable
- Estimated data (flag 'E') may have higher uncertainty
- Some time series have methodology breaks - check metadata

2. Common Pitfalls to Avoid

- **Unit Confusion:** Always verify units (tonnes vs 1000 tonnes, hectares vs 1000 hectares)
- **Geographic Aggregations:** "World" and regional totals are included - avoid double counting
- **Missing Data:** Use appropriate NULL handling in calculations
- **Year Coverage:** Different datasets have different time spans

3. Performance Optimization

- Create indexes on commonly queried columns (year, area_code_id, item_code_id)
- Use materialized views for complex recurring analyses
- Partition large tables by year if needed

- Consider summary tables for dashboard applications

4. Advanced Analysis Ideas

- **Time Series Analysis:** Trend detection, seasonality, forecasting
- **Network Analysis:** Trade flow networks, market integration
- **Efficiency Metrics:** Yield gaps, emission intensities, water productivity
- **Policy Analysis:** Impact of trade agreements, climate policies
- **Food System Analysis:** Link production → trade → food security

5. Visualization Recommendations

- **Maps:** Choropleth maps for country-level data
 - **Time Series:** Line charts for trends, area charts for composition
 - **Networks:** Sankey diagrams for trade flows
 - **Comparisons:** Small multiples for multi-country analysis
 - **Dashboards:** Combine multiple views for comprehensive analysis
-

Next Steps

1. Set Up Your Environment

- Install necessary database tools
- Create indexes for better performance
- Set up a version control system for queries

2. Create Base Views

- Commonly joined reference tables
- Standardized metric calculations
- Data quality filters

3. Build Analysis Templates

- Reusable query patterns
- Standard visualizations
- Export formats for reporting

4. Document Your Findings

- Keep notes on data quirks
- Document any data cleaning steps

- Share insights with the community
-

This guide is a living document. As you explore the database, feel free to add your own insights and query patterns.