



Introducing a Comprehensive AI-Enabled Food Data Hub

Marrying a seven-domain food-system ontology with a federated database ecosystem
Behrouz Arabi and Sonal Choudhary
Birmingham Business School, University of Birmingham

Focus: build a metadata-rich, AI-enabled food data catalogue that links conceptual domains, indicators and heterogeneous data sources for monitoring, scenario planning and anticipatory governance

7 domains

134 platforms

8 families

2 standards

Why this matters

We frame food systems as a poly-crisis that requires resilient data infrastructure, not just better isolated indicators.

Food-system governance is constrained by fragmented data, inconsistent definitions and weak integration across nutrition, access, sustainability, risk and justice.

828M
People hungry in 2021

2.3B
Moderate or severe food insecurity

~1/3
Global GHG emissions come from food systems

70%
Freshwater use is linked to food production

Persistent design gaps identified in the paper

- Fragmented and inconsistent datasets limit reuse
 - Weak disaggregation obscures vulnerable groups
 - Reactive monitoring weakens anticipatory governance
 - Local and community knowledge often remains uncatalogued
- The result is an extendable, monitorable and policy-facing data backbone

What the paper builds

We explicitly link the ontology and metadata logic to a broad database landscape.

Abstract

A metadata-rich food data catalogue can translate dispersed food-system concepts into a practical, searchable infrastructure that links indicators, datasets and decision-support uses.

1 Systematic mapping
Review + coding build the indicator library

2 Seven-domain ontology
Security, resilience, risk, sustainability, governance, justice, transformation

3 Federated catalogue
Metadata, source links and AI-ready discovery over linked repositories

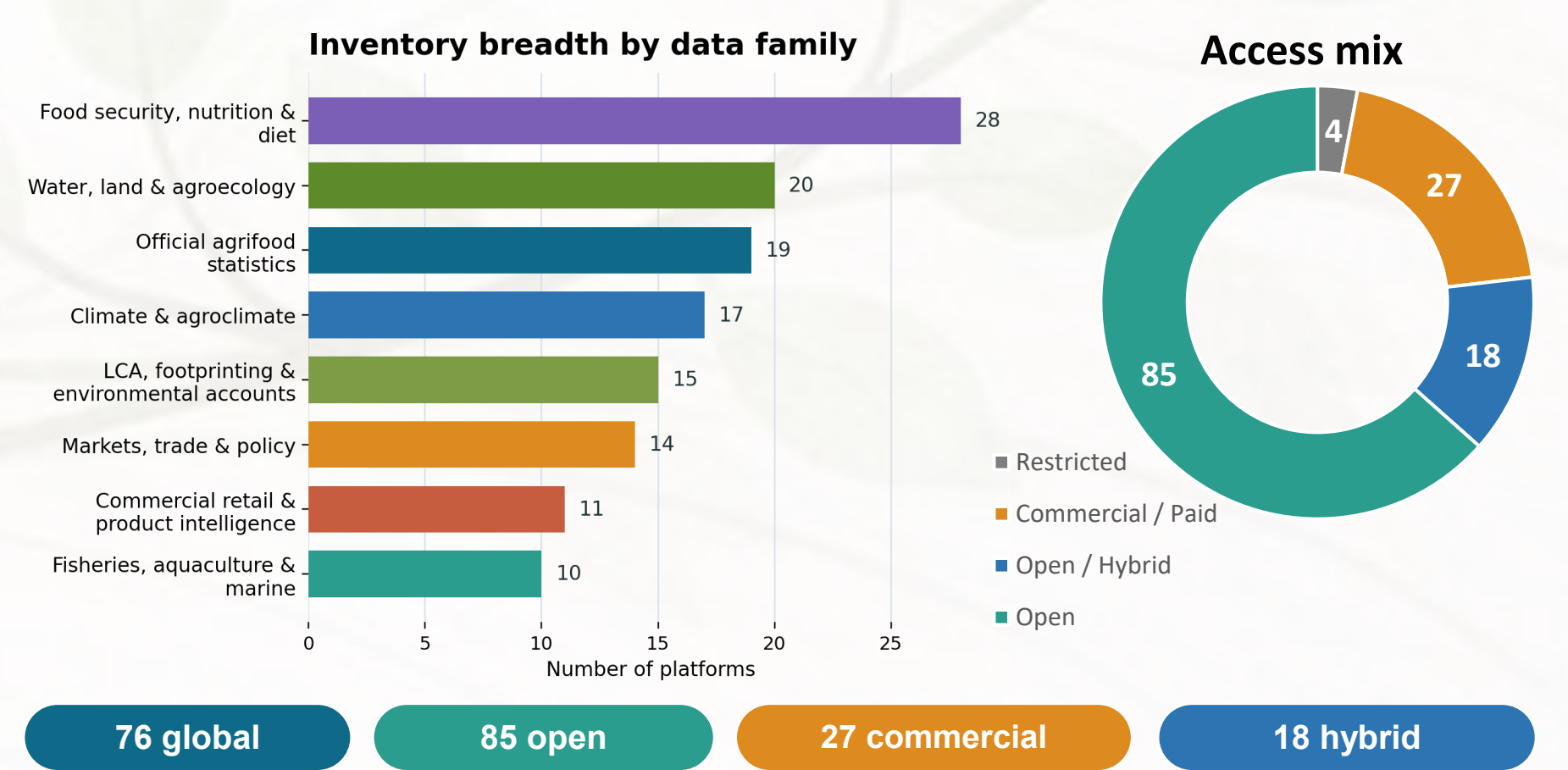
Metadata standards and searchable fields

| domain / theme | keywords | geography | temporal coverage |
|----------------|----------|--------------|-------------------|
| data level | units | format / API | access rights |
| source / owner | licence | update cycle | disaggregation |

Take-home logic: concepts → indicators → metadata → linked sources → search, analytics and governance use.

Database inventory summary

We widen the concepts from a few illustrative public sources to a no-filter ecosystem spanning public, regional, farm-level and commercial platforms.



Ontology and federated catalogue architecture

This is the bridge between the conceptual paper and the database inventory: domains define what matters; metadata and source connectors make it operational.



1 Indicator ontology

7 domains, concepts and cross-domain relations built from systematic mapping

2 Metadata harmonisation

Dublin Core + ISO 19115 + local extensions for title, scale, time, unit, API and access rights

3 Federated source connectors

Representative mapped sources include FAOSTAT, Eurostat, OECD, AQUASTAT, CDS, CCKP, FishStat and ecoinvent.



4 Search, AI discovery and scenario analytics

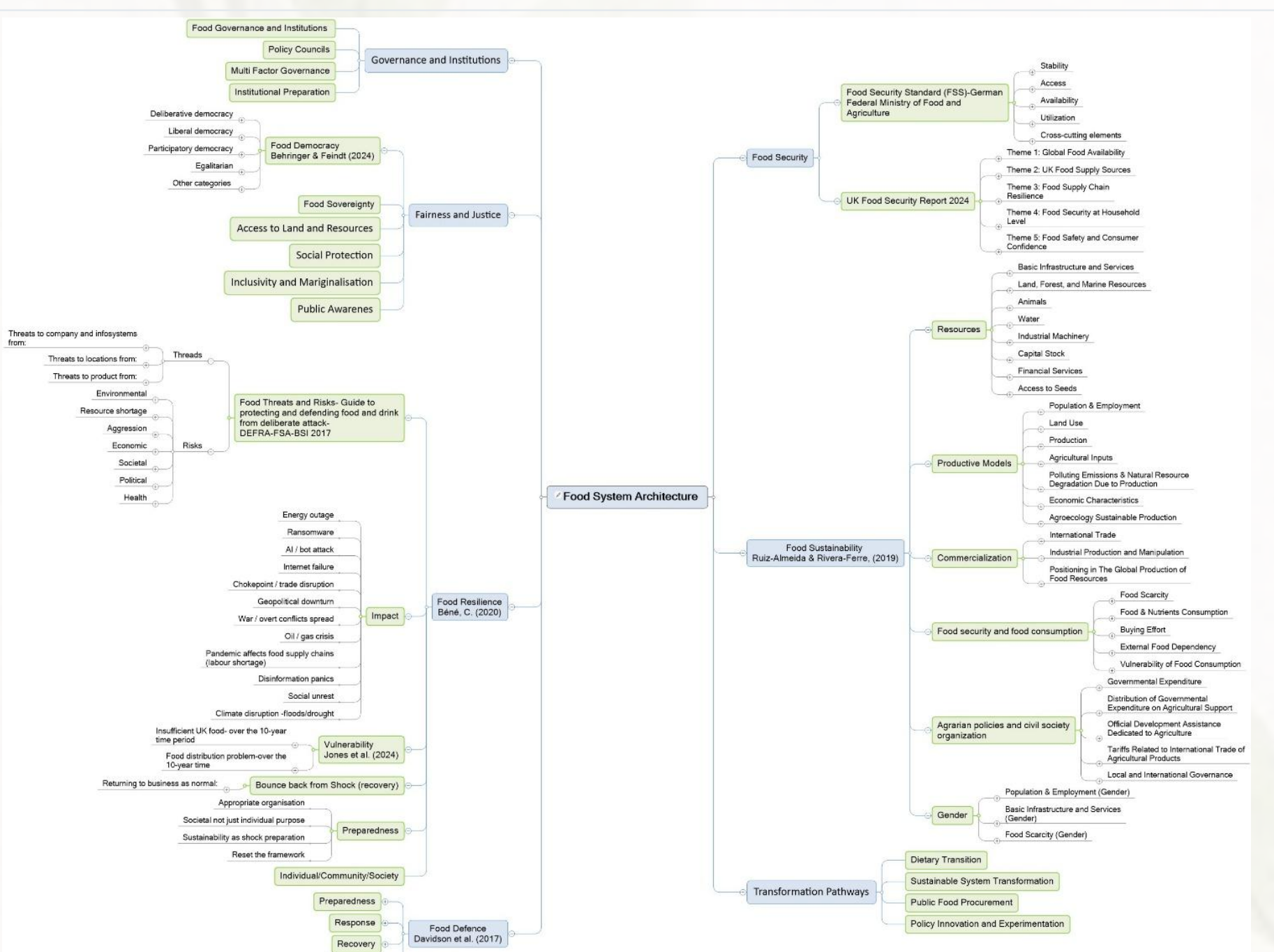
Metadata-driven retrieval, ontology-aware filtering and cross-scale comparison

5 Monitoring, governance and transformation

Supports anticipatory planning, resilience analysis and evidence-led food-system change

Detailed conceptual map

Richer branching structure behind the streamlined poster ontology.



The poster simplifies this large conceptual tree into seven operational domains while preserving cross-domain linkages in the catalogue design.

Metadata schema and example fields

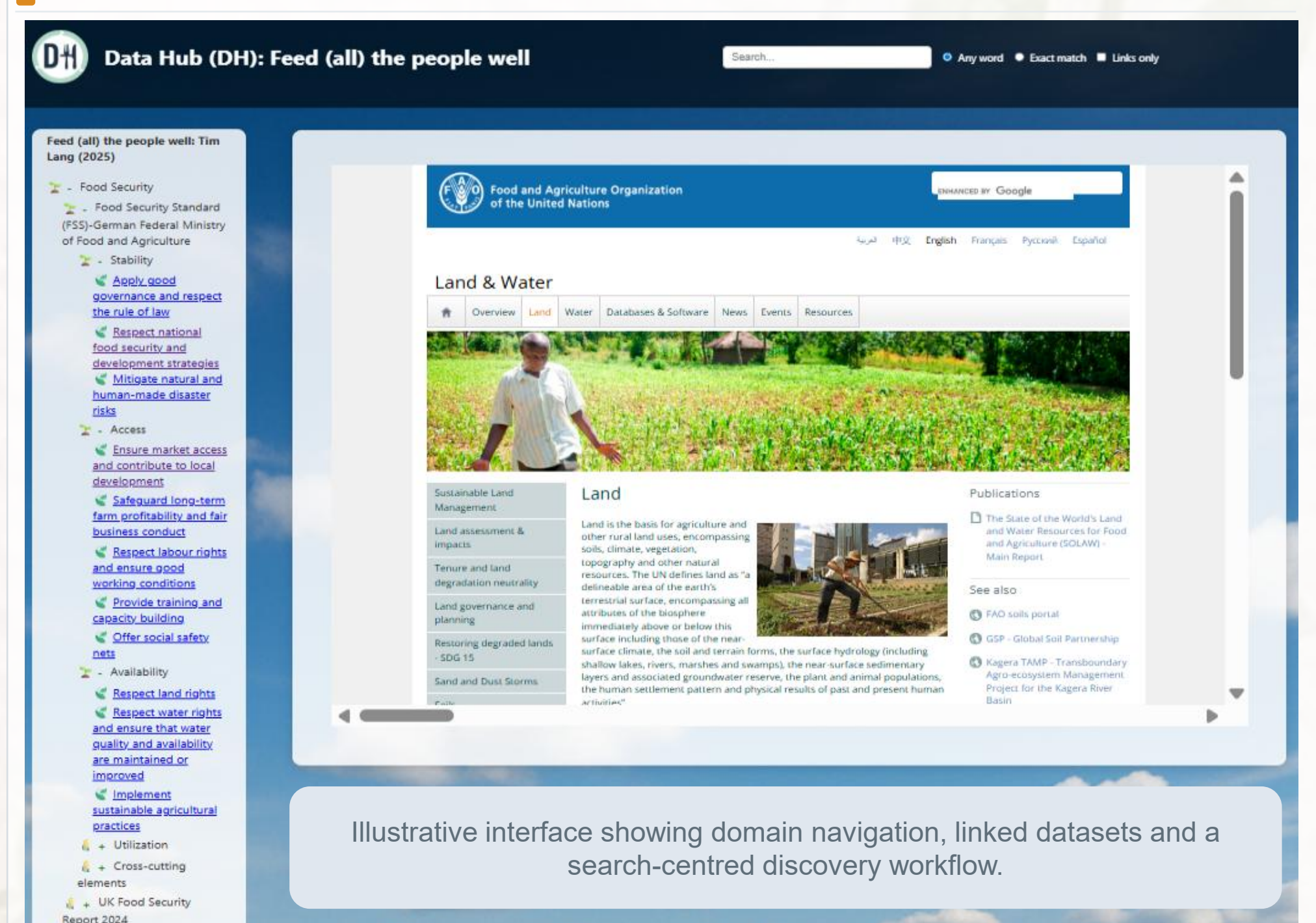
The catalogue needs standardised descriptive fields so that heterogeneous sources remain searchable, linkable and interpretable.

| Dublin Core | ISO 19115 | local extensions |
|---------------------|-----------|----------------------|
| Title | | Description |
| Domain | | Keywords |
| Geographic coverage | | Temporal coverage |
| Spatial scale | | Data level |
| Unit / measure | | Format / API |
| Access rights | | Licence |
| Source / owner | | Update cycle |
| Disaggregation | | Contact / provenance |

Metadata design requirement: link each indicator to a discoverable source, a clear scale, a unit, an update logic and an access condition. This is what turns a bibliography into infrastructure.

Prototype Data Hub interface

Selected sources shown here are examples from the companion inventory.



Illustrative alignment of representative databases to the paper's seven domains

Filled cells show where representative databases can populate or support indicators within each domain. The mapping is illustrative rather than exhaustive and sits on top of the larger 134-platform inventory.

Filled cell = representative coverage / linkable indicators

| | FAO FAOSTAT | FAO FAM | EU Eurostat | EU FSDN | OECD OECD | FAO AQUASTAT | FAO WaPOR | CDS CDS / ERA5 | WB CCKP | FAO FishStat | ISRIC SoilGrids | FAO GAEZ | USD A FoodData | LCA ecoinvent |
|---------------------------|-------------|---------|-------------|---------|-----------|--------------|-----------|----------------|---------|--------------|-----------------|----------|----------------|---------------|
| Food Security | ■ | ■ | ■ | ■ | | ■ | | | | ■ | | | | ■ |
| Food Resilience | ■ | | | | ■ | ■ | ■ | ■ | ■ | ■ | | ■ | | |
| Food Risks | ■ | | | | ■ | ■ | ■ | ■ | ■ | | | ■ | | |
| Food Sustainability | ■ | | | | | ■ | ■ | ■ | | | ■ | ■ | | ■ |
| Governance & Institutions | ■ | | ■ | ■ | ■ | | | | ■ | | | | | |
| Fairness & Justice | ■ | ■ | ■ | ■ | | | | | | | | | | |
| Transformation Pathways | ■ | | | | ■ | | | | ■ | | | ■ | ■ | ■ |

Source columns are representative examples selected from our comprehensive inventory. The point is operational linkage, not a claim that any single platform is sufficient on its own.

Implications, design requirements and next steps

This poster presents a strong narrative by showing how the proposed catalogue can be populated from a broad, real-world data ecosystem.

| 1. What the paper contributes | 2. What the inventory adds | 3. Design priorities | 4. Next step for the hub |
|--|---|---|---|
| <ul style="list-style-type: none"> From fragmented concepts to a navigable seven-domain ontology From isolated indicators to a federated, metadata-rich catalogue Local and lived-experience data are often uncatalogued. | <ul style="list-style-type: none"> Expands the source base to 134 platforms across 8 data families Includes open, regional, farm-level, hybrid and commercial sources | <ul style="list-style-type: none"> Standardise scale, time, units, provenance and access rights Keep connectors modular so public and proprietary sources can coexist | <ul style="list-style-type: none"> Strengthen justice-oriented and local/community data coverage Turn the prototype into a maintained interface for scenario planning |

Take-home message: the food data hub becomes much more persuasive when its ontology is explicitly married to a broad, real-world database ecosystem and a transparent metadata schema.



Selected references from the paper: FAO (2023); Crippa et al. (2021); Béné (2020); Béné et al. (2023); Lang et al. (2025); Schneider et al. (2025). Metadata standards highlighted: Dublin Core and ISO 19115. Figures adapted from the uploaded manuscript.