Questionnaire for Database Evaluation

1. General Information

* What is the name of the database (official title, acronym)?
* Which institution(s), organization(s), or research group(s) are responsible for maintaining and updating it?
* Who is the main point of contact (person, team, or office)?
* Is there a public website, DOI, or repository for the database?
* What is the primary purpose or intended use case of this database? (e.g., National reporting, academic research, public information)

2. Scope and Coverage

* What is the geographical/spatial coverage? (Global, continental, national, regional, local, site-based)
* What is the spatial resolution or unit? (e.g., administrative region, river basin, grid cells, point data, parcels)
* What is the temporal coverage (start year – end year)?
* What is the update frequency (real-time, hourly, daily, monthly, yearly, irregular)?
* What is the reference time for the data? (e.g., Calendar year, fiscal year, snapshot as of January 1st)
* For time series, is the data provided as annual snapshots, averages, or cumulative values?

3. Data Content

* What are the main variables/indicators included?
* What are the units of measurement used for each variable?
* Are there standardized definitions or classifications (e.g., IPCC categories, FAO, INSPIRE, ISO standards)?
* How is the data structured? (e.g., Long format vs. wide format, relational database schema)
* What unique identifiers or coding systems are used? (e.g., FAO product codes (CPC), ISO country codes, NUTS codes, GAUL codes, custom IDs)
* How are categories such as crops, livestock, or land cover classified in your database? Do they follow any international standards (e.g., FAOSTAT, IPCC, Eurostat, or other recognized classifications)?
* What are the formats of the database outputs? (e.g., CSV, NetCDF, shapefile, GeoTIFF, JSON, SQL)
* Is the data primarily raw measurements, derived indicators, or model outputs?

4. Accessibility and Use

* Is the database publicly accessible, restricted-access, or internal?
* What are the licensing conditions (open data, CC-BY, CC0, restricted, proprietary)?
* How can users access the data (API, download portal, request form, web services)?
* If accessed via an API, what are the rate limits and query parameters?
* Are metadata and documentation available (user guides, technical reports, metadata standards)?
* Are there any scripts or tools (e.g., in R, Python) provided or recommended by the data providers for accessing, processing, or disaggregating the data?

5. Quality and Validation

* What quality assurance/quality control (QA/QC) procedures are in place?
* Are uncertainties or error margins documented? If yes, how?
* Has the database undergone peer-review, validation, or benchmarking?
* How are missing, confidential, or estimated values handled and flagged in the dataset?

6. Interoperability

* Is the database linked to or harmonized with other datasets (e.g., Eurostat, Copernicus, FAO, IPCC)?
* Which models or analytical tools use this database as an input (e.g., energy system models, hydrological models, GHG inventories)?
* Does the database follow any metadata standards (e.g., INSPIRE, ISO 19115, Dublin Core)?
* Are there APIs or services enabling machine-to-machine communication?
* hat is the process for bridging this data to a different spatial or thematic resolution? (e.g., Are there official cross-walk tables or established methodologies for downscaling national data to regions or upscaling parcel data to national level?)
* How is temporal consistency handled? (e.g., Are historical time series revised when methodologies change, or is the base year updated? This affects model calibration.)

7. Sustainability

* What is the long-term sustainability plan for maintaining the database?
* What are the funding sources supporting the database?
* Is there a roadmap for future updates or expansions (e.g., new variables, improved resolution)?
* What is the versioning policy? (e.g., Are previous versions archived and accessible?)

8. Expert Assessment

* What are the main strengths of this database?
* What are the main limitations or gaps?
* What would be needed to improve its usability in integrated modeling frameworks like the Fable Calculator? (Consider: data structure, formatting, availability of unique keys, metadata completeness, provision of uncertainty data)
* From your expert perspective, what is the most appropriate use of this dataset within a scenario model? (e.g., As a primary input, as a validation source, for deriving elasticities, for setting constraints)
* Are you aware of any major assumptions or methodological choices in the dataset that could significantly influence model outcomes if used as an input? (e.g., Allocation rules, interpolation methods, exclusion of certain categories)