Risk data library project
Technical Overview

The Risk Data Schema

- Is a language for describing hazard, exposure, vulnerability and modelled loss data.
- Provides a single database that stores all four data types together, with explicit links for data generated under the same project.
- Provides a guide to technically compatible datasets, for example, hazard and vulnerability data with matching intensity measures.
- Provides a consistent framework for describing the most common types of data used in risk assessment.
- Helps anyone working with or creating risk data, particularly disaster risk analysts and researchers preparing data for risk assessment.
- Helps by making datasets significantly easier to identify, understand and work with, as they are formatted with a common schema tailored to the requirements of risk modelling.

Schema benefits

- A single consistent nomenclature for hazard, exposure, vulnerability and loss dataset simplifies the task of preparing data for risk assessments.
- A machine-readable format lays the foundation for data search and for future automation.
- Built-in machine- and human-readable metadata makes it easier to understand the data contents.
- Free and open-source for community improvement and development of a project-owned database.
- Enables storage of risk data for accessible sharing with the community.
- The schema can provide a consistent data source for analytical and visualisation applications, accessing data via OpenAPI.

Reasons for confidence in the schema as a solution

- Developed by experts in generating and using hazard, exposure, vulnerability and loss data.
- The solution delivers on findings of community consultation on making risk data easier to work with (Solving the Puzzle report).
- Community feedback has been sought and acted on throughout the project.
- Early expert reviews of the schema have confirmed the value and informed development.
- There has been a proliferation of similar data management and interoperability related projects and initiatives in DRM and insurance sectors.

The importance of adoption & application

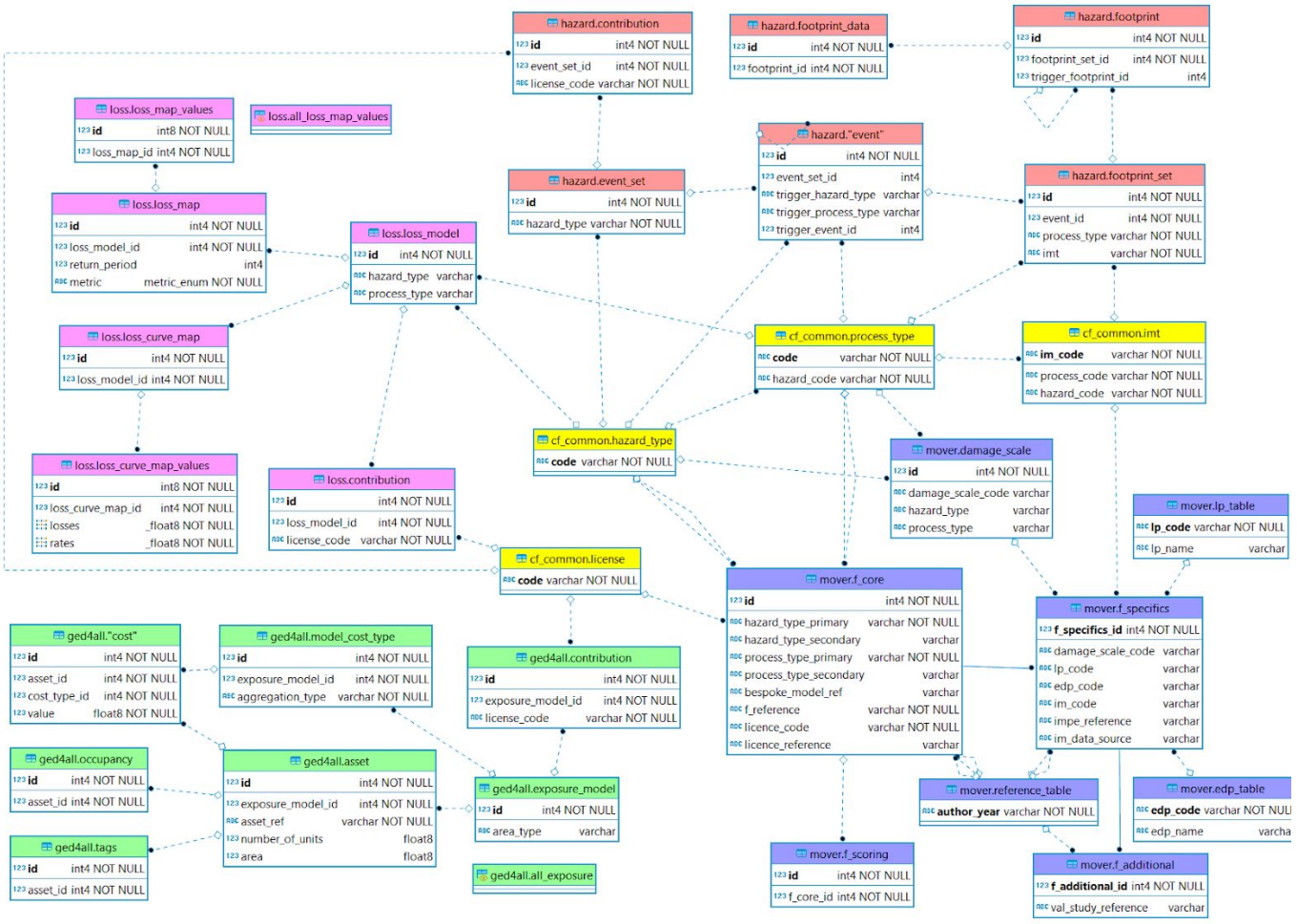
- The value of the schema multiplies with use. The more datasets using the schema, the greater its overall contribution.
- Without widespread adoption, the schema's benefits will be limited to teams and projects that use it.

Technical features

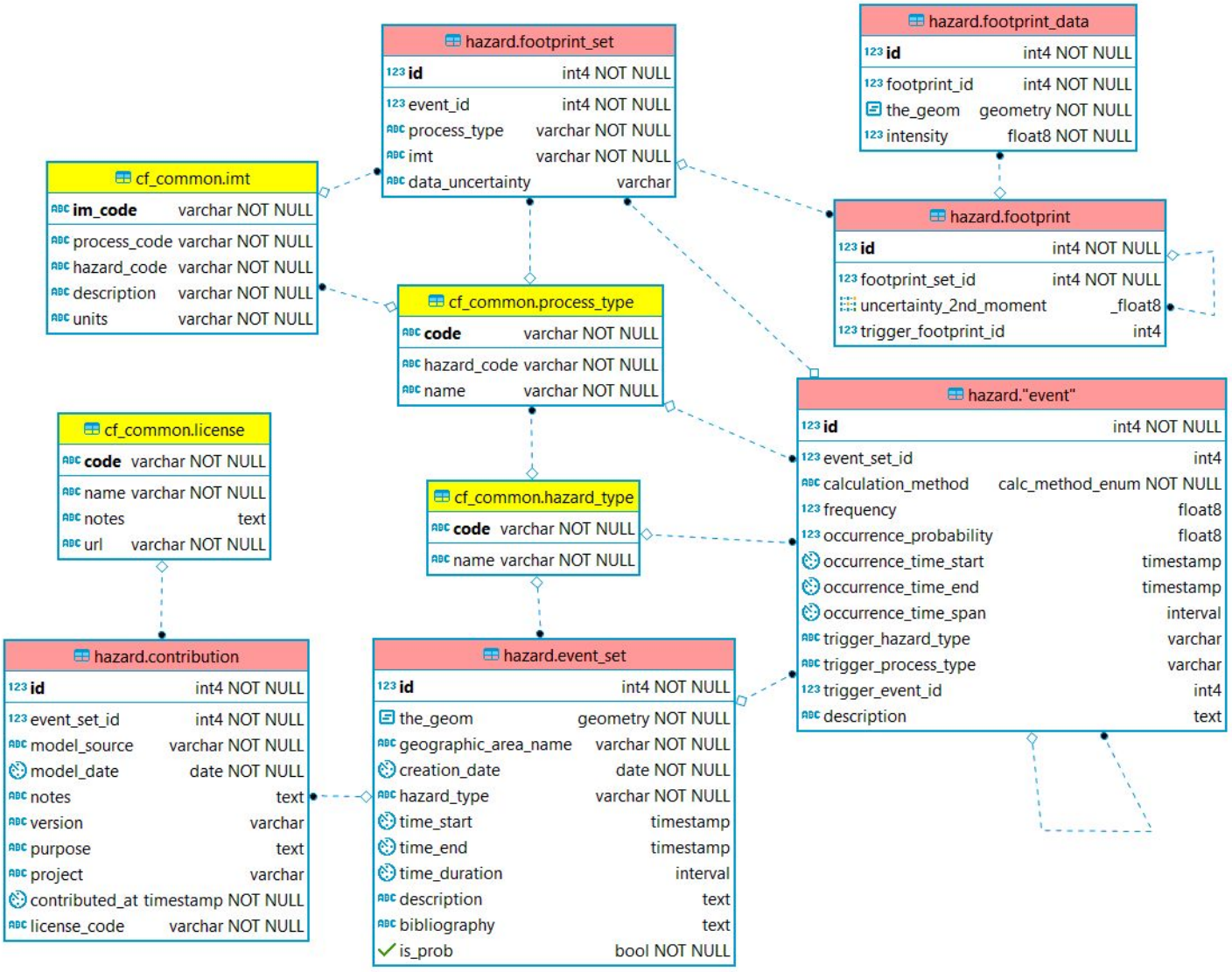
- Hazard scenario footprints and probabilistic hazard maps can be stored for multiple hazards.
- Multiple footprints can be linked to represent hazard uncertainty in an event, for multiple events in an event set.
- Hazard footprints can be linked to another 'trigger event' footprint to represent cascading hazard.
- Multiple physical asset types and social exposure can be stored in the exposure database with attributes for analysing risk from multiple hazards.
- Exposure can be stored at multiple scales from building footprint (polygon) or network segment (line) through single point exposure (a geolocated building) to aggregated/gridded data.
- The vulnerability schema includes physical fragility and vulnerability relationships (single and multi-hazard curves) and will be extended to include socioeconomic indicators and indexes.
- The vulnerability schema provides guidance on geographic relevance relationships.
- Modelled loss data is explicitly linked to the data used to generate the loss estimates; a full project dataset can be accessed together.
- All data is stored with metadata focussed on providing the information required to assess its suitably for risk modelling projects - including source, licence, development method, units, and hazard processes it relates to.

The database structure

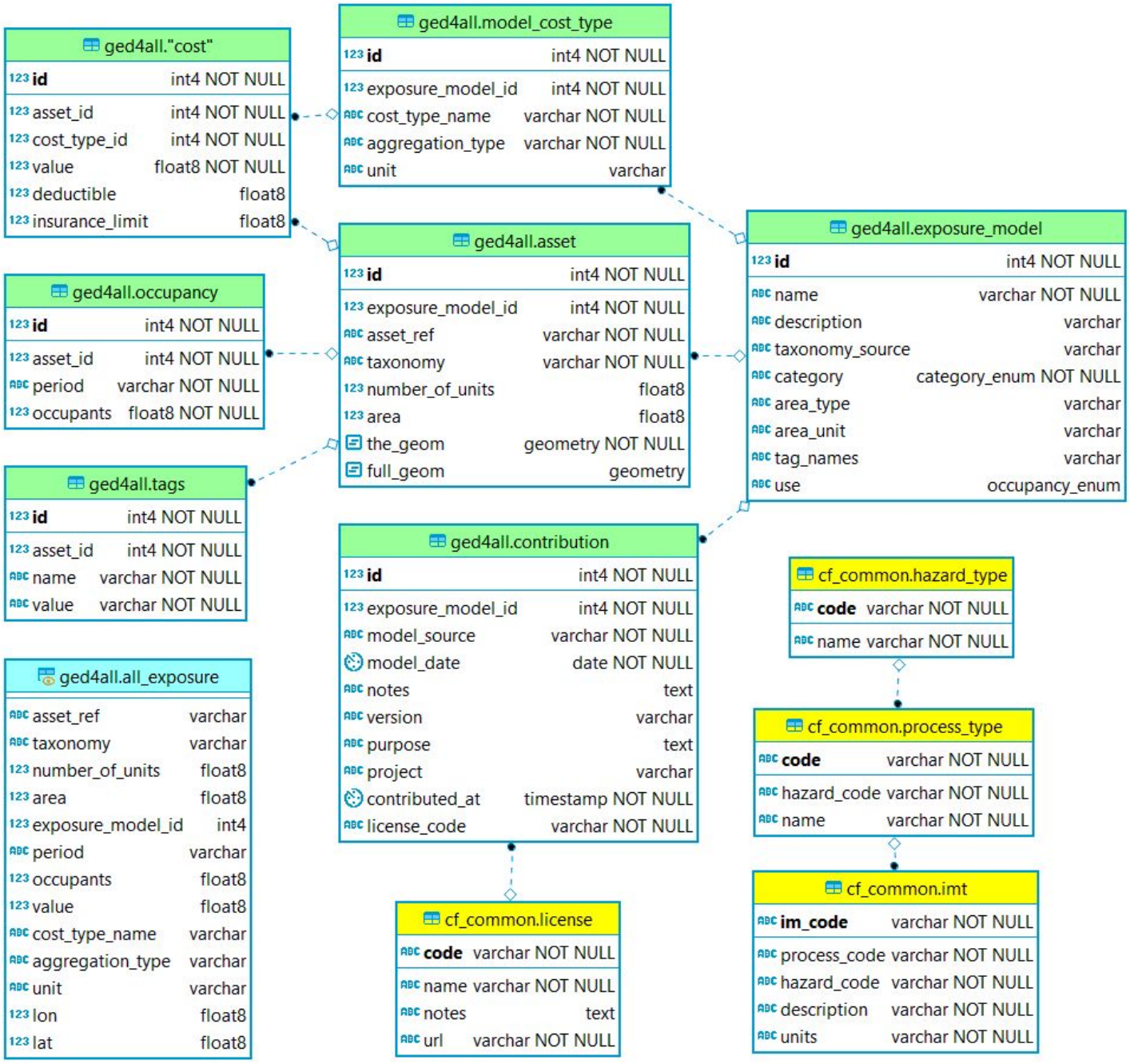
- The database is implemented in PostgreSQL, and uses PostGIS for geospatial operations.
- Data values are stored in point or with full geometry to represent polyline and polygon features.
- No raster data is stored in the database (though PostGIS supports raster and we may explore this in the future).
- Entity-relationship diagrams of the schema are shown below.



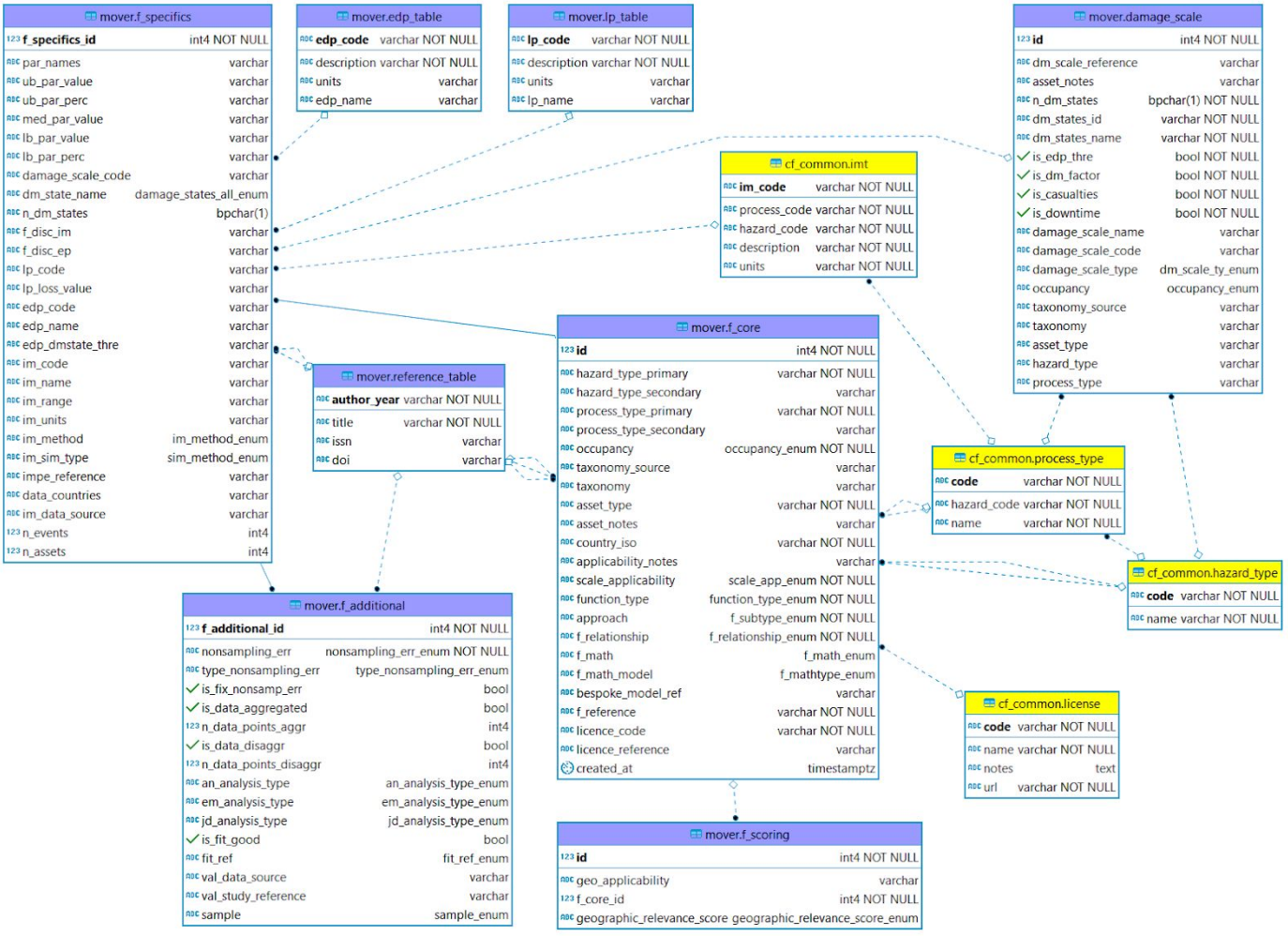
ERD (all schema): Shows the connections between tables of all schema - hazard (red), exposure (green), vulnerability (purple) and modeled loss (violet), and links to each other via common tables (yellow). Contents of each table are truncated (see subsequent ERD for more table detail)



ERD (hazard schema): hazard table contents (red) and links to common tables (yellow)



ERD (exposure schema): exposure table contents (green) and links to common tables (yellow). The schema includes a SQL view (cyan).



ERD (vulnerability schema): vulnerability table contents (purple) and links to common tables (yellow)

ERD (modeled loss schema): loss table contents (violet) and links to common tables (yellow). The schema includes a SQL view (pink).