

## 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 - overall

- Licensing: The schema requires explicit licensing of datasets. Supported open data licenses are provided in the common tables (cf\_common.license), which provides for the following license definitions:
  - [Creative Commons CCZero \(CC0\)](#)
  - [Open Data Commons Public Domain Dedication and Licence \(PDDL\)](#)
  - [Creative Commons Attribution 4.0 \(CC-BY-4.0\)](#)
  - [Open Data Commons Attribution License\(ODC-BY\)](#)
  - [Creative Commons Attribution Share-Alike 4.0 \(CC-BY-SA-4.0\)](#)
  - [Open Data Commons Open Database License \(ODbL\)](#)
- Contribution: Each schema contains 'contribution' attributes that provide information on the developer, source project, and purpose of the dataset. In future, this will enable linking of multiple datasets from the same project, enabling faster retrieval of whole-project datasets.
- 

### Technical features - hazard schema

- Hazard scenario footprints and probabilistic hazard maps can be stored for multiple hazards. Several hazards can be further defined into process types, each with a defined intensity unit:

Hazard_type.name	Process_type.name	IMT.description
Earthquake	Ground Motion	24 intensity measures, including PGA, PGV, SA (multiple period), SD, SV, PGD, IA, EMS, MMI
Flood	Fluvial Flood	Water depth and water velocity
Flood	Pluvial Flood	Water depth and water velocity
Strong Wind	Tropical cyclone	peak gust, 3 x sustained wind measures
Strong Wind	Extratropical cyclone	peak gust, 3 x sustained wind measures
Landslide	Landslide (general)	7 measures including susceptibility index, hazard index, displacement, flow depth, flow velocity, debris flow intensity index
Tsunami	Tsunami	10 measures including runup height, inundation depth, momentum flux, drag force, flow velocity
Volcanic	Ashfall	Ashfall thickness and ash loading
Coastal Flood	Storm Surge	Maximum water velocity and inundation depth
Drought	Agricultural Drought	Crop Moisture Index
Drought	Meteorological Drought	Palmer Drought Severity Index, Standard Precipitation Index

- Uncertainty in the event characteristics and hazard footprint can be represented explicitly, through the inclusion of multiple events in an event set, and many footprints per event.

- Hazard footprints can be linked to another 'trigger event' footprint to represent cascading hazards.
- Hazard footprints can be assigned an occurrence frequency, and can represent a scenario footprint or return period hazard map.
- Inclusion of an event start and end time enables the description of long-duration events.
- Export a dataset as a zipped .csv. This contains hazard intensity at long/lat for each footprint in the dataset, and the occurrence frequency associated with each footprint. We are exploring additional formats in which to present data.

#### Technical features - exposure schema

- Multiple physical asset types and social exposure can be stored in the exposure database with attributes relevant to multiple hazards
- The exposure schema provides 11 defined exposure types, for consistent classification of assets in different datasets: Residential, Commercial, Industrial, Healthcare, Education, Government, Infrastructure, Crop, Livestock, Forestry, and Mixed.
- 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.
- Exposure units can be assigned the number night-time or day-time occupants, to represent population exposure.
- Export a dataset as a zipped .csv. This contains exposure value at long/lat with exposure elements (e.g. buildings, contents), occupancy type, and number of buildings. Key metadata included are dataset name and description, project name, model source and date, purpose of the data and further notes, data license, and the taxonomy source. We are exploring additional formats in which to present data.

#### Technical features - vulnerability schema

- 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 distinguishes key information describing the function, including: function type (i.e fragility, vulnerability, damage-to-loss); development approach (empirical, analytical, judgement, hybrid, code-based); mathematical model used (including exponential, cumulative lognormal/normal); the intensity measure and asset type the function relates to; and loss parameter / engineering demand parameter values.
- The schema defines the countries for which the function was developed, to provide guidance on geographic relevance when applying a function.
- All data is stored with metadata focussed on providing the information required to assess its suitability for risk modelling projects - including source, licence, development method, units, and hazard processes it relates to.
- Export a dataset as a zipped .csv. This contains the function type, development approach and countries it is relevant to. It also communicates the hazard, process type and asset type to which

it relates, and the source of the function and the data license. We are exploring additional formats in which to present data.

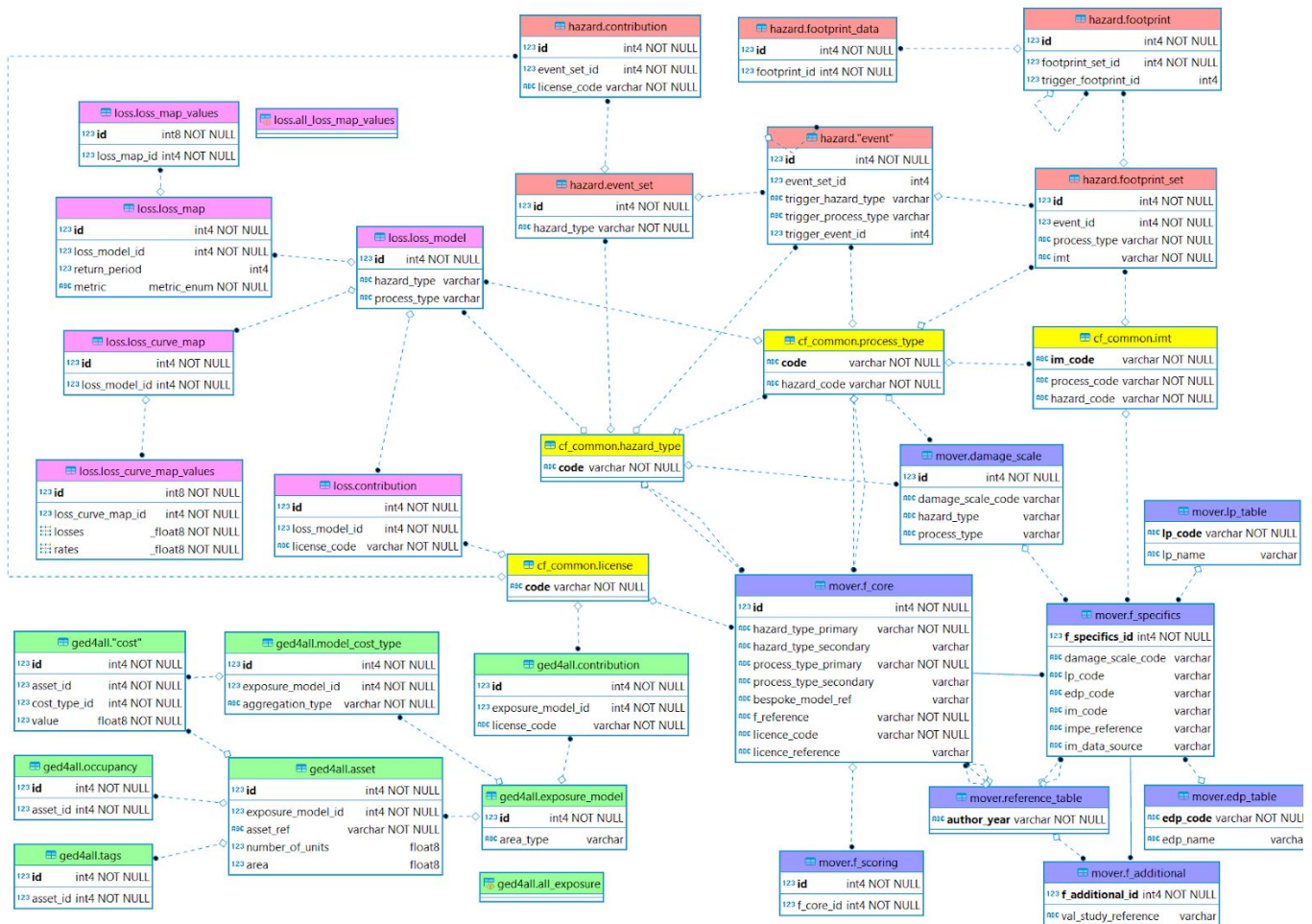
#### Technical features - modeled loss schema

- The modeled loss schema enables return period loss estimates and annual average loss to be stored - and linked to a geographic unit.
- Loss datasets can be directly linked to the hazard, exposure, and vulnerability datasets which were used to create the modeled losses. The dataset also summarises the exposure and hazard types to which the losses relate, independent of those links.
- Enumerated types provide consistent categories for referencing losses as:
  - Component: Buildings, Contents, Direct Damage to other Asset, or Business Interruption
  - Frequency: Return Period, Probability of exceedance, or Rate of exceedance
  - Loss type: Ground Up (economic) or Insured
  - Metric: Annual average loss, Annual average loss ratio, or Probable maximum loss.
- Export a dataset as a zipped .csv. This contains the loss amount and currency, the loss metric and the hazard and occupancy type it relates to. Key metadata included in the file are dataset name and description, project name, model source and date, purpose of the data, data license. We are exploring additional formats in which to present data.

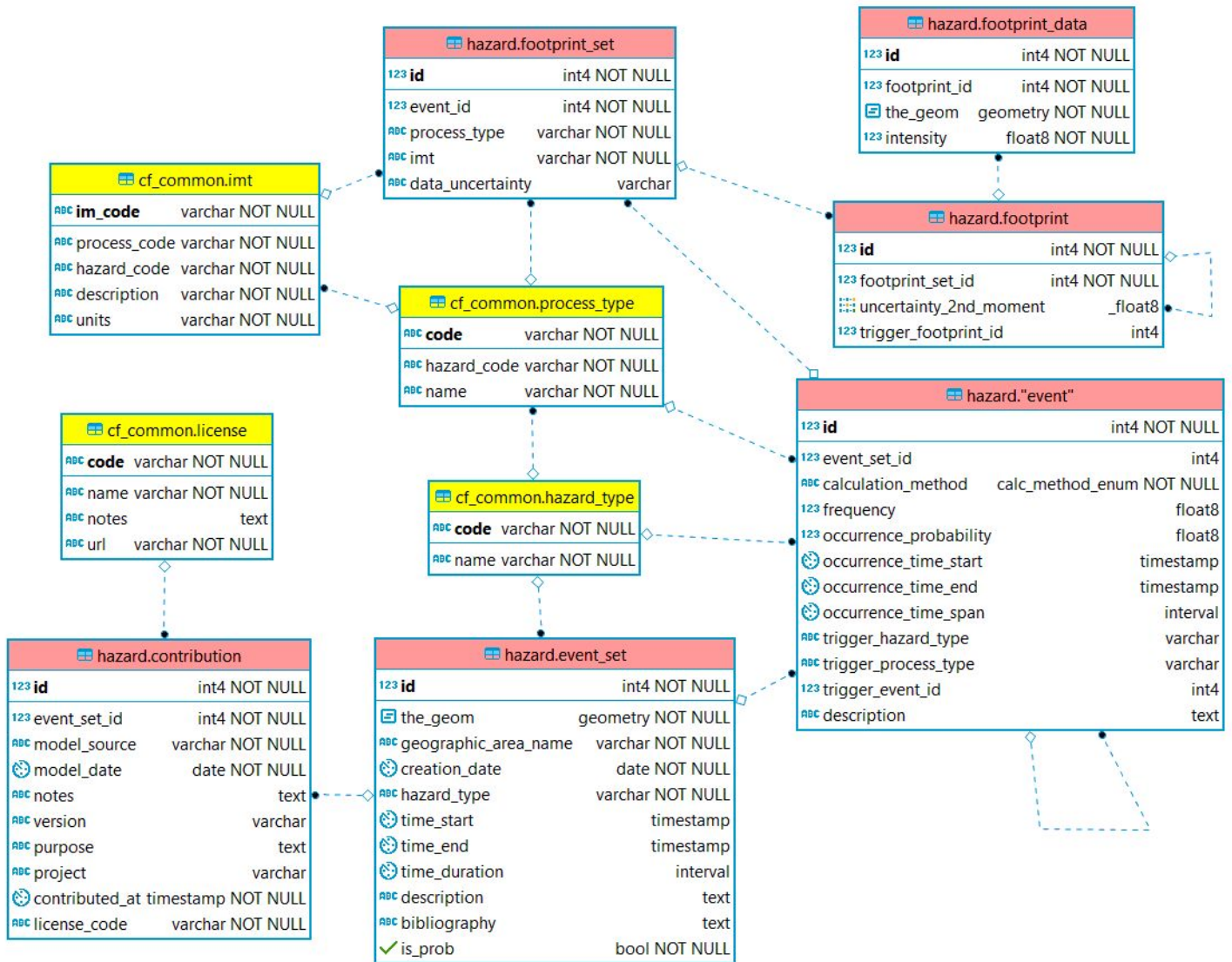
## The database structure

- 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 intend to explore adding this capability).

Entity-relationship diagrams of the schema are shown below.

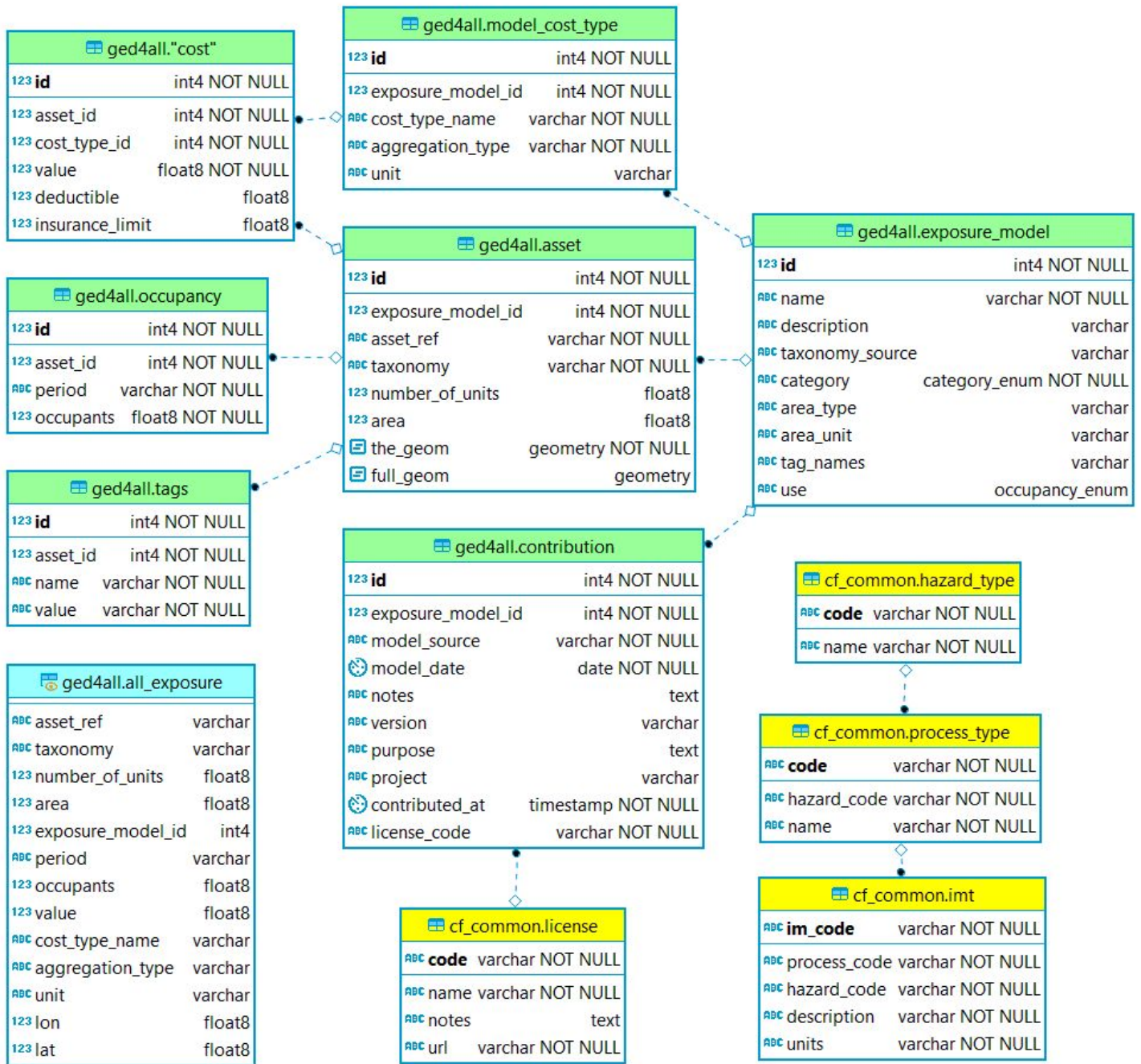


ERD (all schema): Shows the connections between tables of all schema - hazard (red), exposure (green), vulnerability (purple) and modeled loss (violet), and links to each other via common tables (yellow). Contents of each table are truncated (see subsequent ERD for more table detail)

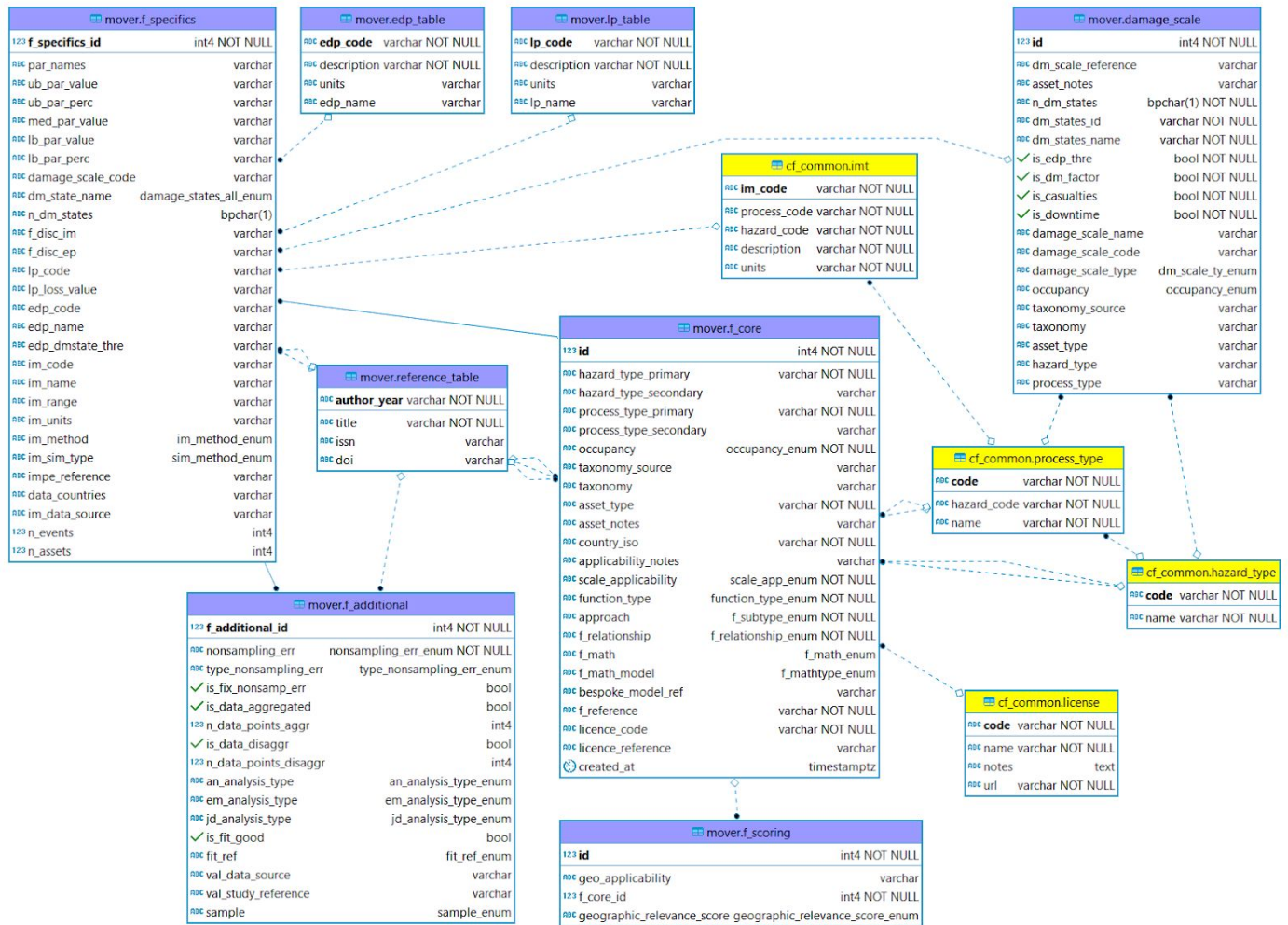


ERD (hazard schema): hazard table contents (red) and links to common tables (yellow)



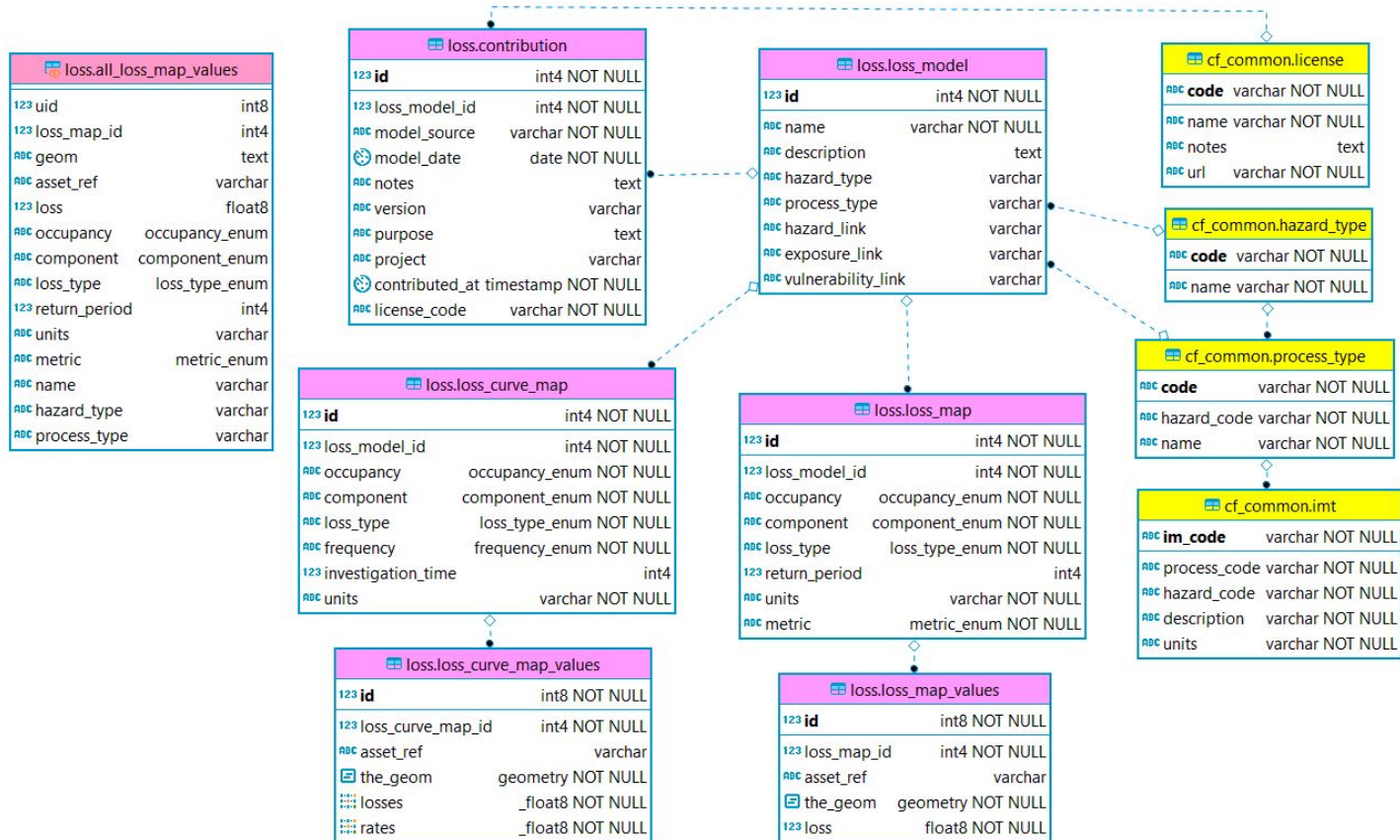


ERD (exposure schema): exposure table contents (green) and links to common tables (yellow). The schema includes a SQL view (cyan).



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).