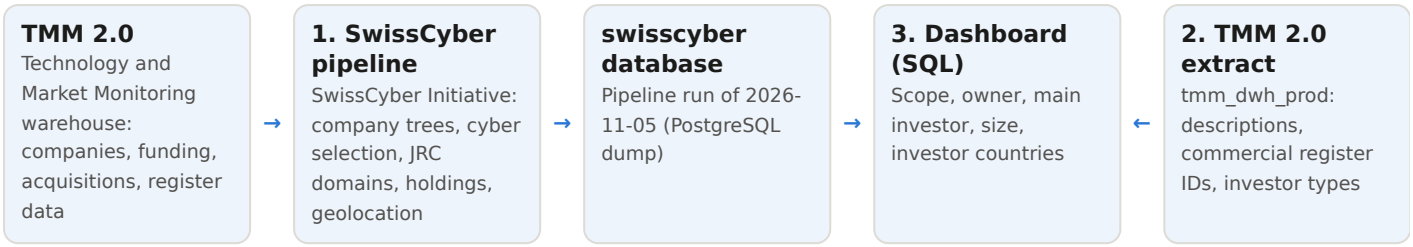
How the data was enriched

Swiss cybersecurity companies dashboard. Three contributions, from raw data to the dashboard: the SwissCyber Initiative's pipeline, a direct extract of TMM 2.0, and SQL computations for the dashboard.



1. Added by the SwissCyber Initiative, through the SwissCyber pipeline

Input: TMM 2.0 tables in the Crunchbase format, and web pages. Output: the swisscyber database.

Company trees

- Parent links from TMM 2.0 form a forest: loops ignored, first parent kept. For a company bought several times, the latest acquisition is kept.
- Each tree is merged into one **organization** named after its root entity. Merging stops at acquisitions: a bought company stays separate and points to its acquirer.

Cybersecurity selection

- A tree is kept when it has a Swiss entity and a cybersecurity entity: one of **19 Crunchbase categories** (Cyber Security, Network Security, Privacy…), or a **register purpose** deemed cybersecurity.
- Purpose classifier: TF-IDF + logistic regression (release 2026-06-11_v1), threshold 0.55, precision 0.64, recall 0.75.

JRC domains

- **Classified**: two LLM passes read the company's web pages (FineWeb, Common Crawl) and assign JRC cybersecurity domains, with high or low confidence. Matched to companies by name.
- **Inferred**: when none was found, 8 Crunchbase categories imply a domain (e.g. Privacy gives Data Security & Privacy). "Security" implies none.

Other enrichment

- **Holding**: a logistic model scores name keywords (capital, group…), categories, employees, sites, acquisitions, investors and amounts.
- **Locations**: addresses geolocated (GeoNames, Overpass, manual fixes). The headquarters is the root entity's site.
- **Funding**: rounds, amounts and named investors with their country, taken from TMM 2.0.
- **Employees**: TMM 2.0 company sizes, summed over the merged entities (e.g. 1002-5010).

2. Extracted from TMM 2.0

Production database tmm_dwh_prod, schema tmm_mart

- **Description**: `companies.description`, the first non-empty description among TMM 2.0's sources.
- **Commercial register**: the Zefix ID (CHE…) in `company_external_ids`. A company is registered when it has one.
- **Investor types**: `investors.investor_types` and whether the investor is a TMM 2.0 company.
- Linked by ID only, never by name: the pipeline keeps each TMM 2.0 company and investor key in the `uid_cb` of its sites. The headquarters' key and the investor's key are matched.

3. Computed for the dashboard

SQL queries on the swisscyber database

- **Scope**: headquarters in Switzerland and a cybersecurity signal of the company's own.
- **Owner**: HQ country of the direct acquirer stored by the pipeline, as SwissCyber does; filters and charts use it, so figures match the platform. The **ultimate owner**, top of the acquirer chain, is shown beside it (differs for 48 companies).
- **Size groups**: for the size filter and chart, each employee range is put in a standard group (1-10, 11-50…) by its lower bound. The range itself is shown as the pipeline wrote it.
- **Investor country**: a round with N investors adds 1/N of its amount to each one.
- **Main investor**: the investor with the largest total of these 1/N shares; ties keep all.

In numbers

Organizations in the database (all countries)	**48,223**
Headquarters in Switzerland	**6,494**
of which cybersecurity: dashboard and CSV	**5,648**
of which with a JRC domain (12 classified, 414 inferred)	**426**

Limits

- No revenue and no shareholdings: the owner is the direct acquirer, not the majority shareholder.
- Founding dates: 38 % are January 1st, so only the year is reliable.
- Some source matches look wrong (a law firm with USD 1.75 billion raised). Figures are not corrected.

of which in the commercial register (Zefix ID)	**4,720**
of which with a main investor (122 with ties)	**290**
Investors in the database	**18,212**

- The cybersecurity signal can be wrong: the purpose classifier has a precision of 0.64.

Where each dashboard column comes from

P = SwissCyber pipeline, T = extracted from TMM 2.0, D = computed for the dashboard

Company	**P**	Legal name of the merged organization, named after its root entity.
Website, City	**P · D**	Domain and city of the sites; the dashboard picks the headquarters' site.
Description	**T**	TMM 2.0 description of the headquarters' company, as is.
Founded	**P · D**	Founding date from the pipeline; the dashboard keeps the year only.
Employees	**P · D**	Range summed over merged entities by the pipeline, shown as is; the dashboard derives the size groups of the filter and chart.
JRC domains	**P · D**	Bitmasks from the pipeline (LLM classification, or implied by categories); the dashboard turns them into names and picks classified, else inferred.
Crunchbase categories	**P · D**	Bitmask from the pipeline, turned into names.
Direct owner	**P · D**	The direct acquirer stored by the pipeline, as on the SwissCyber platform; the dashboard adds its HQ country.
Ultimate owner	**P · D**	The dashboard follows the pipeline's acquirer chain to the top and takes its HQ country. For information only.
Money raised	**P · D**	Round amounts from the pipeline, summed per company.
Main investor	**D**	Computed from the rounds and their investors (1/N).
Status, holding	**P**	Closed flag and holding model, as is.
Commercial register	**T · D**	Yes when TMM 2.0 holds a Zefix ID for the company.
CHE ID	**T**	The Zefix ID(s) from TMM 2.0, as is.

Key figures and charts are counts and sums of these columns: foreign-owned means an owner country other than Switzerland; funds per year use the round's announcement year.

Sources: SwissCyber pipeline documentation (pipeline/README.md, purpose model card, process/tmm.py), the swisscyber database dump of 2026-11-05 and the TMM 2.0 documentation of tmm_mart. Dashboard: public/index.html; queries: tmm.sql, data.sql.