The ecosystem around data.gouv.fr

Pierlou Ramade
Data engineer
Jan 2024

What is data.gouv.fr?

-French platform for open data: all French administrations have to publish their public data there

-Open data fostering: we help administrations and other actors produce and publish data

-Improving data quality: we (help) create data schemas to improve reusability and interoperability

-Reusing data for citizens: build apps/websites to make data speak to everyone

Data quality: the use of schemas

➜A schema is a file (json) that describes the structure of other files (csv or json): fields, descriptions, types, constraints (pivot columns)...

➜Benefits of standardization:

◆integrated documentation

◆generic tools

◆easier data opening

◆easy aggregation

◆consistency

◆interoperability

◆reusability

Data quality: the use of schemas

The ecosystem around schemas:

➜schema.data.gouv.fr: the catalog of public schemas, which is enriched whenever a new schema emerges, after rounds of concertation and drafting (all our schemas are hosted on Github)

➜publier.etalab.studio: website to assist data producers in publishing data according to a selected schema (many producers are not data experts)

➜validata.fr: verify whether a file conforms to a selected schema. The underlying API is used in many of our production scripts to consolidate data and add metadata to data.gouv.fr resources

Data quality: the use of schemas

A few examples:

➔IRVE (electric vehicle charging station): to get a subsidy, entities that have charging stations installed have to declare them on data.gouv.fr, following the dedicated schema (the link to the schema is in the law); more than 1100 files are consolidated into a single file recreated everyday with potential newly uploaded ones

➔Repairability index: manufacturers will have to publish data about how easy their products are to repair, with more than 150 precise fields including grades in all tests and availability of the replacement parts
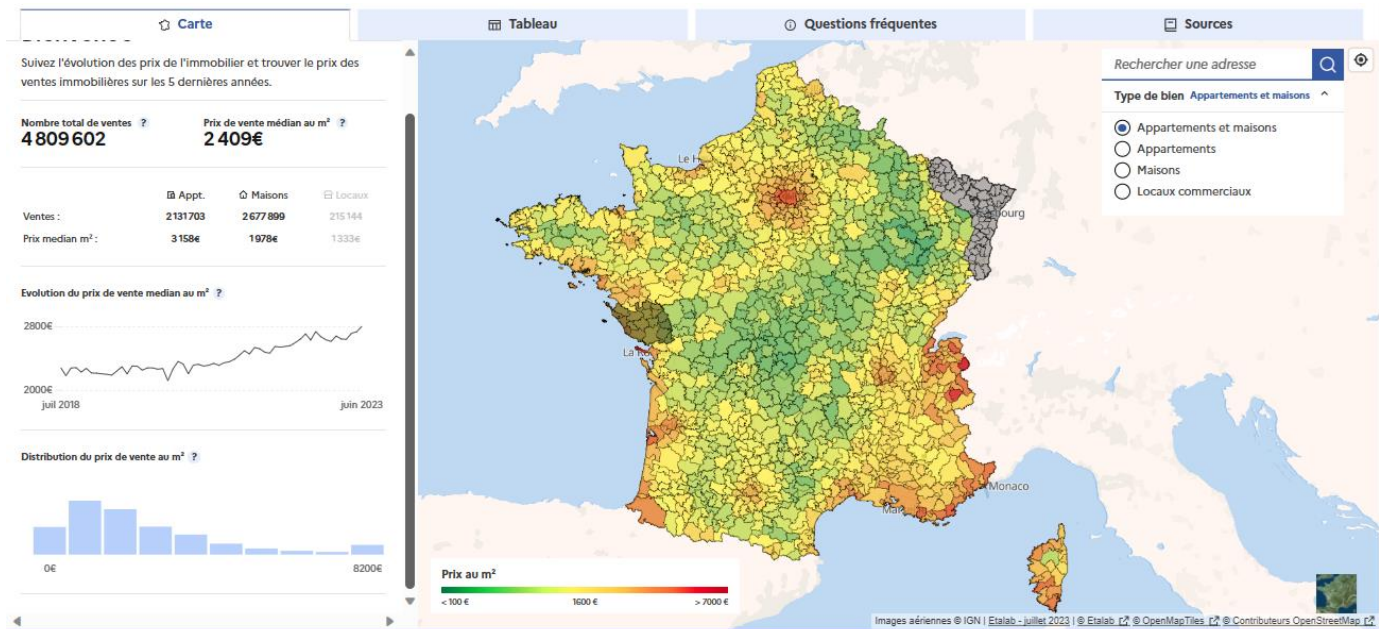
Data reuse: make data graspable to everyone

On data.gouv.fr, data is published in the rawest way possible to prevent biases and allow for the widest range of reuses.

In specific cases, when we consider that a dataset is relevant and important, we can build an exploration of the data available to all.

Data reuse: make data graspable to everyone

DVF : real estate prices across the past 5 years (explore.data.gouv.fr/immobilier)

Data reuse: make data graspable to everyone

National gas prices (explore.data.gouv.fr/prix-carburants)