External Data Sources in Analytics

Thank you for downloading this guide!

Where to find open and commercial data sources relevant to your business? How do you bring external data into existing analytics and processes? And what are the challenges involved in utilizing external data?

The answers to these questions can be found in this guide. The guide also provides links to various types of free open data sources.

We wish you a pleasant reading!

Tip: You can use the table of contents on the next page to easily navigate to the section you want.

Content

What is external data?

External data holds one of the largest unexplored gold mines in modern business intelligence. External data can be defined as all data that is produced and acquired from outside your own organization. External data is also:

✓ Data you don't have readily at hand
✓ Data stored outside the current databases
✓ Data you cannot use within minutes from your everyday sources
✓ Data you download
✓ Data not currently in use

External data sources can be divided into several different categories. The table on the next page lists examples of different types of external data. We have also gathered links in the table, where you will find examples of different types of open data sources.

In addition to open data, a lot of data is available from paid commercial sources. For example, company data can often be found in paid sources and services (such as company databases Finder or Vainu.io).

NATURAL ENVIRONMENT

BUILT ENVIRONMENT

ADMINISTRATION

ENTITIES

TRAFFIC / FLOWS

STATISTICS

IMAGERY

SENSOR DATA

Examples of external data sources

Category	Subcategory and example data sources	Example use cases
Natural Environment	**Weather** **Land use** **Elevation** **Forestry & Natural Resources**	Flood Risk Modelling Storm Risk Modelling Run-off analysis Predictive maintenance
Built environment	**Real estates** **Buildings** **Networks** – roads, rails, utilities	Market analysis, Property valuation Facility allocation, Utility service Routing, Network Analysis
Administration	**Spatial boundaries** (postal codes) **Funding usage**	Data aggregation, facility allocation Public offer targeting
Entities	**People** **Companies** Vehicles	Highly targeted marketing Market modelling & segmentation Property valuation
Traffic / flows	People **Roads & Rails** **Shipping** **Aviation**	Facility & Service planning Infrastructure development and maintenance Traffic estimations Outdoor marketing
Statistics	**Demographics** **Economic key figures**	Large scale modelling Enrichment of existing registries
Imagery	**Remote sensing (satellites etc.)** Photography / Video	Feature Recognition Algorithms, Ground movement analysis People Flow Analysis
Sensor data	Vehicles IoT Devices	Real time process control Security & surveillance

How to map out external data sources?

Thanks to the Open Data movement, which has awakened in recent years and has become more and more powerful, materials can be found more easily and often for free.

As you may have noticed from the table on the previous page, external data can be divided into several different categories and there is a great deal of different data available.

What is there then? There is the natural environment around us with its forests, waterways, altitude differences and weather conditions and on the other hand the built environment we have created with its properties, buildings and roads.

Streams flow in the above-mentioned environments: water, traffic, and telecommunications. Streams, on the other hand, carry several mobile and hard-working actors: people, companies and vehicles. All this data is available thanks to comprehensive imagery-based remote sensing data from satellites and airplanes.

At the ground level, on the other hand, there is device-based information that continuously measures all this in machine form, so-called sensor data. In addition, information is available in the form of materials published by the governing bodies. Finally, there is binding, pre-chewed and so-called traditional, structural statistical data of the whole entity.

How do I find out what kind of data is right for me? When mapping out data sources, it is worth considering:

- ✓ How achievable is the data I crave? Is it behind a payment wall or registration, or is it completely free?
- ✓ At what resolution, temporal or regional, do I need information?
- ✓ Do I need to know how many people walk through the entrance of a particular store on an hourly basis, or would an annually updated zip code area-level statistics of the population be sufficient?

How is external data combined with other data?

How do you bring external data into your existing processes?

The most common challenge after finding suitable data sources is probably linking these data to existing data. When processing several data sources, it is not always possible to have a key that connects different entities (for example, a name or a postal code), and the more data sources are used, the more complex the work steps become. What to do?

Almost all data is proportional to location in one way or another. Mobile data provides indicative information on people's movements, buildings and the actors located in them can be placed on a map by their address, and statistics are often created based on some regional division. Thus, location information can function as a powerful tool for combining and relating different data sources.

For example, the weather conditions in a particular environment can be imported to support route planning based on the coordinates of the nearest weather station. Consumer flows moving close to business area X can be mapped by comparing address information and coordinate pairs from mobile data.

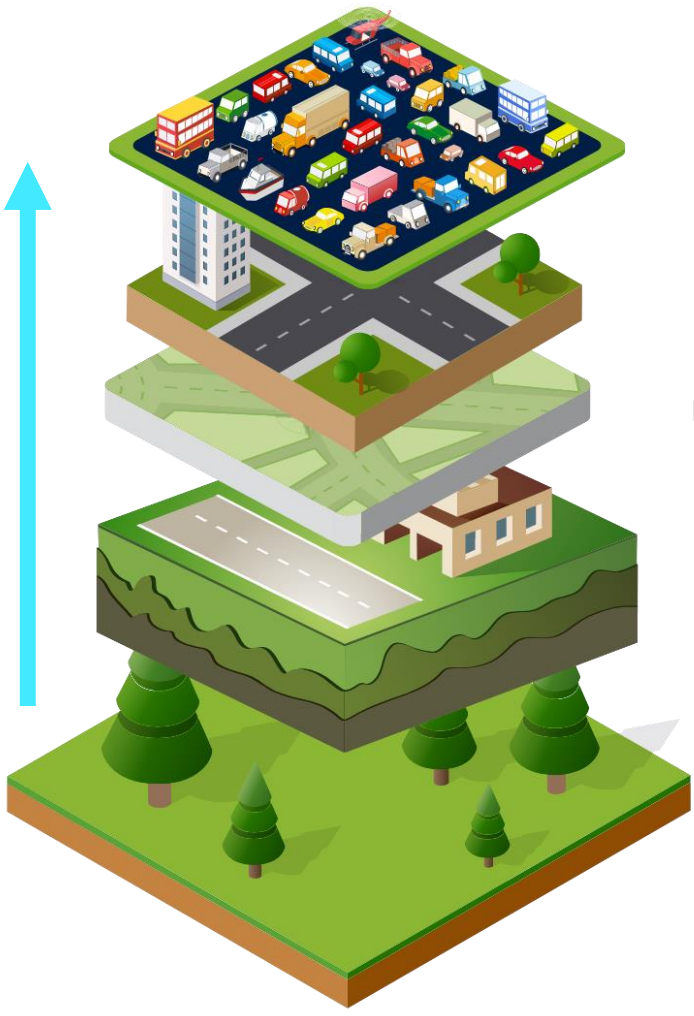
Almost all data has a location component

- ✓ People can be located by their phone location
- ✓ Real estate / building locates by their addresses
- ✓ Photos have a location in their metadata
- ✓ Vehicles location can be determined by their coordinates at several points in time

Location Intelligence: How to combine data?

Location intelligence (also sometimes called location analytics) acts as a link between data silos. If you want to combine two separate data sets or enrich existing data with external data, the solution is location data. Location data glues a company's own data into external data. By comparing locations, two separate data sources can also be combined without common variables.

Current data and the layers of external data have spatial dimensions. Utilizing location intelligence methods, existing information can be enriched with layers of external data.



How do I get started?

When utilizing external data, it is a good idea to start by defining a use case and identifying data gaps.

Define the use case and need

✓ What is the question that needs an answer?
✓ Which operating models currently seem ineffective?
✓ Ideation and innovation: would model A work better if I knew more about B?

It is worth taking a moment mapping the sources of external data and following even the tiniest bursts of ideas. From the combination of the ideas and existing data sources, the need and purpose of the data may become clear very fast. Thus, filtering the right data sources according to your own capabilities and usage is the final step!

Determine your data needs

✓ Determine what kind of data needs you have: what kind of internal data exists, what kind of external data do I need in addition?
✓ What type of data do I need?

Select the right datasets

✓ The best result is usually achieved by using multiple sources, not just one source.
✓ What tools do I have, what can I do myself?
✓ What is supposed to be done with the data?

Some External help?

However, this is not always easy. You don't have to and shouldn't do everything yourself. Relying on the help of experts can ease and speed up your own work process. Contact us for a free consultation here and find out how your organization could benefit from external data!

Use cases for external data

EXAMPLE 1

Assessment of forest damages after a storm

Data sources:
- ✓ Commercial satellite data (interpretation by machine learning)
- ✓ Open forest data (see e.g. **Luke**)
- ✓ Open weather data
- ✓ Open data on residential areas
- ✓ Open data on real estate ownerships

EXAMPLE 2

Near-real-time analysis of customer flows within and around a store

Data sources:
- ✓ Using video analytics - data collected from store entrances and analyzed using machine learning methods: the data can be used to identify customers' gender, age, and emotional state. Individuals cannot be identified from such data, which complies with GDPR regulations.

EXAMPLE 3

Planning of place of business - Real estate / Facility design and location decisions

Data sources:
- ✓ Internal data on visitors / customers of similar sites
- ✓ Commercial data on past and present traffic flows near the target locations
- ✓ Open demographic data and population data near the target locations

What are the challenges with external data?

External data offers tremendous opportunities to enrich your own data sources and make analytics more effective. Why is there still little use of external data?

Often, the use of external data requires location intelligence methods - and expertise in them. Even if the value of external data is recognized in the organization, it is difficult to find top experts. Location intelligence is a separate discipline and entity that requires, among other things, expertise in remote sensing, 3D modeling, geoinformatics and geography as well as methodological and technical knowledge.

The availability and quality of data varies

In Finland, the availability and quality of external data for both open and commercial data is relatively good. At the international level, however, the data are fragmented and incomplete or difficult to access. Exceptions are remote sensing data and traditional statistical data (UN, World Bank), which are also available internationally.

Contact us

CHRIS WINQUIST

Chief Commercial Officer
+358 50 387 8670
christoffer.winquist@advian.fi

Are you interested in spicing up your analytics with external data? Let us help! Ask more and book an appointment from Chris' calendar here:

Our services

Consulting, innovation and design
Mapping and searching suitable data sources
Combining & enriching data sources
Automations of external data processing
Advanced analytics models based on external data
Analyzes & automations

www.advian.fi/en