



OPEN DATA FOR SCIENTIFIC INVESTIGATION BY STUDENTS

Using open data to extend scientific investigations



As part of the SteamCity project, the use of open data represents an opportunity to extend scientific investigations conducted in the classroom into genuine citizen science projects. This practical sheet aims to support teachers in integrating the open data approach into their teaching practices, by showing how these resources allow both to deepen investigative approaches and to enhance students' work through sharing and collaboration between peers.

Understanding Open Data in the School Context

Open data refers to all digital data that is freely accessible, reusable without technical or legal restrictions, and shareable by all.

Originating from historical scientific practices, the open data movement became structured and democratized in the 2000s with the arrival of the Internet and gradually spread to other areas such as public administration and education.

Today, to be considered “open,” a dataset must have terms of use that respect the principles established by the Open Knowledge Foundation, such as availability and access, reuse and redistribution, and universal participation. These terms of access are embodied in licenses, which are legal documents representing the contract that binds the user to the producer of a piece of data.

In the context of the SteamCity project, open data, in addition to representing an external source of knowledge, transforms the way students approach scientific investigation. They are no longer simply consumers of information but also become producers of data that enrich a common heritage of knowledge about their territory.

This approach is part of SteamCity's ambition to develop an experimental approach for learning territories. Students can contextualize their observations within a larger data set, compare their results with those of other classes or other schools, and actively contribute to understanding the environmental and societal issues in their immediate environment. This approach is in line with UNESCO's recommendations on open science education and the Sustainable Development Goals, which advocate for more inclusive and participatory science.

UNESCO Recommendation on Open Science

This document aims to provide an international framework for open science policies and practices. It sets out a definition, as well as internationally shared values, principles and standards, and proposes a set of actions to promote the fair and equitable implementation of open science for all.

https://unesdoc.unesco.org/ark:/48223/pf0000379949_fre



DEEPENING THE INVESTIGATIVE APPROACH USING OPEN DATA

Initial exploration: building on what already exists

Before beginning the experiment, exploring existing data is a fundamental step in the scientific process. Open data portals, whether local, national, or European, provide information that allows students to contextualize their problem and better construct their research questions. Local authorities regularly publish data on air quality, transportation, energy, and biodiversity.

This exploration phase allows students to formulate more precise hypotheses and identify the missing data they will need to collect. It develops their critical thinking by confronting them with the diversity of sources and the need to evaluate the reliability and relevance of the available information.

The available data must be adapted to each situation and each territory. Generally speaking, the student can be encouraged to start with the most local data sources and gradually move up to broader data.

Here are some multi-level examples of open data platforms that can be given to learners:

Local and regional platforms	National platforms
<ul style="list-style-type: none">La Rochelle Agglomeration: https://opendata.agglo-larochelle.fr/accueilGreater Lyon: https://data.grandlyon.comToulouse Metropole: https://data.toulouse-metropole.frParis : https://opendata.paris.frÎle-de-France region: https://data.iledefrance.frSouth Region: https://datasud.fr	<ul style="list-style-type: none">National Open Data Portal: https://www.data.gouv.frGeorisks: https://www.georisques.gouv.frGéoportail IGN : https://www.geoportail.gouv.frINSEE: https://www.insee.frATMO France (air quality): https://www.atmo-france.orgADES (groundwater): https://ades.eaufrance.frOpen Data Education: https://data.education.gouv.fr
European platforms	International platforms
<ul style="list-style-type: none">European Data Portal: https://data.europa.euEuropean Environment Agency: https://www.eea.europa.eu/data-and-mapsAir quality in Europe: https://www.eea.europa.eu/themes/air/air-qualityWater Information System (WISE): https://water.europa.euNoise Monitoring Service: https://noise.eea.europa.euCopernicus Climate Data Store : https://climate.copernicus.eu/climate-data-storeCopernicus Atmosphere : https://atmosphere.copernicus.euCopernicus Land Monitoring : https://land.copernicus.eu	<ul style="list-style-type: none">OpenStreetMap: https://www.openstreetmap.orgWorld Bank Open Data : https://data.worldbank.orgOur World in Data : https://ourworldindata.orgGlobal Forest Watch : https://www.globalforestwatch.org

Enrichment during experimentation

During the data collection phase, access to open data enriches students' observations by contextualizing their measurements within a broader framework and revealing correlations. This perspective transforms a simple data collection into a truly multidimensional scientific investigation.

Here are several examples of protocols that could be enriched using open data:

- On the study of air quality around the school, a project particularly relevant to working on well-being in its territory. The measurements of fine particles taken by students with sensors can be immediately compared with hourly data from the stations of the nearest air quality agencies, making it possible to validate the proper functioning of the instruments and to identify possible hyperlocal phenomena. Cross-referencing with road count data available on municipal portals often reveals correlations between pollution peaks and peak traffic times. The integration of meteorological data, in particular wind direction and strength from national weather stations, makes it possible to understand the dispersion of pollutants and explain certain apparently random variations.
- An urban biodiversity study project illustrates this synergy between collected data and open data in a different way. While students inventory pollinator-friendly spaces in different green spaces in their municipality, they can enrich their observations with data from the CORINE Land Cover program (<https://land.copernicus.eu/en/products/corine-land-cover>), which characterizes land use on a European scale. This comparison helps understand how habitat fragmentation influences the presence of certain species. Historical urbanization data, often available through digitized cadastral archives, reveal landscape evolution and explain the current distribution of biodiversity.
- ***The analysis of urban heat islands is another particularly demonstrative example of this integrated approach. Temperature readings taken by students in different neighborhoods take on a new dimension when cross-referenced with thermal satellite data from the Copernicus program, which offers a broad view of the phenomenon. Urban databases on building density, building height, and vegetation surface area, generally available as open data, make it possible to establish precise correlations between urban morphology and temperature. Aggregated energy consumption data by neighborhood, published by many cities as part of their climate plans, reveal the link between heat islands and energy demand for air conditioning.***
- A project on sustainable mobility offers further enrichment opportunities. Cyclist counts carried out by students on strategic routes can be compared with data from permanent automatic counters installed by many cities, the results of which are published in real time. Road accident data, available on data.gouv.fr with precise geolocation, helps identify safety black spots and understand resistance to cycling. Public transport journey times, accessible via transport operators' APIs, provide objective comparisons between different modes of travel. Air quality data along roadways helps quantify the exposure of cyclists and pedestrians to pollution.

This approach develops in students a systemic understanding of the phenomena studied. They discover that each environmental and/or territorial issue is part of a complex network of interactions where natural, anthropogenic, historical and socio-economic factors intertwine. The ability to mobilize and cross-reference different data sources then becomes an essential skill for understanding this complexity. Students thus learn that scientific investigation is not limited to direct observation but is enriched by the ability to navigate the open data ecosystem to build a systemic understanding of the phenomena studied.

In-depth analysis of the results

The analysis phase represents a privileged moment to develop students' critical thinking and their understanding of the scientific approach. The exploitation of open data offers an opportunity here to address the fundamental question of data quality and the validity of scientific results.

The results obtained can be validated by comparison with reference data, which allows students to question their methodology and refine their understanding of the scientific protocol. Broadening the perspective by cross-referencing different data sources reveals unexpected correlations and stimulates reflection on causalities. The territorial complexity of the phenomena studied becomes accessible thanks to the multiplicity of possible angles of analysis.

During this stage, it is important for the teacher to question the investigative work so that the student understands its scope but also grasps its biases and limitations. To do this, the teacher can address the following aspects:

- Comparison with reference data: Not as a value judgment, but as an opportunity to analyze the sources of variation (instrument precision, measurement conditions, calibration protocols)
- The importance of metadata: Explain how comprehensive documentation of data production conditions is essential to ensure their reproducibility and comparability.
- Concrete statistical learning: Present the use of the concepts of means, standard deviations and confidence intervals on experimental data to understand the uncertainty inherent in any measurement.
- Spatial and temporal representativeness: Explain how a single measurement cannot claim to represent an entire territory, and emphasize the importance of measurement networks.
- Interoperability and standardization: Present this aspect of data not as a constraint but as a condition for scientific collaboration between peers.
- Promoting scientific transparency: Emphasize that recognizing the limitations of one's data is a sign of scientific maturity, not an admission of weakness.

The teacher plays a key role here by guiding this critical analysis without discouraging students. Every investigation, however modest, contributes to the edifice of knowledge as long as it respects these principles of quality and transparency. This perspective encourages students to continually improve their methodology, refine their protocols, and view each experiment as a step in a process of seeking knowledge rather than as an end in itself.



MAKING INVESTIGATIVE WORK VISIBLE AND SHAREABLE

Structuring data for sharing

Valorizing investigative work begins with rigorous structuring of the collected data. This requirement for rigor is in itself a fundamental learning experience in the scientific approach. The data must be organized according to a standardized format that systematically includes the date, time, precise location of the measurement, the measured value with its unit, the type of sensor used, the identification of the class, and any relevant comments on the measurement conditions.

Using open formats like CSV or JSON ensures data interoperability and reuse by other stakeholders. Documenting the methodology used, through a detailed explanatory sheet, ensures the reproducibility of the experiment and understanding of the collection context. Assigning a Creative Commons license explains the conditions of reuse and raises students' awareness of intellectual property issues in the scientific field.

Visualization and dissemination tools

Transforming raw data into understandable visualizations is a key step in scientific communication. Tools like Datawrapper (<https://www.datawrapper.de/>) make it easy to create interactive graphs and maps tailored to students' learning levels. For more ambitious projects, uMap (<https://umap.openstreetmap.fr/fr/>) offers the ability to create custom maps based on OpenStreetMap, while Grafana (<https://grafana.com/>) allows you to build dynamic dashboards for real-time measurement monitoring.

The choice of tool depends on the level of the students and the complexity of the project, but the objective remains constant: to make the data accessible and understandable to different audiences, from peers to local decision-makers.

Creating value through collaboration

Data sharing opens up particularly enriching opportunities for collaboration. Synchronous comparison, where multiple classes simultaneously take measurements in different institutions, neighborhoods, or cities, allows for real-time mapping of territorial variations in a phenomenon. This approach fosters a sense of belonging to a broader scientific community and gives meaning to efforts to standardize protocols.

The international dimension of the project takes on its full meaning here. The exchange of data with European partners allows for a comparative analysis of the impact of different environmental policies and an understanding of how geographical, cultural, and regulatory contexts influence territorial issues.

Contributing to the citizen data heritage represents the culmination of this approach. Publishing the data on local open data portals allows it to be used by other local stakeholders: environmental associations, university researchers, and municipal services. Students thus become aware of their ability to actively contribute to collective knowledge and influence local decisions at their level.



PRACTICAL IMPLEMENTATION: THE EXAMPLE OF SOUND MAPPING

To illustrate this approach concretely, consider a sound mapping project of the school's neighborhood. This project takes place in four distinct phases over approximately six sessions.

The preparation phase begins by exploring existing noise maps, generally available on municipal open data portals in accordance with the European Directive on Environmental Noise. This analysis helps identify areas of interest and relevant points of comparison. Students then define a rigorous measurement protocol, specifying the timing, duration, and conditions of measurement.

The data collection phase involves students in the field using sound level meters or calibrated mobile applications. Structured data recording is accompanied by photographic capture and qualitative observations of the sound environment. This phase develops observation skills and experimental rigor.

Data processing involves importing the data into a spreadsheet for statistical analysis, creating an interactive map to visualize noise levels, and comparing it with current regulatory standards. This phase leverages students' numerical and mathematical skills.

The sharing and promotion phase sees the results published on the SteamCity or Vittamap platform, presented to other classes in the network, and potentially submitted to the city council in the form of proposals for improving the sound environment. This restitution develops communication skills and civic engagement.



THE SKILLS DEVELOPED

The integration of open data into scientific investigations develops a coherent set of essential skills to train the citizens of tomorrow.

On the scientific level, students acquire methodological rigor in data collection and processing. They develop their ability to critically analyze information sources and learn to interpret complex data sets by identifying trends, correlations, and anomalies. Exposure to real data raises their awareness of the uncertainties and limitations inherent in any scientific measurement.

The digital skills mobilized go far beyond the simple use of tools. Students learn to manipulate different file formats, transform raw data into relevant visualizations, and publish responsibly on the internet while respecting the principles of open data and the associated legal constraints.

Citizenship skills are built through a concrete understanding of local issues. Active participation in citizen science projects develops a sense of personal and collective effectiveness in the face of environmental and societal challenges. Involvement in local life, through the production of data useful to the community, strengthens the bond between the school and its local area.



POINTS OF VIGILANCE AND RECOMMENDATIONS

The implementation of projects involving open data requires particular attention to certain legal and educational aspects.

Legally, compliance with the General Data Protection Regulation requires ensuring that personal data is never collected or published. Students must be made aware of these issues from the design stage of the experimental protocol. Verifying the licenses for the data used ensures respect for intellectual property and teaches good knowledge-sharing practices. Obtaining the necessary authorizations for measurements in public spaces requires dialogue with local authorities, which can itself be instructive.

From a pedagogical perspective, adapting complexity to the students' level remains essential. It is important to prioritize understanding a few aspects rather than superficially processing a large amount of data. Valuing the scientific approach, regardless of the spectacular nature of the results, encourages perseverance and intellectual honesty.



CONCLUSION

The integration of open data into school scientific investigations represents much more than a simple methodological enrichment. It fundamentally transforms students' relationship with scientific knowledge and their territory. By becoming producers of open data, students actively contribute to the construction of an informational commons. By exploiting existing data, they place their observations in a broader context that gives meaning to their work. By sharing their results, they participate in a collaborative dynamic that goes beyond classroom walls and national borders.

The SteamCity project provides the ideal framework for this pedagogical transformation. It provides the tools, resources, and network necessary to ensure that each school investigation becomes a meaningful contribution to understanding territorial issues. In this context, the teacher becomes a facilitator who supports students in their transformation into citizen-scientists capable of understanding, producing, and sharing data to take action in their territory.

This approach prepares students for the challenges of the 21st century by simultaneously developing their scientific culture, digital skills, and civic awareness. It makes the school a true laboratory of territorial innovation where the solutions of tomorrow are built.