**3. GIS and Open Street Map for participatory actions**

**Introduction**

This course is the absolute crashing module for those ones who have never done anything with open data and how to exploit them through GIS and Open Street Map systems. First, we want to explain to you why it is important to make data open and how we can do that. Second, some nice ways to do activities around them. Plus, it gives concrete tips on technical actions that can be done with open data within youth organisations.

This course introduces you to:

• the essentials of open data technical features

• the essentials of copyright and copyleft licenses

• how to clean data for beginners and main strategies

• how to organise mapathons and mapping data through different applications

• how to use QGIS application for data visualisation & analysis

**Intro**

Why open data? And what do they have to do with young people? Well, the usage of open data can be applied to many fields and it is not just about statistics. For example, if you are into archaeology studies, you can use open source and open-data-based technologies to reconstruct and classify cultural items. One of the most powerful tools is QGIS in this! Check it out in the next paragraphs.

Second. If you want to develop some nice research on social problems, it would be great to render such a work into a map with different pins and be able to make some analysis between them. How? Using Excel, for instance, and Google Maps or Open Street Maps. You can enrich your work with videos that are released in Creative Commons (how? Simply click the option on YouTube).

Open data are more and more used for any app, like for detecting statistics and using real-time data. Indeed, we can be good users of those if we get to understand why open data can be so crucial in terms of individual and organisational contributions to the local society.

1. **Open data**

Open data are data that anyone can access, use and re-use and share according to licenses. Their main characteristics are related to availability and their format, readable by machines. They are so important because many organisations can access them and creatively use them to analyse problems, develop apps, figure out local problems and trends to understand strategies and solutions.

Open data are mainly produced by States, individuals or organisations, profit or non-profit. Their main impact is referred to transparency and to enhance governance, making citizens aware of what happens in terms of public expenditure, statistics regarding crime, environmental data (like air pollution, water quality etc.), traffic, richness etc.

Open data are generally published in public repositories and have different formats that guarantee certain standards. Recognising them is a first fundamental step to use, re-use or produce them.

Making data open means for societies to grow, to collaborate, above all to know what it is going on in the places we live.

1. **Open data have to be clean and with clear licenses**

As we learned, open data must have specific features which make them ‘open’, particularly because we are discussing facts and figures which every machine and every human should understand. So, what is the first order of problems for a data really to be open and usable? Its level of cleanness is a crucial element.

Unfortunately, even if the majority of data are defined open, basically all of them are technically not cleaned. What is this cleaning?

Cleaning data means to set up consistent data ready for machine reading and for human analysis – let’s make an example. If we have a table full of percentages and statistics, we would expect numbers in all the cells.

What happens if in one of the cells we have some alphanumeric expressions like “a” or “love” , instead of “0.5%”? In this case we will have a problem to deal with and that is why we need to clean such data and to make it uniform and coherent. Having clean data means easiness in combining them with different datasets and gain deeper understanding.

In this way, if one has a clean criminality dataset per neighbourhood of a town and a clean dataset on wage distribution per area, it is possible to inter-relate those data and smartly act according to the obtained results.

As a second level of issues, we have the licensing. Simply because you put your data somewhere on the web, like a photo on a map, it does not mean it is ‘open’. Open does not mean something is simply online somewhere.

1. **Open data culture**

Open data culture is all about sharing, discussing, reusing and producing innovative products around a community and shared values on governance and digital social innovation.

One of the main rules around open data use, re-use and distribution is about giving the same licensing of the data we have retrieved to the data we have manipulated (sharealike principle).

For example, we have downloaded a hotel geographical distribution dataset. Once we have used it, we release the new product with the same licensing of the data we have manipulated.

Open data licenses can be found very easily and good standards focus on the traceability of what it is produced, so that any description should be self-evident.

1. **How to orienteer yourself with licenses**

Giving licenses is always a pain. To orienteer yourself in what you should do when you upload any of your creative products, you should really ask yourself “What is my purpose in sharing this content?”. In this sense, you have to be aware of the rationale of your activities and of the desired outcomes.

In fact, if you are the ‘maker’ of something, it is up to you to decide which freedom other users will have on your content. To make a practical example: if I want my music to be remixed or cut without further authorisations, you can release it on YouTube under Creative Commons licensing, in this way, users can edit it and quote your previous work. If you release it under normal copyrights, it means people cannot do anything but listen to and share your work.

This is an example, and it can be seen as a true discussion around openness of data and what to do with them and, above all, why.

A very good website to get a good grasp of licenses on open data is Creative Commons, giving you valuable definitions and guidance on how to give credit and set up your licensing.

1. **Open data cleaning**

When we talk about cleaning of data, we basically refer to formatting of statistics data, so we have to take care of the quality and the coherence of how data are drafted in the file we have downloaded (for reference, check Unit 1: main platforms).

Cleaning is a draining process and selection of information and it can be done automatically or manually. In order to do it automatically, one can use apps or bots that must be programmed. For our scopes, a manual check on the files is enough, certainly it takes some time, but it is the ideal mechanism.

Some advantages of cleaning:

• Cleaning makes the reading process intuitive, simple, coherent and adds specific nuances that ‘dirty’ raw data fail to have. Cleaning is crucial and it is fundamental in an open-data-driven project;

• A cleaned open dataset will raise the reputation of the entity that released it, therefore putting the basis for a community as well as making the open dataset usable;

• Cleaning is not just about cutting the information, but it is also about enriching the existing open dataset, for instance adding details and identifiers that users will much appreciate.

• Cleaning means also preparing data to be processed and analysed to deliver coherent analysis/reports

But how or what to do with a cleaning set up? It depends on our aim and how we want to manipulate our datasets.

So, it is worth looking at the following section dealing with the main mistakes on the datasets.

1. **Wrong data formats**

Many times, data can be written without consistency, for example there are many ways to mention a specific date:

06.10.2019 or 10.06.2019 can mean very different dates whether we mean DAY.MONTH.YEAR or MONTH.DAY.YEAR - this is a fundamental difference between US format vs European format.

Plus, one could use letters instead of numbers, like 06 October 2019 – from a machine point of view letters can lead to errors.

**Multiple representations**

Similar to the previous type of error, abbreviations can lead to consistency faults. In fact, differences in capitalisation, how spaces are placed or even adjective position can cause all the possible errors when ‘chewing’ data.

**Duplicated records**

These mistakes are common when datasets get combined and the same value is present. In other cases, it occurs when the same piece of data has been entered more than once in the same open dataset.

**Redundant data**

Data not relevant to your work is considered redundant. Very common mistakes are rows that represent total amounts or other columns which are put to assist human readability.

Mixed numerical scale

Common errors are inconsistencies in units used (pounds or euros? Kilos or grams? Thousands or millions?). Plus, writing numbers in different ways like “1.2m” instead of “1,200,000” can lead to very diverse results. For instance, if one writes “800,000” and compare it to “1.2m”, the machine can read it as a smaller value and cause an error.

Mixed ranges

To avoid errors in terms of value, it is good to separate low and high values, so that datasets are ordered.

Generic spelling errors

Running data through a spell-checker is fundamental – in this way one can detect problems in how data is written.

GIS stands for Geographic Information System, which is a computer-based tool used for capturing, storing, manipulating, analyzing, and presenting geographical or spatial data. It helps users to visualize, understand, and interpret data in the form of maps, graphs, and tables. In simple terms, GIS allows us to take data related to a particular location, such as population, land use, elevation, and combine it with maps to reveal patterns, relationships, and trends that can help us make better decisions.

OpenStreetMap (OSM) is a collaborative project to create a free and open map of the world. It is built by a community of mappers who contribute and maintain data about roads, buildings, landmarks, and other features using GPS devices, aerial imagery, and other mapping tools. Anyone can edit and use OSM data, subject to certain license terms.

HOTOSM (Humanitarian OpenStreetMap Team) is a non-profit organization that works to support humanitarian and development efforts by using open mapping data. HOTOSM works with partners around the world to provide mapping support in response to natural disasters, health emergencies, and other crises. HOTOSM tools and resources are designed to enable volunteers, communities, and organizations to work together to create and share geospatial data for the benefit of all. HOTOSM uses OpenStreetMap data as a foundation to support their humanitarian missions.

GIS (Geographic Information Systems) and OpenStreetMap are powerful tools that can be used to facilitate participatory processes in rural areas. Here are some ways in which they can be used:

**How they can be used for participatory processes**

1. Mapping: GIS and OpenStreetMap can be used to map rural areas and identify important features such as water sources, farmland, and forests. These maps can be used as a basis for discussion in participatory processes, helping to identify areas of concern and potential solutions.
2. Data collection: GIS and OpenStreetMap can be used to collect data from local communities. This can be done using mobile apps or GPS devices. This data can be used to identify patterns and trends, and to inform decision-making in participatory processes.
3. Visualization: GIS and OpenStreetMap can be used to visualize data in a way that is easily understandable for local communities. For example, maps can be used to show the impact of climate change on crop yields, or the distribution of health services in rural areas. This can help to engage local communities and encourage their participation in decision-making.
4. Planning: GIS and OpenStreetMap can be used to help plan and design rural development projects. This can include identifying suitable locations for infrastructure such as roads and water supply systems, and analyzing the impact of different development scenarios on the local environment and community.

**Mapathons and benefits**

A mapathon is an event where a group of people come together to improve maps for a particular area using OpenStreetMap. OpenStreetMap is a free and open source map of the world, created and maintained by a global community of volunteers.

A mapathon can be useful in many ways. It can help create accurate maps of rural areas that may not have been mapped before. Accurate maps can be useful for emergency response, resource management, and community planning. For example, accurate maps can help responders locate and reach people in need during natural disasters. They can also help local governments and organizations plan and manage resources, such as water or transportation. In addition, maps can help communities identify and address issues such as land use, access to services, and environmental concerns.

Participating in a mapathon can also be a fun and rewarding experience. It can bring people together to work towards a common goal, learn new skills, and connect with others who care about their community. It can also provide an opportunity to contribute to a global effort to improve maps and make them accessible to everyone.

**How to organise a mapathon**

1. Plan your event: Decide on the date, time, location, and format of your mapathon. Will it be in-person or virtual? Will it be a one-time event or a series of events? Determine your target audience and reach out to them through social media, email, or other channels.
2. Choose your mapping project: Visit the HOTOSM Tasking Manager to find a mapping project that fits the interests and skills of your participants. The Tasking Manager provides a list of projects that need mapping support, including disaster response, community development, and environmental conservation. Choose a project and familiarize yourself with its goals and mapping instructions.
3. Prepare for the event: Create a detailed agenda for the event, including introductions, mapping instructions, breaks, and debriefs. Determine what mapping tools and software you will use and make sure all participants have access to them. Prepare any necessary training materials or tutorials to help participants get started with mapping.
4. Host the mapathon: On the day of the event, welcome your participants and provide an overview of the project and mapping process. Provide instructions on how to use the mapping tools and software, and facilitate the mapping process. Encourage collaboration and answer any questions or concerns that participants may have.
5. Debrief and follow-up: At the end of the event, debrief with participants and reflect on the mapping process and accomplishments. Discuss any challenges and opportunities for improvement. Share the results of the mapping project and celebrate the collective effort of the participants. Follow up with participants and share any future opportunities for mapping or engagement with the project.