Report on the role and responsibilities of managers of data cooperatives, citizen observatories, and open data archive

Based on Deliverable D4.3

by
Michael O'Grady, Niccolò Tempini

Report on the role and responsibilities of managers of data cooperatives, citizen observatories, and open data archives

Considering novel models of participation and deliberation, deliverable 4.3 focuses on the data dimension, taking into consideration three forms of data management that encapsulate different but complementary models of governance: data cooperatives (bottom-up deliberation and delegation of data management), public data repositories often driven by international agencies (archetypical top-down model), citizen observatories (may follow either model, however, in practice, are mostly top-down initiatives).

More in depth:

i) the first task of the study focuses on health data cooperatives, where a qualitative case study identified some of the key organizational challenges and best practices associated with such models;

ii) the second one focuses on citizen observatories and classic citizen science, undertaking a quantitative survey of data collection, archiving, and dissemination, with specific emphasis on open data practices;

iii) the third and last task focuses on open data repositories, evaluating the usability and accessibility of several well-known data repositories from the perspective of the average citizen.

Health data cooperatives

Emerging over the last 10 years, a health data cooperative is an organization that enables individuals to pool, manage, share, and benefit from the re-use of their health data. It is meant to collect, store, process, and share health data, which can be used for a variety of purposes such as research, disease prevention, and personalized medicine.

It benefits both individual members and the broader healthcare community, all while ensuring the privacy and security of the members' data. In principle, and discounting for the deep-seated demographic inequalities, this model of data management and sharing can help democratize management of health data.

Research shows that Health Data Cooperatives (HDCs) present an innovative approach to data management that could, in principle, significantly shift the status quo in health research. They operate on the fundamental principles of individual agency and autonomy over personal data, thus aligning with the increasing demand for individual control over personal data, and offer a potential solution to the prevailing challenges surrounding consent policies in health research, the ethically problematic blanket consent and the scientifically restrictive single-use consent.

Their potential extends beyond research to a wider societal context, as they can potentially serve as an important tool for guiding government policymaking. The rich data collected within a HDC can also include, and reflect, priorities and research questions formulated by the member community.

Some HDCs best practice recommendations include:

- fostering transparency about 3rd party data sharing;
- prioritizing public good in research;
- implementing robust accountability and governance principles;
- encouraging bottom-up development;
- establishing an ethics oversight function.

Citizen observatories

Citizen observatories (Cos) are considered a potential tool for promoting democratic participation. Through this study, several recommendations concerning the design and implementation of COs in the citizen science domain were identified. These correspond to each of the standard Responsible Research and Innovation (RRI) keys, thus regarding open access, public engagement, science education, gender, ethics, and governance. From an ISEED perspective, the most interesting finding concerned the sixth RRI key – governance. At the time of writing, governance within the context of COs is being actively researched outside of ISEED, with results anticipated within the next few years.

The survey of citizen scientists resulted in 100 submissions for analysis after the quality control check. The gender breakdown was 53% female, 45% male with 2% not answering. The age profile was dominated by the 35-44 cohort (31%). 95% of participants were from Europe with 15 countries being represented. Over 50% described themselves as active citizen scientists while others classified themselves as managers or project leaders. Biodiversity, earth science, and environmental science were the most popular specializations. Most projects were scoped at the country (23%) and regional levels (32%). 46% of projects were driven by an academic institution. Most participants were active in citizen science for less than 5 years.

To those considering harnessing the citizen science concept and community in various democratic activities, here are some brief remarks on the dataset:

- 36% of participants would freely share their data with a for-profit organization;
- 64% of participants were familiar with Open Data but only 54% with Open Science;
- 87% of participants understood GDPR but only 37% encountered the FAIR data principles;
- 49% of participants identified Civic and community benefits as a motivation to contribute to Citizen Science.

Open data repositories

Fifty participants contributed to the last survey, concerning the usability of open data portals, with a Cronbach Alpha score of 0.87 confirming that the internal consistency of the survey was good. A total of 200 answers were provided (5 portals x 4 Questions x 10 participants). Results were as follows:

	n=200	*%*
Incorrect	56	28%
Correct	127	65%
Good Effort	17	9%

The overall sentiment score for the first dimension of participants' experience - how they rate their own experience, was 3.49 on a scale of 1(negative) to 5 (positive). Extrapolating the results would suggest that perhaps up to 70% of the EU population would succeed in finding data on an open data portal. However, two limitations must be acutely kept in mind – the small nature of this study and the population being computer literate.

Observations and recommendations

Those considering promoting data cooperatives, citizen observatories, or open data repositories as part of an initiative to contribute to policy definition at the local or national levels, must first develop a deep understanding of some fundamental principles about the participating population.

The following constitute a series of observations and implicit recommendations that may be considered by managers during planning:

- Observations:
  - Citizen scientists are not representative of a population. They represent a subgroup, often distinguished by age profile. Other characteristics may also arise, for example, gender, educational profile, and so forth.
  - Citizen scientists tend to contribute to projects operating at regional or greater levels, rather than being restricted to their immediate locality.
  - Citizen scientists tend to contribute to longer-term initiatives.
  - Many citizen scientists contribute to the decision-making, even though initiatives tend to be hierarchical and top-down driven.
  - Though data collection is a ubiquitous activity amongst the citizen science community, other activities including problem definition and research question specification are actively contributed to.
  - While many citizen scientists are data literate, a smaller but still significant group is agnostic to data management activities.
  - A good understanding of GDPR and data protection is manifested by citizen scientists.
  - There is a significant lack of training, both formal and informal, made available to citizen scientists.
  - Open Data Repositories are known to citizen scientists but have not been used extensively.
  - If computer-literate, open data portals are generally accessible and usable by the public.
- Recommendations
  - Citizen scientists are motivated by many factors; however, civic and community benefit is a key motivation and could be effectively harnessed.
  - While citizen scientists may be regarded as a distinct population, other activities e.g., bird watching, where similar activities are undertaken but participants do not identify themselves as citizen scientists, should be considered.
  - Risks associated with sharing both personalizable information as well as location should be considered and communicated.
  - Open science, and to a lesser extent, open data, are not fully understood by the broad citizen science community, indicating a need for better operationalization of open science.
  - When adopting a citizen observatory model, significant attention must be paid to the issue of governance.
  - Inclusion needs consideration under many headings, including technology literacy.
  - Responsible Research and Innovation (RRI) is not widely understood in the broad citizen science community, indicating a need for better operationalization of RRI.