IRoSiE

Training Materials for Responsible Open Science

**Case study 1**

Open access biodiversity data

SOURCE: Quinn, A. (2021). Transparency and secrecy in citizen science: Lessons from herping. *Studies in History and Philosophy of Science Part A*, *85*, 208-217. https://doi.org/10.1016/j.shpsa.2020.10.010

eBird is an online platform for posting observations of birds that was launched in 2002 by the Cornel Lab of Ornithology. Now eBird is among the world's largest biodiversity-related science projects with more than 100 million bird sightings contributed annually by eBirders around the world. In 2008, the success of eBird inspired three students to develop a platform for natural history observations of any organism - iNaturalist. Now iNaturalist is maintained by the California Academy of Sciences and the National Geographical Society. As of 2020, iNaturalist hosted 42 427 731 observations contributed by 1 149 886 observers. Both platforms bring substantial epistemic benefits by contributing data for natural sciences and life sciences research, as well as for educational purposes. As A. Quinn points out: "Citizens contributing observations in a casual manner can rapidly expand the scale of the dataset far beyond what is possible using traditional research methods. Moreover, it is not even possible to predict what kind of knowledge can be pulled from long-term, huge datasets." Besides that, the platforms encourage many people to experience nature. However, the popularity of the platforms also may create a danger for many species. Smartphone photos uploaded by volunteers contain location coordinates, and, for example, iNaturalist by default makes observation locations visible to all users. While users can choose an option to hide observation locations when uploading data, it is easy to forget to enable this function and there still is a potential for overrides. Many users also are unaware of the risks tied to sharing location information. While iNaturalist automatically hides location data for some species in need of conservation, users are primarily accountable for hiding location data.

The main concern is that the information might be used by poachers. For many species, the more severe threat is the destruction or degradation of habitat. A record of an interesting species on iNaturalist might attract many people who might go to look for the species. But as many people are ignorant of what interventions can destroy habitat, such visits might turn out to be fatal for the animals. This concern is especially raised by the community of herpers as herps (amphibians and reptiles) are very sensitive to any changes in their habitat.

**Questions for discussion:**

1) How to responsibly implement the principle 'as open as possible and as closed as necessary' regarding biodiversity data?
2) Platforms like iNaturalist promote citizens' interest in the natural world by enabling participation. However, by doing that they also increase interest in finding rare species, which in turn might pose a risk to habitats that are

This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No 101006430

necessary for their survival. Are there any ways to avoid or at least minimize the risks created by these platforms?

3) Should there be stricter ethical guidelines for sharing location-specific data of rare or sensitive species on platforms like iNaturalist?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No 101006430

IROSiE

Training Materials for Responsible Open Science

**Case study 2**

Data privacy in water sciences

| Zipper, S. C. et al. (2019). Balancing open science and data privacy in the water sciences. *Water Resources Research*, *55*(7), 5202-5211. https://doi.org/10.1029/2019WR025080 |
| --- |
| Open science practices, like sharing data, research results and code, are providing new opportunities to scientists working in the field of water science, e.g., by enhancing data availability and reproducibility. However, as physical and social science domains are merging in fields like socio-hydrology, researchers may inadvertently compromise privacy and security when sharing sensitive information in open access. High-resolution spatial data, including satellite data, hydrological model outputs, and other geospatial datasets, are widely used in hydrologic sciences. This type of data can be sensitive, despite not meeting traditional human subject research definitions. For example, 30% of farmers in Iowa, US in a survey stated that collecting geospatial data on private property compromises privacy. Also, other groups of data, like water or electricity consumption, are potentially sensitive. The possibility of commercializing the data amplifies these concerns by adding issues of data ownership.<br>Privacy and other risks may increase when researchers lack cultural understanding and sensitivity. There have been cases when powerful groups or companies have misused open data "at the expense of the intended beneficiaries of the shared data". For example, the digitization of land records in Karnataka, India which was intended to democratize information access, instead empowered wealthy landowners. These concerns may be particularly important in environmental justice contexts, especially when working with indigenous communities and historically disadvantaged groups. |

**Questions for discussion:**

1) How sharing the different types of data mentioned in the case description might violate the privacy and security of individuals or communities?
2) Do you agree with the authors' statement that: "Natural scientists have little guidance to deal with privacy concerns for open science, which are inherent in socio-environmental research"?
3) What should the scientists do to protect data privacy and security?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No 101006430

IROSiE

Training Materials for Responsible Open Science

**Case study 3**

Open geospatial data in agriculture research

Prince Czarnecki, J. M., & Jones, M. A. (2022). The problem with open geospatial data for on-farm research. *Agricultural & Environmental Letters*, 7(1), e20062. https://doi.org/10.1002/ael2.20062

On-farm research in agriculture may involve collection of spatially referenced farm data, such as fertilizer application, plant populations, and yield, which can be traced back to individual properties and private collaborators. At the same time, this type of research lacks comprehensive approaches for de-identifying geospatial data. Commonly used techniques for general geospatial data, like random perturbation and temporal cloaking, are not well-suited for farm data. Shifting point locations and altering time stamps may distort statistical measures of the data without ensuring adequate privacy. Another approach involves removing geospatial references and rescaling points to a spatially correct grid. While this helps to keep spatial relationships, it hinders contextual analyses as features with geographic concurrence cannot be identified and it is not possible to use, e.g. climate data for analysis.

In the situation where publishers and funders require researchers to share research data in open access, researchers in the field of on-farm research face challenges to ensure open access and at the same time safeguard the privacy of farm owners. Some researchers state that insisting on the publication of the data may make the involvement of collaborators more complicated, and the requirement of de-identification of data may make on-farm research more challenging.

**Questions for discussion:**

1) How sharing the geospatial data mentioned in the case description might violate the privacy and security of individuals or communities?
2) Do natural scientists have the necessary guidance, knowledge and skills to deal with privacy concerns in the context of open science?
3) What should the scientists do to ensure data privacy and security?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No 101006430