

Ethical and legal concerns regarding user-generated data: experiences from a GDPR-compliant research project

Christoph Fink^{1*}
Enrico Di Minin¹
Anna Hausmann¹

¹ Helsinki Lab of Interdisciplinary Conservation Science, Department of Geosciences and Geography, University of Helsinki, Finland.

* corresponding author:

christoph.fink@helsinki.fi

Pieteri Kalmin katu 5, 00560 Helsinki, Finland

Social media data and other crowd-sourced information are being increasingly used in conservation science, for instance to study human-nature interactions. User-generated social media content, such as images, video, text, and audio, and the associated metadata can be used to assess such interactions at a fine spatio-temporal scale. Many of the addressed research questions are highly sensitive in nature: They touch upon societal conflicts (such as conservation priorities clashing with development goals) or even uncover illegal activities (such as trade in protected species). The risks to individual users' privacy and well-being are significant when it comes to a potential misuse of these data.

In this contribution we report from a recently established research project in Conservation Geography that is centred around the use of crowd-sourced information, especially social media data, and that benefited greatly from a meticulous planning of data handling and data processing procedures. We explain how a Data Protection Impact Assessment helped us to implement a risk-based approach to data protection in order to identify and minimize privacy risks to social media users, to demonstrate accountability, and to comply with data protection legislation, e.g. the European Union's (EU) new General Data Protection Regulation (GDPR). We also introduce a framework to collect, store, protect, share, and manage social media and other crowd-sourced data in a way that reduces the risk of profiling and identifying social media users. Finally, we discuss the legal basis for processing social media data while ensuring data subject's rights, using EU's GDPR as a case study.

The experiences we gained during setting up this project are of great value to all researchers working with user-generated data, such as Volunteered Geographic Information (VGI), social media data, or any other data source that potentially contains personal information.