ETHICS AND INTEGRITY IN OPEN SCIENCE

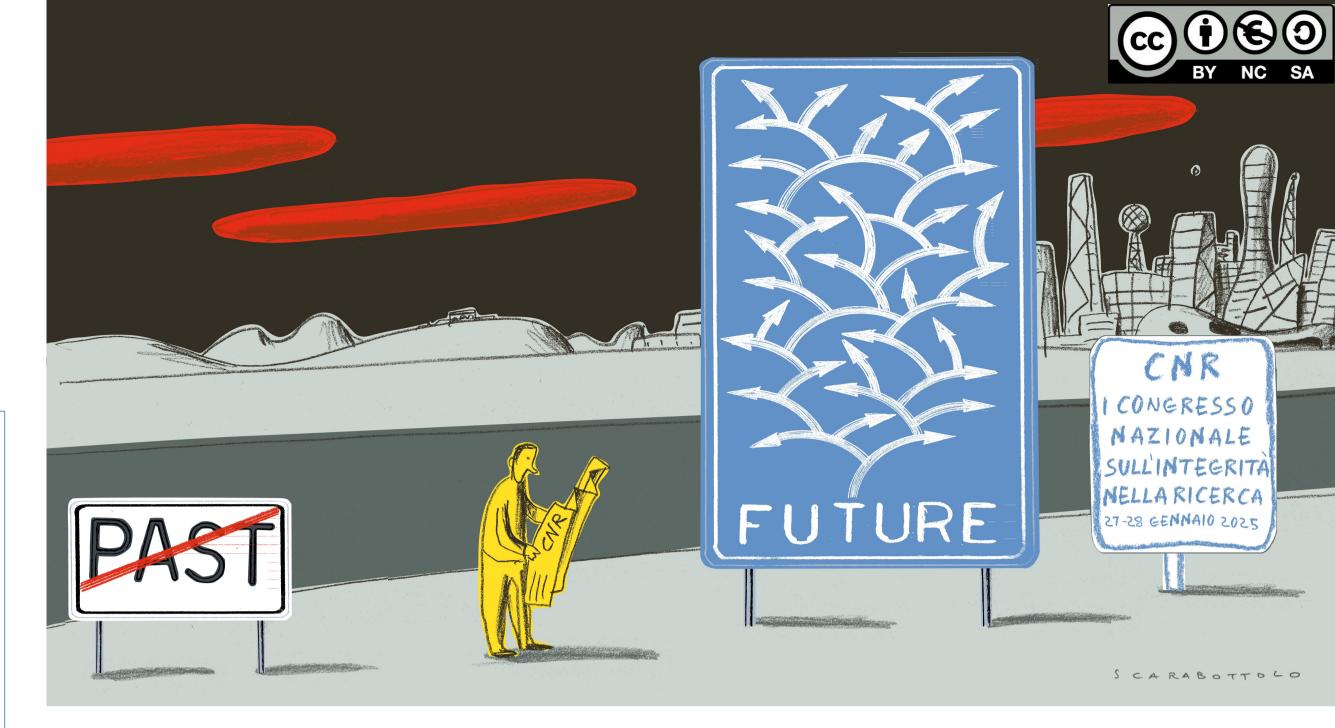## IMPLICATIONS FOR RESEARCH INFRASTRUCTURES IN THE SOCIAL SCIENCES

Giulia Sciolli and Roberta Martina Zagarella

Centro Interdipartimentale per l'Etica e l'Integrità nella Ricerca, Consiglio Nazionale delle Ricerche

Email: giulia.sciolli@cnr.it; roberta.zagarella@cnr.it

INTRODUCTION

▪ **Open Science** is a cornerstone for enhancing transparency, reproducibility and integrity in research.

▪ The development and sustainability of **Research Infrastructures (RIs)** are fundamental to foster Open Science. RIs provide free and open services, including laboratories, data networks, advanced technological tools and digital platforms for the collection, analysis, reuse and sharing of scientific data, according to the principles of **Open Science**. In the field of **social sciences**, RIs facilitate **in-depth analysis of complex issues** such as population aging, inequalities, migrations and the **implementation and evaluation of social policies**, offering researchers the opportunity to access comparable data, develop interdisciplinary collaborations, and offer **scientifically grounded solutions**.

• In this context, the goal of the **PNRR** funded project **FOSSR - Fostering Open Science in Social Science Research** is to create an **Italian Open Science Cloud** dedicated to the social sciences, which can both strengthen the existing Italian nodes of such RIs and provide the national scientific community with innovative tools and services.

▪ Against this background, existing literature on the **ethical issues** relevant to **Research Infrastructures**, especially for **empirical social sciences**, is still poor and hard to find, tends to be repetitive and country-specific, and often does not offer constructive critique but only points out problems.

METHODS

▪ The FOSSR project has been developed following an ***ethics by design* approach,** through the establishment of a specific task dedicated to ethical issues and of an independent Ethics Committee.

▪ As part of the activities of the FOSSR Project WP2, Task 2.5 "FOSSR Ethical Issues", a **scoping literature review** was conducted to identify critical ethical issues, existing practices, research gaps and implications for the creation and management of Research Infrastructures for empirical social sciences.

▪ A total of **39 resources** were collected, classified in three categories – scientific articles (21), grey literature (13) and books (5) – and analysed by the two authors.

RESULTS AND DISCUSSION

The analysis identified **three** particularly **recurring** and interconnected **issues**.

▪ The **social sciences** use a **wide variety of data** (images, text collections, data gathered from social media and surveys, etc.) and it is important to consider that "qualitative" data that can be made open might also include, depending on the specific discipline: transcripts of unstructured, semi-structured, or structured interviews, or of focus groups; audio or video recordings; notes (including fieldnotes from participant observation in ethnographic research); maps and geographical data; official documents, reports and meeting minutes; autobiographies, diaries, letters, and photographs. Different disciplines within the social sciences therefore generate different issues regarding the ethics and integrity of "open data." Despite such complexity, there is still **little clarity** on the **concrete meaning and implications of "open data" and "FAIR principles" for the social sciences**, especially for individual disciplines within the social sciences.

▪ The variety and specificity of these data are not recognized, but they are often talked about as if they were quantitative data. The vast majority of the literature pushing towards data openness is indeed based on a 'positivist' conception of data. From this perspective, **data sharing is considered primarily a technical issue** that can be managed through the implementation of appropriate data management tools and policies. While some types of data – such as data collected through surveys – can be more easily shared and reused, other kinds of data – such as interview transcripts or field observations – are strongly tied to the context of collection and to the person collecting them, and suffer from too much **decontextualization** once archived and shared. Consequently, researchers accessing and reusing such data would not be able, for example, to anticipate and consider any undeclared **biases**, which are part of the 'baggage' of those data.

▪ Given their specificity, for these kinds of data it is more complex to balance ethical imperatives (e.g., protecting research participants) with scientific needs (e.g., ensuring that data are indeed reusable). In fact, **data openness** allows for the critical examination and evaluation of results, and facilitates efficiency; but, at the same time, it also **makes data vulnerable to unethical practices**. In qualitative research, even if the details of some content can be modified or removed, there may be other elements that lead to the indirect identification of participants, especially in narrative data. The question thus arises about the extent to which it is necessary to modify data to make the transcript of an interview, for example, shareable without any risk to participants, and whether at that point, after such modifications and removals, the transcript will still be usable and useful to other researchers.

CONCLUSIONS AND NEXT STEPS

• For RIs in the social sciences, the problem is how to ensure robust ethical frameworks that enable a balance between openness and the protection of research subjects. In fact, the social sciences involve topics and methods that result in data that are dialogic, emergent, abundant and relational (so-called **DEAR data**), and that, therefore, need to be treated not only following the **FAIR principles** (findable, accessible, interoperable, reusable), but also the **CARE principles** (collective benefit, authority to control, responsibility, ethics). Awareness and knowledge of these issues and of their implications are fundamental for researchers to ensure that their projects are conducted with attention to integrity, and that they promote responsible Open Science. **FOSSR**, as the future **Italian Infrastructure for the social sciences**, will also have to deal with these issues. Therefore, the considerations emerged in the literature review constituted the springboard for structuring an "ethical toolkit" that will provide operational guidelines for research carried out using the resources made available by FOSSR.