

Chapter 4

Ethical Implications in Building Longitudinal Data Infrastructures



Loredana Cerbara , Dario Germani , and Michele Santurro 

4.1 Introduction

The transition to open science has been increasingly recognized as a transformative movement in the research landscape. It goes beyond merely providing open access to publications and data, promoting instead new norms of collaboration, inclusivity, and public accountability. The promise of open science lies in its ability to democratize knowledge production, enhance reproducibility, and enhance public engagement with science (Fecher & Friesike, 2014; Haven et al., 2022).

As open science initiatives expand in scale and complexity, they raise important ethical questions. How can openness be balanced with privacy protection? What are the implications of making basic personal data more widely accessible? How can transparency coexist with the need for confidentiality, particularly when working with vulnerable populations? What safeguards are needed to mitigate risks associated with personal data processing? These questions highlight the necessity of embedding ethics at the core of open science practices right from the design stage.

Rather than treating ethics as a checklist of requirements, contemporary approaches advocate for an embedded ethics model that integrates ethical reflection into all stages of the research process (Iphofen, 2017). This framework shapes study design, consent procedures, data sharing strategies, and infrastructure development, ensuring legal compliance alongside genuine respect for participants' autonomy and dignity. In open social science research, this means embedding respect for autonomy, privacy, and fairness throughout the research lifecycle, from design to data dissemination (European Commission, 2023). Key components include transparent consent, strong data protection, and inclusive communication tailored to diverse populations (Sieber, 2012).

L. Cerbara (✉) · D. Germani · M. Santurro
Institute for Research on Population and Social Policies (CNR-IRPPS), Rome, Italy
e-mail: loredana.cerbara@cnr.it; dario.germani@cnr.it; michele.santurro@cnr.it

© The Author(s) 2026
L. Taddei, M. Paolucci (eds.), *Longitudinal Data Infrastructures in Europe*,
https://doi.org/10.1007/978-3-032-07005-0_4

In this chapter, these overarching ethical challenges are addressed through a comparative analysis of prominent European longitudinal studies—namely, the LISS Panel (Netherlands), the UK Household Longitudinal Study (UKHLS), and the German Socio-Economic Panel (SOEP). These cases offer concrete illustrations of how data protection, informed consent, and trust-building mechanisms are implemented in practice, in accordance with the General Data Protection Regulation (GDPR) and broader ethical standards. The discussion then turns to the Italian experience of the FOSSR project (Fostering Open Science in Social Science Research), which illustrates how open science principles can be embedded in the development of new research infrastructures. Special attention is given to four longitudinal surveys developed within FOSSR—GUIDE, GGS-II, SHARE, and IOPP—each characterized by a specific thematic and demographic focus. The chapter outlines the tools and materials adopted to ensure legal compliance, ethical accountability, and transparent communication across these initiatives. Together, these experiences underscore the importance of designing longitudinal data infrastructures that are not only technically sound and legally compliant, but also ethically proactive and socially responsive.

4.2 Implementing Ethics and Privacy Standards: The Cases of LISS, UKHLS, and SOEP

The ethical and privacy management of longitudinal survey data in Europe requires robust frameworks that protect participant confidentiality while enabling meaningful social science research. A primary challenge is safeguarding sensitive personal information, including health data, against disclosure risks. Over recent decades, considerable attention has been devoted to statistical disclosure limitation techniques, which reduce the possibility that individuals can be re-identified from anonymized datasets (e.g., Lambert, 1993; Fuller, 1993; Jabine, 1993; Rubin, 1993; Fienberg & Willenborg, 1998; Doyle et al., 2001). This concern is especially relevant in longitudinal studies, where repeated measurements may create unique patterns facilitating re-identification. As Couper et al. (2008) highlight, participants' willingness to engage in surveys depends not only on technical and legal protections but also on their perceptions of privacy risks and confidentiality. This highlights the importance of transparent communication and trust-building in ethical and privacy-friendly survey design.

Statistical disclosure risk refers to the potential for linking anonymized records to external databases with overlapping attributes containing identifiable information, thus compromising confidentiality even without direct identifiers like names or addresses. Documented re-identification cases (Winkler, 1997; Malin & Sweeney, 2000) demonstrate the need for ongoing vigilance. However, empirical evidence quantifying disclosure risks in public-use datasets remains limited.

Beyond technical safeguards, ethical frameworks in European longitudinal surveys must also address participants' perceptions and concerns about privacy and