

● LA CASSETTA DEGLI ATTREZZI

STRUMENTI PER LE SCIENZE UMANE

ITALIAN ONLINE PROBABILITY PANEL

Ideazione, realizzazione e contesto
del primo panel probabilistico online in Italia

a cura di

Loredana Cerbara, Claudia Pennacchiotti,
Ilaria Primerano, Luciana Taddei

FrancoAngeli 



La cassetta degli attrezzi. Strumenti per le scienze umane

Direttore

Giovanni Di Franco, Università di Roma “La Sapienza”

Comitato editoriale

Elena Battaglini, Ires-Cgil

Sara Bentivegna, Università di Roma “La Sapienza”

Claudio Bezzi, valutatore professionista

Orazio Giancola, Università di Roma “La Sapienza”

Alberto Marradi, Università di Firenze

Federica Pintaldi, Istat

Luciana Quattrociochi, Istat

Luca Salmieri, Università di Roma “La Sapienza”

Michele Santurro, Cnr - Roma

Marta Simoni, Iref-Acli

La collana, rivolta a ricercatori accademici e professionisti, studiosi, studenti, e operatori del variegato mondo della ricerca empirica nelle scienze umane, si colloca sul versante dell’alta divulgazione e intende offrire strumenti di riflessione e di intervento per la ricerca.

Obiettivo è consolidare le discipline umane presentando gli strumenti di ricerca empirica, sia di raccolta sia di analisi dei dati, in modo intellegibile e metodologicamente critico così da consentirne l’applicazione proficua rispetto a definiti obiettivi cognitivi.

I testi sono scritti da professionisti della ricerca che, attingendo alla personale esperienza maturata in anni di attività, offrono ai lettori strumenti concettuali e tecnici immediatamente applicabili nella propria attività di ricerca.

Tutti i volumi pubblicati sono sottoposti a referaggio.



Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più: [Pubblica con noi](#)

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "[Informatemi](#)" per ricevere via e-mail le segnalazioni delle novità.

ITALIAN ONLINE PROBABILITY PANEL

Ideazione, realizzazione e contesto
del primo panel probabilistico online in Italia

a cura di

Loredana Cerbara, Claudia Pennacchiotti,
Ilaria Primerano, Luciana Taddei

La cassetta degli attrezzi
Strumenti per le scienze umane/10119.3

FrancoAngeli 

Il contenuto di questo volume riflette esclusivamente le opinioni degli autori. La Commissione europea e il Ministero dell'Università e della Ricerca (MUR) non possono essere ritenuti responsabili per qualsiasi uso delle informazioni in esso contenute.

Questo lavoro è stato realizzato nell'ambito del progetto **FOSSR – Fostering Open Science in Social Science Research**, finanziato dall'Unione europea – NextGenerationEU nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNRR), Grant Agreement n. MUR IR0000008, Missione 4 “Istruzione e Ricerca”, Componente 1, CUP B83C2200395000.



Progetto grafico di copertina di Maria Teresa Pizzetti

Isbn cartaceo: 9788835185741

Isbn e-book: 9788835193357

Isbn e-pub Open Access: 9788835193364

Copyright © 2026 by FrancoAngeli s.r.l., Milano, Italy.

Publicato con licenza *Creative Commons*
Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale
(CC-BY-NC-ND 4.0).

Sono riservati i diritti per Text and Data Mining (TDM), AI training e tutte le tecnologie simili.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore.
L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni
della licenza d'uso dell'opera previste e comunicate sul sito
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

Gli eventuali link attivi e QR code inseriti nel volume sono forniti dall'autore. L'editore non si assume alcuna responsabilità sui link attivi e QR code ivi contenuti che rimandano a siti non appartenenti a FrancoAngeli.

119. La cassetta degli attrezzi. Strumenti per le scienze umane

Volumi pubblicati:

1. Giovanni Di Franco, *L'analisi dei dati con SPSS. Guida alla programmazione e alla sintassi dei comandi.*
2. Silvia Cataldi, *Come si analizzano i focus group.*
3. Federica Pintaldi, *Come si analizzano i dati territoriali.*
4. Giovanni Di Franco, *Il campionamento nelle scienze umane. Teoria e pratica.*
5. Lucia Coppola, *NVivo: un programma per l'analisi qualitativa.*
6. Simone Gabbriellini, *Simulare meccanismi sociali con NetLogo. Una introduzione.*
7. Giovanni Di Franco, *Dalla matrice dei dati all'analisi trivariata. Introduzione all'analisi dei dati.*
8. Giovanni Di Franco, *Tecniche e modelli di analisi multivariata.*
9. Federica Pintaldi, *Come si interpretano gli indici internazionali. Istruzioni per l'uso a favore di ricercatori, giornalisti e politici.*
10. Maria Paola Faggiano, *Gli usi della tipologia nella ricerca empirica.*
11. Danilo Catania, *Dati e rappresentazioni territoriali con ArcGis.*
12. Claudio Bezzi, *Fare ricerca con i gruppi. Guida all'utilizzo di focus group, brainstorming, Delphi e altre tecniche.*
13. Giovanni Di Franco (a cura di), *Il poliedro coesione sociale. Analisi teorica ed empirica di un concetto.*
14. Elena Battaglini, *Sviluppo territoriale. Dal disegno della ricerca alla valutazione dei risultati.*
15. Claudio Bezzi, *Domanda e ti sarà risposto. Costruire e gestire il questionario nella ricerca sociale.*
16. Elena Pavan, *La Network Analysis con Nodexl*
17. Maria Paola Faggiano, *L'analisi del contenuto di oggi e di ieri. Testi e contesti on e offline.*
18. Giovanni Di Franco, *Usi e abusi dei sondaggi politico-elettorali in Italia. Una guida per giornalisti, politici e ricercatori.*
 - 1.1. Sergio Mauceri, *Qualità nella quantità. La survey research nell'era dei Mixed Methods*
 - 1.2. Alberto Marradi, *Tutti redigono questionari. Ma è davvero così facile?*
 - 1.3. Giovanni Di Franco, *Introduzione alla sociologia empirica. Metodi e tecniche della ricerca sociale.*
 - 1.4. Vittorio Cotesta, *Millennials. Avere vent'anni a Latina.*
 - 1.5. Claudio Bezzi, *Manuale di ricerca valutativa.*

1.6 Luca Salmieri (a cura di), *Servizi sociali e misure di contrasto alla povertà*.

1.7 Gianmaria Bottoni, *Introduzione all'analisi multilivello. Teoria e pratica dei modelli di regressione multilivello*.

1.8 Michele Santurro, *La coesione sociale in Europa. Un'analisi ecologica e diacronica*.

1.9. Orazio Giancola, Luca Salmieri, *Disuguaglianze educative e scelte scolastiche*.

1.10 Giovanni Di Franco (a cura di), *Disuguaglianze intergenerazionali in Italia*.

1.11 Angela Delli Paoli, *La netnografia nella ricerca sociale*.

1.12 Giovanni Di Franco, *L'analisi dei dati nelle scienze sociali al tempo dei Big Data, del Machine Learning e dell'intelligenza artificiale*.

Volumi pubblicati in Open Access

1. Giovanni Di Franco (a cura di), *Giovani a tempo indeterminato. Valori e atteggiamenti dei giovani romani*.

2. Orazio Giancola, Luca Salmieri (a cura di), *Le disuguaglianze educative come processo. Risultati, scelte e aspettative*.

Indice

Prefazione , di <i>Loredana Cerbara, Claudia Pennacchiotti, Ilaria Primerano e Luciana Taddei</i>	pag.	11
Introduzione , di <i>Loredana Cerbara, Dario Germani e Rocco Paolillo</i>	»	13
Riferimenti bibliografici	»	18
1. Perché un <i>Online Probability Panel</i> in Italia? , di <i>Ferruccio Biolcati Rinaldi e Cristiano Vezzoni</i>	»	19
1.1 Introduzione	»	19
1.2 Le infrastrutture di ricerca <i>survey</i> per le scienze sociali in Europa e in Italia	»	19
1.3 <i>L'Italian Online Probability Panel</i>	»	24
1.4 Implicazioni e conclusioni	»	27
1.5 Riferimenti bibliografici	»	29
2. Gli <i>Online Probability Panel</i>: caratteristiche, potenzialità e sviluppi , di <i>Luciana Taddei, Gabriella D'Ambrosio e Francesco Visconti</i>	»	30
2.1 Introduzione	»	30
2.2 Definizione, vantaggi e limiti degli OPP	»	31
2.3 Struttura e caratteristiche di un OPP	»	32
2.4 Peculiarità specifiche degli OPP	»	38
2.5 La progressiva affermazione degli OPP	»	40
2.6 Prospettive future	»	44
2.7 Riferimenti bibliografici	»	45

3. L'Architettura di IOPP tra sfide scientifiche e operative , di <i>Nicolò Marchesini, Francesco Piacentini, Ilaria Primerano e Luciana Taddei</i>	»	48
3.1 Introduzione	»	48
3.2 Struttura e gestione di IOPP	»	49
3.3 Dal disegno del panel alla somministrazione dell'indagine	»	52
3.4 Riflessioni conclusive	»	56
3.5 Riferimenti bibliografici	»	56
4. La strategia di campionamento , di <i>Ilaria Primerano, Nicolò Marchesini e Michele Santurro</i>	»	58
4.1 Introduzione	»	58
4.2 Il campionamento nella ricerca sociale	»	59
4.3 La strategia di campionamento per IOPP	»	64
4.4 Riflessioni conclusive	»	67
4.5 Riferimenti bibliografici	»	68
5. Reclutamento e Mantenimento , di <i>Francesco Piacentini, Claudia Pennacchiotti e Michele Santurro</i>	»	70
5.1 Introduzione	»	70
5.2 Il reclutamento nei panel probabilistici	»	71
5.3 Strategie per l'inclusione degli <i>offliners</i>	»	73
5.4 Incentivi e motivazione alla partecipazione	»	75
5.5 Fidelizzazione e strategie <i>anti-attrition</i>	»	77
5.6 Etica, trasparenza e sostenibilità	»	78
5.7 Riflessioni conclusive	»	79
5.8 Riferimenti bibliografici	»	80
6. Core Questionnaire , di <i>Michele Santurro, Angela Paparusso, Frank Heins, Caterina Ambrosio e Valentina Tudisca</i>	»	82
6.1 Introduzione	»	82
6.2 Come abbiamo progettato il <i>Core Questionnaire</i> di IOPP	»	84
6.3 La struttura attuale del <i>Core Questionnaire</i>	»	86
6.4 Le principali criticità incontrate e le sfide future da affrontare	»	90
6.5 Conclusioni	»	93
6.6 Riferimenti bibliografici	»	94

7. Archiviazione e disseminazione , di <i>Domingo Scisci, Giovanna De Santis e Francesco Piacentini</i>	»	96
7.1 Il ciclo di vita dei dati	»	96
7.2 Gli archivi dati per le scienze sociali	»	97
7.3 La <i>Data Curation</i>	»	99
7.4 La <i>Data Curation</i> e il paradigma <i>Data as Code</i>	»	107
7.5 Conclusioni	»	108
7.6 Riferimenti bibliografici	»	109
8. Verso un approccio Integrato alla Sostenibilità per le Infrastrutture di Ricerca nelle Scienze Sociali: il caso dell'<i>Italian Online Probability Panel (IOPP)</i> nell'infrastruttura FOSSR , di <i>Claudia Pennacchiotti, Loredana Cerbara, Ferruccio Biolcati Rinaldi e Cristiano Vezioni</i>	»	112
8.1 Introduzione	»	112
8.2 Per una definizione di sostenibilità delle Infrastrutture di Ricerca nelle scienze sociali	»	114
8.3 Il caso dell' <i>Italian Online Probability Panel</i> nel contesto dell'infrastruttura di ricerca FOSSR	»	120
8.4 Conclusioni	»	131
8.5 Riferimenti bibliografici	»	133
Conclusioni , di <i>Mario Paolucci</i>	»	135
Riferimenti bibliografici	»	137
Gli autori	»	139

Prefazione

di *Loredana Cerbara, Claudia Pennacchiotti, Ilaria Primerano e Luciana Taddei**

Questo libro è rivolto a quanti tra studenti e studentesse o ricercatori e ricercatrici nelle scienze sociali, ma anche portatori di interesse in ambito politico e sociale vogliono conoscere il percorso metodologico e strategico che ha condotto alla realizzazione dell'*Italian Online Probability Panel* (IOPP) e comprenderne appieno il potenziale derivante dall'adesione ai principi di Scienza Aperta a cui si ispira e in cui opera. La costruzione di IOPP ha comportato la gestione di numerose sfide metodologiche, tecniche ed etiche, tipiche dell'innovazione digitale applicata alla ricerca sociale. L'utilizzo di soluzioni tecnologiche di ultima generazione, e l'ideazione di una architettura digitale improntata alla gestione di strumenti di comunicazione efficienti a disposizione dei panelisti permettono la produzione di dati di elevata qualità in tempi estremamente brevi rispetto agli usuali standard di raccolta dati.

Il libro, dunque, non si limita a fornire una descrizione tecnica del progetto IOPP, ma stimola anche una riflessione critica sulle implicazioni etiche e sociali dell'uso di strumenti digitali e pratiche nella ricerca. La possibilità di accedere a dati di alta qualità in tempo reale rappresenta una risorsa preziosa per i *policy maker*, la comunità scientifica e la società civile, facilitando l'identificazione tempestiva di tendenze sociali e la risposta a sfide emergenti, promuovendo al

* Il presente capitolo è il frutto di una riflessione comune da parte di tutte le autrici.

contempo un approccio integrato nella soluzione di problemi complessi. Il progetto IOPP diventa così l'occasione per aprire una riflessione più ampia sul futuro della ricerca sociale, nella convinzione che l'apertura metodologica possa contribuire alla democratizzazione della conoscenza e a una partecipazione più diffusa alla produzione scientifica.

Introduzione

di *Loredana Cerbara, Dario Germani e Rocco Paolillo**

È ormai condivisa, nella comunità scientifica internazionale, l'idea che una scienza permeabile a diversi approcci ed aperta e trasparente nella possibilità di scambio di risultati e dati abbia maggiori possibilità di progredire e mantenere il passo con l'evoluzione della società moderna. Quest'orientamento si è affermato negli ultimi anni anche nelle scienze umane e sociali grazie alla standardizzazione dei principi della Scienza Aperta (Wilkinson *et al.* 2016; Piwowar e Vision 2013) ed alla crescente diffusione di dati panel e longitudinali (Groves 2004). L'accesso libero ai dati e alle pubblicazioni, la riproducibilità delle ricerche (Munafò *et al.* 2017) e l'adozione di infrastrutture condivise rappresentano elementi essenziali per garantire una maggiore trasparenza, affidabilità e comparabilità dei risultati, in linea con i principi FAIR, *Findable, Accessible, Interoperable, and Reusable* (Wilkinson *et al.* 2016) e con l'impegno verso una cultura della trasparenza (Nosek *et al.* 2015).

In questo contesto, il libro si propone di esplorare il ruolo cruciale delle infrastrutture di ricerca sociale, descrivendo, con riferimento a esperienze concrete di respiro nazionale e internazionale, l'importanza della raccolta e dell'archiviazione di dati longitudinali e di panel (Menard 2005), quali strumenti indispensabili per l'analisi delle dinamiche sociali complesse.

Nell'ambito del progetto FOSSR (*Fostering Open Science in Social Science Research*, descrizione reperibile alla pagina

* Il presente capitolo è il frutto di una riflessione comune da parte dell'autrice e degli autori.

<https://www.fossr.eu>) del Consiglio Nazionale delle Ricerche (CNR) finanziato con fondi PNRR (Piano Nazionale di Ripresa e Resilienza approvato nel 2021 dall'Italia per rilanciare l'economia dopo la pandemia di COVID), grazie alla collaborazione attivata tra CNR e Università degli Studi di Milano La Statale, è nato l'*Italian Online Probability Panel* (IOPP). Progettato a partire da esperienze già realizzate in ambito internazionale, esso rappresenta un esempio innovativo di panel probabilistico che eleva l'Italia al livello di altre esperienze virtuose nel panorama internazionale.

Grazie all'impiego di tecnologie digitali avanzate, IOPP consente il monitoraggio in tempo reale dei fenomeni sociali, garantendo elevati standard qualitativi e la rappresentatività dei dati raccolti. Le scelte in fase di progettazione iscrivono il disegno dell'indagine nell'ambito dei campioni probabilistici (Cicchietelli *et al.* 1992; Di Franco 2010), con una struttura complessa che si basa su un'estrazione casuale di nominativi dall'Anagrafe Nazionale della Popolazione Residente (ANPR).

L'applicazione di sistemi informativi avanzati per la costruzione del primo panel probabilistico italiano ha comportato anche l'implementazione di protocolli rigorosi per la tutela della privacy e la sicurezza dei dati, l'adesione alle normative nazionali ed europee e ai loro aggiornamenti recenti in materia di protezione dei dati e libera circolazione e alle nuove regole deontologiche per i trattamenti personali di dati a fini di ricerca scientifica e statistica. Sono pertanto state adottate le misure necessarie in conformità alle indicazioni del Garante per la protezione dei dati personali, che è stato consultato e che ha espresso un parere favorevole relativamente alle scelte effettuate ritenendo adeguate le misure di sicurezza tecnica previste per IOPP.

Ugualmente, la restituzione delle informazioni raccolte alla comunità scientifica e alla società civile è basata su procedure di anonimizzazione avanzate e sul rispetto di standard etici elevati. Una serie di elementi fondamentali per bilanciare le esigenze di progresso scientifico, sempre più dipendente dalla disponibilità di informazioni di alta qualità, e con l'assicurazione di una adeguata protezione dei diritti dei partecipanti. Queste misure contribuiscono all'instaurazione di

un clima di fiducia tra ricercatori e cittadini, indispensabile per il successo delle iniziative di Scienza Aperta in cui IOPP si muove.

Il paradigma della Scienza Aperta si configura, inoltre, come un catalizzatore per l'innovazione e la collaborazione interdisciplinare, promuovendo la condivisione di conoscenze e risorse tra diverse comunità scientifiche. Tale approccio, mentre contribuisce a ridurre i *bias* metodologici e a migliorare la replicabilità degli studi, comporta una maggiore trasparenza e responsabilità nella produzione del sapere (Munafò *et al.* 2017). L'esperienza italiana di IOPP in questo senso può fornire un modello replicabile e può rappresentare un punto di riferimento per altri Paesi.

Per questo, a partire dall'esperienza di IOPP il volume offre anche una riflessione sulle prospettive future della ricerca sociale, sottolineando come l'adozione di metodologie aperte possa contribuire alla democratizzazione del sapere e ad una più ampia partecipazione della società alla produzione scientifica (Nosek *et al.* 2015). Come si vedrà nei capitoli che seguono, in questo scenario, il modello proposto da IOPP si configura come un esempio per una ricerca sociale più aperta, inclusiva e innovativa.

La vocazione di IOPP alla Scienza Aperta è anche dimostrata dalla possibilità di confrontarsi con altre realtà di respiro internazionale indotta dall'opportunità di creare un *Life Course Observatory* (LCO), un osservatorio composto da tre importanti indagini: GUIDE (*Growing Up in Digital Europe*), che rappresenta il primo studio longitudinale e comparativo finalizzato a fornire evidenze statistiche a supporto di politiche sociali per il benessere dei bambini e delle bambine, dei giovani e delle giovani fino ai 24 anni e delle loro famiglie, GGS (*Generations and Gender Survey* nell'ambito dell'*Infrastructure Generation and Gender Programme*) che si configura come uno studio internazionale che ha come obiettivo la raccolta di dati sulle dinamiche familiari, le relazioni tra generazioni e la struttura della popolazione, e SHARE (*Survey on Health, Ageing and Retirement in Europe*) che è un'indagine inserita nell'*European Research Infrastructure Consortium* (ERIC) e ha l'obiettivo di raccogliere dati sulla salute, lo stato socio-economico, le relazioni sociali della popolazione di età maggiore o uguale a 50 anni.

IOPP beneficia inoltre di altre infrastrutture collegate all'interno del LCO, che hanno il potenziale di includere metodi innovativi per la ricerca sociale, come la possibilità di sintetizzare la conoscenza delle relazioni tra i dati mediante ontologie semantiche, o altri metodi computazionali come la modellizzazione di popolazioni sintetiche per la simulazione di fenomeni collettivi. Si tratta di possibilità di sfruttare IOPP in ampi studi interdisciplinari per la conduzione di ricerca di base o applicata nelle scienze sociali.

Il primo capitolo, a cura di Ferruccio Biolcati Rinaldi e Cristiano Vezzoni, si concentra sull'importanza di costruire anche per l'Italia, come già accade in numerosi contesti internazionali, una infrastruttura per le scienze sociali dotata di un panel probabilistico online, in grado di consentire molteplici applicazioni e di rispondere ad un bisogno informativo tuttora non soddisfatto.

Il secondo capitolo, a firma di Luciana Taddei, Gabriella D'Ambrosio e Francesco Visconti, introduce in via teorica, le caratteristiche degli *Online Probability Panels* (OPP) addentrandosi nella descrizione delle tecniche più conosciute per la loro realizzazione e nella elencazione di potenzialità e limiti.

Al terzo capitolo, a cura di Nicolò Marchesini, Francesco Piacentini, Ilaria Primerano e Luciana Taddei è affidata la descrizione del primo panel probabilistico online tutto italiano (IOPP), raccontando le scelte che hanno contribuito alla sua realizzazione a partire dalla progettazione e fino alla strutturazione di un sistema di mantenimento nel tempo e di utilizzo da parte di ricercatori e studiosi dell'ambito sociale.

La progettazione del disegno dell'indagine è descritta nel dettaglio nel quarto capitolo, a cura di Ilaria Primerano, Nicolò Marchesini e Michele Santurro. Qui viene descritta la struttura teorica che ha guidato le scelte metodologiche che devono garantire elevata qualità dei dati rilevati e affidabilità delle stime ottenute.

Il quinto capitolo, a firma di Francesco Piacentini, Claudia Pennacchiotti e Michele Santurro affronta invece la questione delle modalità di reclutamento, particolarmente rilevante ai fini della mitigazione di fattori di auto-selezione e in generale di limitazione della qualità del panel, nonché quello del mantenimento nel tempo, cioè

della limitazione della *panel attrition* attraverso strumenti che incentivano la fidelizzazione.

Il sesto capitolo, a cura di Angela Paparusso, Frank Heins, Michele Santurro, Caterina Ambrosio e Valentina Tudisca, è dedicato alla progettazione del *core questionnaire* di IOPP. Il contributo illustra le scelte metodologiche che hanno guidato la selezione e l'organizzazione degli *item*, il confronto con le principali esperienze europee e la struttura modulare dello strumento, concepito come elemento centrale per garantire continuità, comparabilità e qualità dei dati in una prospettiva longitudinale.

Il settimo capitolo, a cura di Domingo Scisci, Giovanna De Santis e Francesco Piacentini, descrive le scelte in materia di archiviazione dei dati, comprendendo la *data curation* ma anche la progettazione e implementazione di una struttura di archiviazione in grado di includere tutte le informazioni necessarie per agevolare l'utilizzo dei dati anche nel tempo, con il rispetto di standard condivisi che consentono la disponibilità secondo i principi FAIR.

Infine, l'ottavo capitolo, a firma di Claudia Pennacchiotti, Loredana Cerbara, Ferruccio Biolcati Rinaldi e Cristiano Vezzoni, si inoltra nella riflessione sugli elementi necessari per garantire la sostenibilità nel tempo. Si utilizza come strumento operativo una comparazione con altre esperienze similari facendo infine uno sforzo costruttivo per progettare il mantenimento nel tempo di standard qualitativi di tenore elevato in grado di portare questa esperienza italiana nella sfera della confrontabilità con le altre esperienze internazionali.

Riferimenti bibliografici

- Cicchietelli G., Herzel A., Montanari G. E., 1992, *Il campionamento statistico*, Bologna, Il Mulino.
- Di Franco G., 2010, *Il campionamento nelle scienze umane. Teoria e pratica*, Milano, FrancoAngeli.
- Menard S., 2005, *Longitudinal Studies, Panel*, in Encyclopedia of Social Measurement, 601-607. <https://doi.org/10.1016/B0-12-369398-5/00007-4>.
- Groves R. M., 2004, *Survey Methodology*, New York, Wiley.
- Munafò M. R., Nosek B. A., Bishop D. V. M., et al., 2017, *A Manifesto for Reproducible Science*, in Nature Human Behaviour, 1, 0021.

Nosek B. A., Alter G., Banks G. C., *et al.*, 2015, *Promoting an Open Research Culture*, in *Science*, 348(6242), 1422-1425.

Piwowar H. A., Vision T. J., 2013, *Data Reuse and the Open Data Citation Advantage*, in *PeerJ*, 1:e175. <https://doi.org/10.7717/peerj.175>.

Wilkinson M. D., Dumontier M., Aalbersberg I. J., *et al.*, 2016, *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, in *Scientific Data*, 3, 160018.

1. Perché un *online probability panel* in Italia?

di Ferruccio Biolcati Rinaldi e Cristiano Vezzoni¹

1.1 Introduzione

Questo capitolo argomenta l'importanza della realizzazione di un *Online Probability Panel* per il rafforzamento delle infrastrutture di ricerca nelle scienze sociali in Italia. Viene dapprima delineato lo scenario delle infrastrutture di ricerca in Europa, per poi descrivere il caso italiano e la lacuna che un *Online Probability Panel* andrebbe a colmare. Successivamente, viene introdotto il contesto in cui il progetto di un panel italiano si è sviluppato, le sfide di lungo periodo e contingenti da affrontare, le opportunità che si sono create. Infine, si evidenzia come tale progetto possa essere funzionale allo sviluppo delle infrastrutture di ricerca e delle scienze sociali più in generale nel nostro Paese.

1.2 Le infrastrutture di ricerca *survey* per le scienze sociali in Europa e in Italia

Per descrivere il panorama delle infrastrutture di ricerca *survey* per le scienze sociali a livello europeo, è utile distinguere tra quelle che hanno carattere nazionale (sia come tipo di organizzazione sia come estensione dei dati raccolti) e quelle con carattere sovra-nazionale, in

¹ Il presente capitolo è il frutto di una riflessione comune da parte degli autori.

cui organizzazioni basate su collaborazioni internazionali raccolgono dati comparati.

Per quanto riguarda le prime, nei diversi paesi si possono osservare esperienze ispirate ad alcuni modelli ricorrenti. Uno dei modelli più comuni si ispira alla *General Social Survey* (GSS), avviata negli Stati Uniti d'America nel 1972, più di cinquant'anni fa (<https://gss.norc.org/>). Questo modello si basa su rilevazioni svolte a scadenze abbastanza ravvicinate (solitamente annuale o biennale) su campioni probabilistici di individui che vengono intervistati su temi sociali e politici, in termini sia di atteggiamenti sia di comportamenti. Ogni rilevazione viene svolta su individui differenti, con l'obiettivo di avere sempre un campione rappresentativo della popolazione che cambia per le principali dinamiche demografiche (natalità, mortalità, immigrazione, emigrazione). In altre parole, si tratta di un disegno di ricerca trasversale ripetuto che consente di cogliere al meglio il cosiddetto cambiamento aggregato, ossia l'andamento dei diversi fenomeni nella popolazione e nei suoi sottogruppi (Duncan e Kalton 1987). Un esempio europeo di questo modello è il progetto ALLBUS (*Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*) che dal 1980 indaga ogni due anni diversi ambiti della società tedesca, come ad esempio: comportamenti devianti e sanzioni, etnocentrismo e minoranze, salute, reti sociali e capitale sociale, orientamenti valoriali, lavoro e occupazione, famiglia, stili di vita e personalità, religione e visione del mondo, disuguaglianza sociale (<https://www.gesis.org/en/allbus/>).

Un secondo modello ricorrente è quello dell'*household panel*. In questo caso l'unità di campionamento è costituita dalle famiglie – e non dagli individui – per cui vengono intervistati tutti i membri di un nucleo familiare su temi che hanno solitamente carattere economico e demografico. Il disegno di ricerca è di tipo panel, per cui gli stessi individui vengono intervistati più volte attraverso il tempo, consentendo così l'analisi del cambiamento a livello individuale. Per tenere conto delle modifiche nella popolazione dovute alle principali dinamiche demografiche, il campione viene periodicamente aggiornato inserendo nuove famiglie. Anche qui abbiamo un modello americano che è quello del *Panel Study of Income Dynamics* (PSID) che dal 1968 studia la povertà, le dinamiche reddituali, occupazione, ricchezza, consumi, salute, matrimonio e fertilità, sviluppo infantile,

istruzione, e altri temi ancora (<https://psidonline.isr.umich.edu/>). Un esempio europeo è rappresentato dal progetto britannico *Understanding Society*, avviato nel 2009 raccogliendo l'eredità del *British Household Panel Survey* (BHPS). Tra gli elementi innovativi di questo progetto troviamo la disponibilità di dati genetici per molti dei partecipanti al progetto e la possibilità di collegare i dati di intervista a dati amministrativi e geografici (<https://www.understandingsociety.ac.uk/>).

Il terzo modello è appunto quello dell'*Online Probability Panel* (OPP) che può essere visto come un'ibridazione tra i due modelli precedenti. Il disegno di ricerca è panel mentre l'unità di campionamento è solitamente individuale. I temi indagati hanno carattere maggiormente sociale e politico, anche se un profilo sociodemografico dei partecipanti è sempre rilevato. L'innovazione apportata da questo modello sta nel coniugare il disegno di campionamento probabilistico con la modalità di raccolta dati online – o meglio *Computer-Assisted Web Interviewing* (CAWI) – che dalla sua apparizione sulla scena della ricerca *survey* era stata sempre associata a disegni campionari di convenienza (Callegaro *et al.* 2015). Per una trattazione più approfondita degli OPP si rimanda al capitolo successivo.

Passando alle infrastrutture di ricerca *survey* basate su collaborazioni internazionali che raccolgono dati comparati, è opportuno distinguere tra quelle in cui i processi di armonizzazione (codifica, campionamento, etc.) sono precedenti alla raccolta dei dati (*ex ante*) e quelli in cui sono – per quel che è possibile – successivi (*ex post*). Un esempio di armonizzazione *ex post* è il progetto *True European Voter* (TEV) che ha integrato diversi studi elettorali europei (Schmitt 2021).

Per le più diffuse infrastrutture di ricerca – quelle a carattere *ex ante* – si può forse parlare di tre diverse generazioni che si distinguono su due diverse dimensioni, la struttura organizzativa e l'obiettivo.

La prima pionieristica generazione è quella delle infrastrutture nate negli anni Ottanta, come *European Values Study* (EVS - <https://europeanvaluesstudy.eu/>), *World Values Survey* (WVS - <https://www.worldvaluessurvey.org/wvs.jsp>), *International Social Survey Programme* (ISSP - <https://issp.org/>). L'obiettivo fondamentale di queste infrastrutture è la raccolta dei dati attraverso un coordinamento di tipo *bottom-up* basata sulla collaborazione volontaristica

tra diverse università, istituti di ricerca, organizzazioni di *fieldwork* a livello nazionale.

La generazione successiva sancisce il consolidamento dell'idea di infrastruttura di ricerca nelle scienze sociali. Si tratta di infrastrutture di ricerca, nate all'inizio del XXI secolo o nella decade successiva, che hanno ancora come *core business* la raccolta dei dati ma hanno una organizzazione a carattere maggiormente *top-down*, che adotta o aspira allo status di ERIC (*European Research Infrastructure Consortium*). Questa è infatti una forma giuridica europea che facilita il coordinamento e la gestione congiunta di un'infrastruttura di ricerca internazionale, garantendo riconoscimento legale, semplificazioni amministrative e più facile accesso a fondi governativi ed europei.

I principali esempi sono *European Social Survey* (ESS - <https://www.europeansocialsurvey.org/>), *Survey of Health, Ageing and Retirement in Europe* (SHARE - <https://share-eric.eu/>) e, più recentemente, *Growing Up in Digital Europe* (GUIDE - <https://www.guidecohort.eu/>).

Infine, si inizia a intravedere una terza generazione di infrastrutture di ricerca che aspira sempre a una forma organizzativa di tipo *top-down*, dove però l'obiettivo centrale non è più quello della raccolta dei dati ma piuttosto quello della fornitura di servizi di ricerca a vantaggio della comunità scientifica (ma anche del dibattito pubblico). Un esempio di questa terza generazione è l'infrastruttura *Monitoring Electoral Democracy* (MEDEM - <https://www.medem.eu/>), per la quale si veda il Riquadro 1.

Riquadro 1. Monitoring Electoral Democracy (MEDEM)

(a cura di Francesco Piacentini)

Monitoring Electoral Democracy (MEDEM) è un'infrastruttura di ricerca europea che ambisce a superare la frammentazione del quadro della ricerca e della raccolta dei dati sulla democrazia in Europa, collegando e integrando *ex post* fonti diverse, che comprendono *survey* elettorali, dati testuali dei media, *élite survey* e statistiche istituzionali. MEDEM punta a favorire l'interconnessione e l'armonizzazione dei dati e delle strategie di ricerca, superando i limiti dei singoli gruppi e progetti di ricerca. L'infrastruttura, riconosciuta dall'ESFRI *Landscape Analysis 2024* e candidata per l'inclusione nella ESFRI *Roadmap 2026*, si propone di creare una piattaforma unificata per mettere in relazione, armonizzare e standardizzare i dati nel tempo e tra paesi, seguendo i principi FAIR. L'obiettivo è inoltre

quello di garantire l'accessibilità a lungo termine ai dati sulle democrazie. Composta da otto istituzioni partner e con diversi accordi di collaborazione con altre infrastrutture di ricerca, MEDEM intende offrire servizi di armonizzazione dati, formazione, consulenza specialistica e una *dashboard* per ricercatori e giornalisti, coordinando al contempo la comunità scientifica europea che studia la democrazia trasversalmente in diverse discipline, paesi e istituzioni.

Se muoviamo la nostra attenzione al panorama italiano delle infrastrutture di ricerca *survey*, la situazione risulta meno vivace e sviluppata. A livello nazionale, la situazione è diversificata ma, in generale, la capacità di implementare infrastrutture di ricerca risulta limitata, se si eccettuano esperienze quali l'Indagine sui Bilanci delle Famiglie Italiane di Banca d'Italia (<https://www.bancaditalia.it/pubblicazioni/indagine-famiglie/>) o l'*Italian National Election Studies* (ITANES - <https://www.itanes.it/>). Ovviamente prezioso è il contributo offerto da Istat che mette a disposizione diverse indagini (Aspetti della vita quotidiana, Famiglie e soggetti sociali, Rilevazione sulle forze di lavoro, etc.) sebbene nel tempo la facilità di accesso e il livello di apertura non sia sempre stato omogeneo. Al di fuori di questo perimetro, non esistono in Italia infrastrutture di ricerca a carattere nazionale ispirate ai modelli sopra illustrati: *general social survey*, *household panel*, *online probability panel*. Questa è una grave carenza nello scenario nazionale della ricerca sociale che limita le opportunità a disposizione della comunità scientifica.

Passando alle infrastrutture di ricerca *survey* basate su collaborazioni internazionali che raccolgono dati comparati, fino al 2016 la partecipazione dell'Italia è stata caratterizzata da una sostanziale discontinuità. Un progetto di ricerca longitudinale e comparato (basato su procedure di armonizzazione *ex post*) ci offre la possibilità di quantificare il *gap* nella partecipazione a tali infrastrutture tra l'Italia e altri paesi europei di analoga dimensione (Biolcati *et al.* 2022). La tabella 1.1 riporta il numero di occasioni in cui i quattro più grandi paesi europei hanno partecipato a quattro tra le più rilevanti infrastrutture di ricerca *survey* tra il 1973 e il 2016. Germania e Regno Unito sono i paesi col maggior numero di partecipazioni (rispettivamente 45 e 42) seguiti dalla Francia (31) e dall'Italia (18).

Tab. 1.1 – Partecipazione di Francia, Germania, Italia e Regno Unito a infrastrutture di ricerca survey (Eurobarometer - EB, European Social Survey – ESS, International Social Survey Programme - ISSP, European Values Study - EVS, World Values Survey – WVS) nel periodo 1973-2016 (adattata da Biolcati et al. 2022).

	ESS	ISSP	EVS	WVS	Totale
Francia	7	19	4	1	31
Germania	7	31	4	3	45
Italia	3	10	4	1	18
Regno Unito	7	30	4	1	42

Il gap tra l'Italia e i paesi con maggiore partecipazione è rilevante: per sintetizzare, è possibile affermare che l'Italia ha partecipato meno della metà delle volte che hanno partecipato Paesi come la Germania o il Regno Unito. C'è tuttavia da rilevare che dal 2016, la situazione ha cominciato a migliorare, grazie agli sforzi fatti da diverse istituzioni per garantire la partecipazione italiana alle grandi infrastrutture di ricerca comparata della ricerca sociale: ESS è stato garantito dal sostegno di INAPP – <https://www.inapp.gov.it/rilevazioni/rilevazioni-periodiche/european-social-survey-ess>, mentre ISSP, EVS e WVS sono stati sostenuti dal Dipartimento di Scienze sociali e politiche dell'Università degli Studi di Milano tramite il laboratorio spsTREND - <https://www.spsrend.it/>.

1.3 L'Italian Online Probability Panel

La proposta di costruire anche in Italia un *Online Probability Panel*, l'*Italian Online Probability Panel* (d'ora in poi IOPP), nasce in questo contesto e cerca di rispondere a diverse sfide poste dalle rapide trasformazioni in atto sia nella ricerca che nella società nel suo complesso. Da una parte, ci sono sfide a lungo termine e specifiche del campo della ricerca *survey*: si fa qui riferimento in particolare alla somministrazione del questionario con modalità faccia-a-faccia che presenta tassi di risposta in tendenziale declino e costi in tendenziale crescita. Dall'altra parte, ci sono sfide più generali che si sono manifestate in maniera istantanea e inattesa. Un esempio molto chiaro è emerso in occasione della pandemia da COVID-19 che ha avuto rilevanti conseguenze negative per moltissimi settori economici e in par-

ticolare per quelli basati su un lavoro di relazione come è appunto la somministrazione del questionario con modalità faccia-a-faccia.

In questo quadro, è chiaro che gli strumenti della ricerca sociale debbano essere in grado di rispondere a cambiamenti che sono spesso imprevisti e diventano sempre più rapidi e trasversali – si pensi appunto alle crisi pandemiche (Vezzoni *et al.* 2020) ma anche all’affermazione dirompente dell’intelligenza artificiale nel giro di pochissimi anni e al suo impatto trasversale su tutti gli ambiti della vita, tanto individuale quanto associata. In questo contesto, il fondo straordinario di investimenti *Next Generation EU*, che nel caso italiano è stato declinato nel Piano Nazionale di Ripresa e Resilienza (PNRR), ha offerto enormi opportunità in molti settori, tra cui quello specifico delle scienze sociali. In particolare, il PNRR ha fornito un ammontare di risorse mai visto prima per la costruzione di infrastrutture di ricerca e opportunità di collaborazione tra le istituzioni accademiche e il Consiglio Nazionale delle Ricerche (CNR) che svolge un ruolo fondamentale nello scenario delle infrastrutture di ricerca nei diversi campi scientifici. Inoltre, uno degli effetti più generali della pandemia è stata la forte spinta alla diffusione delle modalità online di lavoro e studio (Naumann *et al.* 2025) che ha reso il *field-work* delle ricerche *survey* sempre più adatto anche in Italia per una applicazione estensiva del CAWI, ovvero la rilevazione attraverso questionari autosomministrati online.

In questa miscela di sfide e opportunità, IOPP e il più ampio progetto *Fostering Open Science in Social Science Research* (FOSSR) in cui esso è inserito (<https://www.fossr.eu/>), si pongono come veri e propri *game changer* sia per lo scenario delle infrastrutture per le scienze sociali in Italia sia per lo sviluppo delle attività scientifiche post COVID-19. Diversi sono gli obiettivi dell’infrastruttura IOPP.

1) Il primo obiettivo di IOPP è il monitoraggio del cambiamento sociale attraverso la raccolta sistematica di dati basati su campioni probabilistici. Grazie alla struttura panel, IOPP consente di osservare dinamiche individuali e collettive nel tempo, offrendo strumenti analitici particolarmente adatti allo studio delle trasformazioni sociali, politiche ed economiche che attraversano la società italiana. L’uso di procedure di campionamento rigorose e di standard elevati di qualità garantisce l’affidabilità delle stime e la loro rilevanza scientifica.

2) IOPP è progettato per fornire un flusso continuo di dati destinati alla ricerca secondaria e alle attività di insegnamento. La disponibilità regolare di dati aggiornati e documentati rappresenta una risorsa fondamentale per la comunità accademica, facilitando analisi comparative, replicabilità dei risultati e uso didattico nei corsi universitari di metodologia, statistica e scienze sociali. Questi dati vengono forniti seguendo i principi FAIR che rispondono all'idea di fondo di promozione dell'*Open Science* (Commissione Europea 2023), che ha informato tutto il lavoro svolto nel contesto del progetto FOSSR.

3) Un ulteriore obiettivo centrale è quello di fungere da *benchmark* metodologico per la ricerca *survey* nel contesto nazionale. IOPP si propone come riferimento per la valutazione della qualità dei dati, la sperimentazione di innovazioni metodologiche e il confronto tra diverse modalità di raccolta delle informazioni. In questo senso, il progetto contribuisce al rafforzamento degli standard della ricerca empirica in Italia.

4) IOPP svolge inoltre una funzione formativa, offrendo opportunità di *training* per giovani ricercatori interessati alla ricerca *survey* di alta qualità. Attraverso il coinvolgimento diretto nelle fasi di progettazione, raccolta e analisi dei dati, il panel favorisce lo sviluppo di competenze metodologiche avanzate e la diffusione di buone pratiche di ricerca.

5) Infine, il progetto mette a disposizione della comunità scientifica un'infrastruttura che consente la realizzazione di indagini originali secondo standard elevati di qualità e trasparenza, ampliando le possibilità di raccolta dati per studi competitivi a livello nazionale e internazionale.

Rispetto a quest'ultimo punto, è opportuno chiedersi se il mercato della ricerca sociale in Italia esprima una effettiva domanda per una infrastruttura come IOPP. A nostro avviso, vi sono almeno tre fonti che possono alimentare la domanda per infrastrutture come IOPP. Innanzitutto, vi sono progetti di ricerca avanzata che richiedono dati in grado di soddisfare i più elevati standard di qualità ma che possono fare affidamento su *budget* limitati. Ci riferiamo qui soprattutto ai progetti PRIN (Progetti di Rilevante Interesse Nazionale). Il *budget* messo a disposizione di questi ultimi si è negli ultimi bandi stabilizzato intorno ai 200 mila euro: tenendo conto che è sempre necessario

investire una parte rilevante del *budget* nelle risorse umane necessarie per lo sviluppo del progetto, rimangono poche risorse per una raccolta dati *ex novo*, che potrebbero essere invece sufficienti nel contesto di una infrastruttura come IOPP. In questa ottica però non bisogna escludere l'interesse anche di progetti finanziati su fondi ERC (*European Research Council*) o FIS (Fondo Italiano per la Scienza) interessati a studiare il nostro paese. Questi progetti, nonostante *budget* più generosi, a volte faticano a trovare modalità affidabili per la raccolta di dati *survey*. In secondo luogo, vi sono progetti di ricerca anche innovativi che sono però privi di *budget*: si pensi ai progetti legati alle tesi di dottorato per i quali la raccolta di una o poche domande (che si vanno a sommare alle centinaia di variabili già disponibili in IOPP) o un semplice *survey experiment* potrebbero essere sufficienti per sviluppare un intero progetto di ricerca. IOPP mira, infatti, a riservare uno spazio in ciascuna *wave* per domande che provengono dai progetti di giovani ricercatori e ricercatrici che verranno selezionati sulla base di una procedura competitiva. Infine, IOPP offre l'opportunità di costruire sinergie con infrastrutture di ricerca consolidate che potrebbero optare per un passaggio dalla classica modalità di somministrazione faccia-a-faccia a soluzioni alternative quali quelle online. Infatti, sempre più spesso infrastrutture come ISSP o EVS aprono alla possibilità di raccogliere dati sfruttando panel probabilistici esistenti: tale possibilità viene concessa a patto che l'*Online Probability Panel* soddisfi i più alti standard di qualità rispetto al reclutamento, mantenimento e gestione del panel.

1.4 Implicazioni e conclusioni

Si vuole infine chiudere con una considerazione più generale. Il progetto IOPP, per quanto ambizioso e metodologicamente solido, non sarà in grado di incidere sullo scenario infrastrutturale sopra descritto se non sarà accompagnato dalla capacità di condividere e consolidare una vera e propria "cultura infrastrutturale" all'interno della comunità scientifica nazionale. Tale cultura non riguarda un singolo progetto, ma costituisce una condizione necessaria per il funzionamento e la sostenibilità di tutte le infrastrutture di ricerca nelle scienze sociali.

Un primo elemento cruciale riguarda la definizione di assetti istituzionali in grado di garantire un flusso costante di risorse economiche. Le infrastrutture di ricerca richiedono investimenti continuativi e prevedibili, incompatibili con una logica esclusivamente progettuale o basata su finanziamenti a breve termine. Senza meccanismi stabili di finanziamento, anche le infrastrutture più avanzate rischiano di rimanere fragili o discontinue.

Accanto alle risorse economiche, è indispensabile investire nelle risorse umane, attraverso la costruzione di strutture permanenti capaci di assicurare continuità operativa, competenze specialistiche e memoria istituzionale. Le infrastrutture di ricerca non possono poggiare esclusivamente su lavoro temporaneo o su contributi individuali non strutturati, ma necessitano di personale dedicato e professionalizzato.

Un ulteriore aspetto riguarda il riconoscimento del lavoro svolto per le infrastrutture di ricerca. La progettazione, la gestione e il mantenimento di strumenti come IOPP producono valore scientifico collettivo, ma questo contributo che deriva dall'attività di una moltitudine di ricercatrici e ricercatori è spesso sottovalutato nei sistemi di valutazione accademica. Promuovere una cultura infrastrutturale implica riconoscere tali attività come parte integrante del lavoro di ricerca.

Fondamentale è inoltre il rafforzamento della collaborazione istituzionale. Le infrastrutture di ricerca funzionano efficacemente solo se sostenute da una rete cooperativa di università, enti di ricerca e istituzioni, capace di superare frammentazioni e competizioni sterili.

Le condizioni sopra citate non riguardano soltanto IOPP, ma rappresentano un presupposto fondamentale per il rafforzamento complessivo delle infrastrutture di ricerca nelle scienze sociali e, più in generale, per la loro capacità di produrre conoscenza di alta qualità nel lungo periodo.

Attraverso questo lavoro si auspica lo sviluppo di una cultura infrastrutturale che potrà contribuire in modo decisivo alla costruzione di una comunità nazionale della ricerca più coesa, fondata su standard comuni, fiducia reciproca e obiettivi condivisi.

1.5 Riferimenti bibliografici

Biolcati Rinaldi F., Molteni F., Quandt M., Vezzoni C., 2022, *Church Attendance and Religious change Pooled European dataset (CARPE): a survey harmonization project for the comparative analysis of long-term trends in individual religiosity*, in *Quality & Quantity*, v. 56, 1729–1753, ISSN 0033-5177, DOI 10.1007/s11135-020-01048-9.

Callegaro M., Manfreda K., Vehovar V., 2015, *Web survey methodology*, Londra, SAGE Publications Ltd, <https://doi.org/10.4135/9781529799651>.

Commissione Europea, 2023, *European Charter for Access to Research Infrastructures: Principles and guidelines for access and related services*, Publications Office of the European Union. <https://doi.org/10.2777/353942>.

Duncan G. J., Kalton G., 1987, *Issues of Design and Analysis of Surveys across Time*, in *International Statistical Review*, Vol. 55, No. 1, 97-117, DOI <https://doi.org/10.2307/1403273>.

Naumann E., Biolcati Rinaldi F., Nandi A., Recchi E., (a c. di), 2025, *Social Life during COVID-19 in France, Germany, Italy and the UK*, Palgrave Macmillan, ISBN 9783031993480, DOI 10.1007/978-3-031-99349-7.

Schmitt H., 2021, *The True European Voter* (ZA5054; Version 1.0.0) [Data set], GESIS, Cologne. <https://doi.org/10.4232/1.13601>.

Vezzoni C., Ladini R., Molteni F. B., Dotti Sani G. M., Biolcati Rinaldi F., Chiesi A. M., Maraffi M., Guglielmi S., Pedrazzani A., Segatti P., 2020, *Investigating the social, economic and political consequences of Covid-19: A rolling cross-section approach*, in *Survey Research Methods*, v. 14, n. 2, 187-194, ISSN 1864-3361, DOI: 10.18148/srm/2020.v14i2.7745.

2. Gli *Online Probability Panel*: caratteristiche, potenzialità e sviluppi

di *Luciana Taddei, Gabriella D'Ambrosio e Francesco Visconti**

2.1 Introduzione

Gli *Online Probability Panels* (d'ora in poi OPP) sono strumenti di ricerca sociale che hanno assunto un'importanza e diffusione crescente nel corso degli ultimi anni grazie alla loro capacità di combinare campionamento probabilistico e raccolta di dati digitale, integrando la metodologia di ricerca tradizionale alla ricerca online. Questa combinazione rende, dunque, gli OPP molto più vantaggiosi rispetto ad altre tecniche di raccolta dati (interviste faccia-a-faccia, sondaggi telefonici, questionari postali, etc.) perché maggiormente efficienti in termini di tempi e costi, oltre a mantenere elevati standard di qualità e rappresentatività del campione. La loro struttura longitudinale consente, inoltre, di monitorare cambiamenti a livello individuale di lungo periodo, rendendoli particolarmente adatti allo studio delle trasformazioni sociali, del comportamento politico, delle dinamiche economiche, di opinioni, atteggiamenti e valori di una determinata società.

In questa prospettiva, questo capitolo presenta le caratteristiche principali degli OPP (struttura, funzionamento, vantaggi e limiti), of-

* Il presente capitolo è il frutto di una riflessione comune da parte delle autrici e dell'autore. In ogni caso, il paragrafo 2.2 e il sottoparagrafo 2.3.2 sono da attribuire a Gabriella D'Ambrosio, il paragrafo 2.4 e 2.3 con il sottoparagrafo 2.3.3 sono da attribuire a Francesco Visconti, il paragrafo 2.5 e il sottoparagrafo 2.3.1 sono da attribuire a Luciana Taddei. I paragrafi 2.1 e 2.6 sono stati scritti congiuntamente dalle autrici e dall'autore.

frendo altresì una panoramica dei principali OPP già implementati a livello internazionale e, in particolare, europeo.

2.2 Definizione, vantaggi e limiti degli OPP

Le indagini campionarie costituiscono uno strumento essenziale per la ricerca nelle scienze sociali, in quanto offrono un approccio sistematico alla raccolta di dati su ampie popolazioni, consentendo l'analisi di una vasta gamma di temi, quali struttura familiare, condizioni abitative, situazione lavorativa ed economica, diseguaglianze e vulnerabilità sociale, competenze digitali, stili di vita, comportamenti, opinioni, valori e atteggiamenti dei cittadini *et alii*. La possibilità di integrare tali dati con fonti esterne ne accresce ulteriormente la versatilità, rendendole adatte a esplorare molteplici domande di ricerca. Inoltre, la ricerca sociale può affrontare questi temi adottando un approccio trasversale, che consente di descrivere – ed eventualmente comparare – le caratteristiche di una determinata popolazione in un determinato momento di tempo, oppure adottando un approccio longitudinale, che permette di analizzare l'evoluzione dei fenomeni, individuando processi di cambiamento e traiettorie individuali o collettive.

A differenza delle indagini trasversali, infatti, le indagini longitudinali riconoscono esplicitamente che le popolazioni sono dinamiche (Lynn 2012). Riprendendo la distinzione proposta da Galtung tra ricerca statica e ricerca dinamica (1967), si indica con “trasversale” qualsiasi ricerca effettuata sulla popolazione di riferimento in un preciso periodo di tempo, mentre l'attributo “longitudinale” fa riferimento a ricerche che indagano il cambiamento di una popolazione in periodi più lunghi esplorandone le caratteristiche nel tempo (Menard 2002). Gli studi diacronici o longitudinali «sono dunque specificatamente destinati allo studio del mutamento sociale, per descriverlo, spiegarlo o prevederne lo stato futuro» (Agnoli 2008, 112).

A partire dalla prospettiva di ricerca adottata, è possibile distinguere ulteriormente tra diverse tipologie di studi longitudinali, da applicare a seconda della domanda cognitiva sottostante la ricerca, di seguito riportati (Ruspini 2002):

- studi di *trend*, che analizzano l'evoluzione di una variabile su campioni distinti nel tempo;
- studi di *follow-up*, che replicano uno studio o rilevazione precedente;
- studi di *panel*: studi sull'evoluzione di un fenomeno sociale effettuati sulla stessa popolazione a intervalli più o meno regolari di tempo.

A differenza delle prime due tipologie, gli studi di panel si caratterizzano per l'interrogazione ricorrente degli stessi soggetti, consentendo l'analisi dei cambiamenti a livello individuale, oltre che aggregato. Al riguardo, la costruzione e la gestione di un panel rappresentano un processo metodologicamente complesso, che implica la definizione di criteri di selezione, l'implementazione di strategie di coinvolgimento e mantenimento della popolazione ingaggiata nel tempo, nonché la predisposizione di strumenti adeguati al monitoraggio della partecipazione e al controllo della mortalità, al fine di garantire la validità longitudinale e la robustezza inferenziale dei dati raccolti.

Nella sezione successiva, esaminiamo in dettaglio le fasi che caratterizzano la costruzione di un panel e, nello specifico, dei panel realizzati online.

2.3 Struttura e caratteristiche di un OPP

Gli OPP seguono generalmente una struttura operativa articolata in fasi sequenziali e ben definite (Callegaro *et al.* 2014), ciascuna delle quali riveste un ruolo essenziale nel garantire la qualità del disegno e la solidità dei dati raccolti. In particolare, le fasi principali includono:

1. *Definizione della popolazione e campionamento*: la fase iniziale prevede l'individuazione della popolazione di riferimento e la progettazione di una strategia di campionamento, generalmente basata su registri anagrafici o elettorali, con possibili stratificazioni e campionamenti a più stadi.
2. *Reclutamento*: una volta creata la lista campionaria e definito il questionario di reclutamento, i soggetti selezionati (individui o

famiglie) vengono invitati a partecipare attraverso diversi canali di contatto (posta, web, faccia-a-faccia). L'invito a partecipare è spesso accompagnato da un incentivo incondizionato, utile ad aumentare il tasso di adesione al panel. È, inoltre, prassi fornire supporti tecnologici (tablet, connessione Internet) per includere chi non ha accesso autonomo al web così migliorando la rappresentatività del campione (Scherpenzeel 2009).

3. *Profilazione e iscrizione al panel da parte dei rispondenti*: chi accetta di partecipare all'intervista di reclutamento, compila un questionario preliminare volto a raccogliere informazioni di base utili per la profilazione. Al termine, viene chiesto all'intervistata/o se desidera effettivamente iscriversi al panel.
4. *Somministrazione periodica delle indagini*: i panelisti e le paneliste iscritti partecipano successivamente a ondate regolari di rilevazione, somministrate generalmente online, con eventuali promemoria via e-mail o SMS, per le quali possono essere previsti ulteriori incentivi condizionati. Per giunta, integrare la tecnica di rilevazione online con una strategia *mixed-mode* può risultare più efficace per raggiungere sottogruppi di popolazione meno inclini alla partecipazione online senza compromettere l'attendibilità dell'indagine (Istat 2017).
5. *Mantenimento del panel*: per garantire la partecipazione nel tempo, vengono adottate strategie di fidelizzazione e monitoraggio il tasso di abbandono, al fine di mantenere nel tempo la rappresentatività del panel e la validità inferenziale delle analisi.

Di seguito presentiamo le tre caratteristiche fondamentali che contraddistinguono un OPP: la tecnica di rilevazione (online), la tipologia di campionamento (probabilistico) e il tipo di disegno adottato (di panel).

2.3.1 La tecnica di rilevazione

Con la crescente digitalizzazione della popolazione, le tecniche *web-based* hanno assunto un ruolo sempre più rilevante nella ricerca sociale, favorendo la diffusione di panel online, ovvero liste di indivi-

dui disponibili a partecipare ripetutamente a ricerche scientifiche condotte via web (Görizt 2007). Sebbene i partecipanti ai panel possano essere reclutati mediante procedure *mixed-mode* (per es. interviste faccia-a-faccia, telefoniche, etc.), è opportuno precisare che tutte le successive interviste, volutamente brevi e strutturate per favorire l'*engagement* dei rispondenti, sono condotte prevalentemente tramite autosomministrazione online (tramite *browser* o *software* dedicato). L'uso delle *web survey* si è progressivamente affermato nella ricerca sociale, rispondendo alle esigenze di efficienza, rapidità e flessibilità nella raccolta dei dati (Blom *et al.* 2016).

Tuttavia, l'efficacia di queste tecniche dipende dalle caratteristiche della popolazione oggetto di studio, che potrebbe incontrare difficoltà di comprensione e compilazione, legate a scarse competenze digitali o mancanza di mezzi adeguati. Numerosi OPP, pur adottando prevalentemente la modalità CAWI (*Computer-Assisted Web Interviewing*), integrano ulteriori tecniche di rilevazione per ridurre il *bias* di copertura causato dal *digital divide*, come questionari cartacei via posta o interviste telefoniche. Queste modalità alternative, sebbene permettano di includere rispondenti meno digitalizzati, se non adeguatamente controllati possono introdurre variazioni sistematiche nei *pattern* di risposta e nella qualità dei dati (Yeager *et al.* 2011).

Numerosi studi hanno evidenziato vantaggi e limiti di questa tecnica di rilevazione, oggi sempre più privilegiata nella ricerca sociale empirica di tipo interrogativo (Istat 2017; Mauceri, Faggiano e Di Censi 2020; Veltri 2021; Taddei e Diana 2024). Tra i vantaggi principali figurano: la riduzione dei costi di gestione, resa possibile dall'eliminazione di personale per la somministrazione (intervistatori) e dalla mancata produzione e distribuzione di materiale cartaceo, la rapidità di somministrazione e la capacità di copertura anche in aree difficilmente raggiungibili.

Dal punto di vista tecnico, le indagini online permettono di realizzare questionari con *routine* complesse, favorendo una raccolta dati mirata e personalizzata. Tuttavia, tale complessità può comportare errori di implementazione, con potenziali effetti distorsivi sui risultati. D'altra parte, se adeguatamente progettata e pre-testata, l'automatizzazione permette una codifica immediata delle risposte, riducendo al minimo gli errori di trascrizione e facilitando l'analisi statistica.

Il formato digitale consente inoltre l'integrazione di elementi multimediali e grafici, anche interattivi, aumentando possibilità e modalità di interazione con gli intervistati e le intervistate. La somministrazione di un questionario digitale consente infatti di raccogliere informazioni in formato immagine, audio, video – in forma sincrona o asincrona – moltiplicando esponenzialmente le possibilità di raccolta e di analisi dati.

La relazione con i rispondenti e le rispondenti può inoltre avvalersi di ulteriori strumenti di comunicazione digitali, tempestivi ed immediati (si pensi, ad esempio, alle *chat* sincrone automatizzate, all'utilizzo di *emoticons* o al monitoraggio delle reazioni).

Infine, un ulteriore vantaggio consiste nella possibilità di raccogliere metadati e paradata (es. tempo di risposta, tipo di dispositivo utilizzato, orario di compilazione), utili a valutare la qualità del processo e migliorare la progettazione di future rilevazioni. Questa possibilità risulta particolarmente utile a gestire i rischi legati alla tipica flessibilità delle indagini online che, permettendo ai rispondenti e alle rispondenti di scegliere liberamente tempi e luoghi di compilazione, può comportare distrazioni o interruzioni nella compilazione. In aggiunta, attraverso l'analisi dei paradata, è possibile intervenire rapidamente per validare o adattare i quesiti, eliminando o modificando domande ridondanti o poco significative. In ultimo, il formato digitale facilita l'implementazione multilingue senza richiedere la presenza di intervistatori o intervistatrici con competenze linguistiche specifiche. Pertanto, la raccolta di questi dati e paradata durante la compilazione del questionario crea nuove possibilità di ricerca, consentendo anche, ad esempio, di valutare l'impatto dei fattori di contesto.

Le indagini online risultano particolarmente efficaci anche per affrontare tematiche delicate, grazie all'assenza dell'intervistatore che potrebbe influenzare le risposte o innescare dinamiche di desiderabilità sociale. Tuttavia, questa modalità limita il controllo sul contesto di compilazione, impedendo al ricercatore o alla ricercatrice di fornire supporto interpretativo, intervenire in caso di calo dell'attenzione, rassicurare sull'anonimato o cogliere segnali comunicativi non verbali o para-verbali.

La garanzia dell'anonimato in particolare costituisce un aspetto cruciale, che richiede misure tecniche e organizzative adeguate, non

sempre agevoli per istituti di ricerca di piccole dimensioni o collaborazioni con partnership estese.

2.3.2 La strategia di campionamento

Sul piano metodologico, uno dei limiti più significativi negli OPP riguarda la definizione e rappresentatività del campione. Spesso l'utilizzo di *web survey* è connesso a indagini su popolazioni di riferimento non chiaramente definite o definibili, e dunque soggette a forti dinamiche di autoselezione, con conseguente rischio di *bias* e compromissione della generalizzabilità dei risultati. L'adozione di un campionamento probabilistico rappresenta la soluzione metodologicamente più solida, ma risulta spesso complessa, costosa e in alcuni contesti difficilmente praticabile (Couper 2008).

È proprio su questo aspetto che si distinguono gli OPP che, per assicurare l'alta qualità dei dati rilevati, si fondano su strategie di campionamento probabilistico che garantiscono la rappresentatività della popolazione di riferimento e, quindi, la validità esterna e la generalizzabilità delle analisi (Callegaro *et al.* 2014). A differenza dei sondaggi online basati su campioni di convenienza spesso affetti da *bias* di autoselezione (Cornesse e Blom 2023), gli OPP impiegano rigorose procedure di campionamento che riflettono la composizione demografica della società di riferimento.

Tra i diversi tipi di panel online esistenti, gli OPP di cui trattiamo all'interno di questo volume, sono quelli rivolti alla popolazione generale, che necessitano dunque di campioni ampi e rappresentativi, costruiti con criteri probabilistici.

Altri tipi di panel includono, ad esempio (Callegaro *et al.* 2014):

- i *panel specializzati*, orientati a target specifici – come medici, piccole e medie imprese, o consumatori e consumatrici di un certo prodotto – utili per indagini settoriali o approfondimenti tematici;
- i *proprietary panels* creati invece da aziende o istituzioni per raccogliere dati dai propri utenti o clienti;

- i *panel elettorali*, progettati nello specifico per monitorare il voto della popolazione di riferimento, consentendo l'analisi delle dinamiche sul cambiamento delle intenzioni di voto e in generale sul clima politico che attraversa l'opinione pubblica;
- *passive data panels*, che raccolgono informazioni in modo automatico – ad esempio tramite il tracciamento del comportamento online – e rappresentano una frontiera interessante nello sviluppo della ricerca sociale contemporanea.

Tuttavia, nessuno di questi modelli adotta sistematicamente il campionamento probabilistico, condizione necessaria per generalizzare i risultati su base nazionale.

Il disegno di un panel implica però anche delle criticità in termini di campionamento. Considerate la mortalità del campione (in particolare per alcune categorie di popolazione) e il mutamento demografico, la strategia non si può mai considerare definitiva, ma dovrà prevedere successivi aggiustamenti e integrazioni (Maslovskaya e Lugtig 2022), così come verrà esplicitato nel paragrafo successivo.

2.3.3 Il disegno longitudinale

L'ultima caratteristica di rilievo in un OPP è il disegno longitudinale proprio del panel, che prevede la partecipazione degli stessi individui a più ondate di raccolta dati. Ciò consente di analizzare processi sociali dinamici e monitorare i cambiamenti nel tempo a livello individuale (Menard 2002). Come evidenziato da Biolcati Rinaldi e Vezzoni (2012), questa strategia di ricerca sociale è fondamentale al fine di osservare come e perché atteggiamenti, opinioni, valori o comportamenti dei soggetti si modificano, permettendo di collegare i cambiamenti individuali ad altre variabili di contesto o di analizzare in profondità le dinamiche sociali sottostanti ai mutamenti osservati. Tuttavia, le indagini panel presentano alcune criticità specifiche derivate dalla loro natura longitudinale.

Un primo limite è rappresentato dalla cosiddetta mortalità del campione o logoramento (*attrition*), che consiste nella perdita di soggetti nel corso delle successive rilevazioni (per motivi come rifiu-

ti, trasferimenti o decessi) o solo per alcune di esse, che può compromettere significativamente i risultati e la rappresentatività del campione. La mortalità del campione può essere in parte ridotta ricorrendo ad incentivi condizionati alla partecipazione al panel, oppure ricorrendo periodicamente all'integrazione del campione (*refresh*). Questa operazione consiste nell'aggiornare il campione disponibile aggiungendo un nuovo gruppo di partecipanti al fine di mantenere la rappresentatività e garantire l'accuratezza dei risultati. Questo si rende necessario soprattutto quando la distribuzione dei partecipanti e delle partecipanti che decidono di abbandonare lo studio o risultano irreperibili non è casuale. Inoltre, le caratteristiche demografiche della popolazione studiata possono mutare nel tempo, rendendo necessario un aggiornamento. Nello sviluppo degli OPP è quindi opportuno considerare interventi di *refreshment sampling* per mantenere o migliorare l'efficienza delle stime.

Un secondo problema è la cosiddetta interferenza (*panel conditioning*), ovvero il rischio che la partecipazione ripetuta all'indagine porti gli intervistati e le intervistate a modificare le proprie risposte, per un effetto di memoria, apprendimento o desiderabilità sociale. La tendenza dei e delle partecipanti a ricordare le risposte fornite in precedenti rilevazioni o a adattarle in base a ciò che percepiscono essere atteso, rischia di alterare la rappresentazione dei reali cambiamenti sociali.

2.4 Peculiarità specifiche degli OPP

Oltre alle caratteristiche principali già descritte, gli OPP si distinguono per alcune peculiarità che ne rafforzano il valore scientifico e ne ampliano le potenzialità applicative. Tipicamente, gli OPP sono indagini multiscopo che offrono la possibilità di includere al loro interno esperimenti integrati e di collegare i dati raccolti con fonti esterne, come dati amministrativi o registri ufficiali, ampliando così le prospettive analitiche. Questa flessibilità progettuale consente di approfondire temi specifici, testare ipotesi causali e migliorare la precisione delle stime attraverso l'integrazione di diverse fonti informative.

Inoltre, gli OPP di più recente sviluppo aderiscono ai principi FAIR (*Findable, Accessible, Interoperable, Reusable*) e si inseriscono pienamente nell'orizzonte dell'*Open Science*, promuovendo la trasparenza della ricerca e la condivisione dei dati secondo standard internazionali oltreché la possibilità per la comunità scientifica di collaborare attivamente nella produzione stessa dei dati. La struttura modulare degli OPP consente, infatti, la combinazione di moduli *core* – che raccolgono dati su ambiti centrali della vita come famiglia, istruzione, lavoro, salute, risorse economiche, partecipazione civica, valori e atteggiamenti – e moduli aggiuntivi tematici, che possono essere proposti da Enti o gruppi di ricercatori attraverso bandi competitivi. Questa struttura favorisce la democratizzazione della ricerca, poiché fornisce l'opportunità a tutta la comunità scientifica di utilizzare infrastrutture utili a condurre e sviluppare indagini di elevata estensione e qualità.

Tuttavia, queste potenzialità comportano anche sfide significative, prima fra tutte quella della sostenibilità a lungo termine (cfr. cap. 8). Gli OPP richiedono, infatti, risorse ingenti soprattutto nella fase iniziale di avvio (reclutamento, infrastrutture digitali, sviluppo dei moduli), ma è fondamentale garantire anche un impegno costante nel tempo in termini di risorse economiche e di personale qualificato, al fine di mantenere la continuità, la qualità e la tempestività delle rilevazioni.

Nonostante questi limiti però, i vantaggi operativi e scientifici degli OPP tendono a superare le criticità, soprattutto in contesti caratterizzati da una crescente alfabetizzazione digitale della popolazione che, di conseguenza, vede una riduzione progressiva del rischio di esclusione digitale. In sintesi, come sostiene Scherpenzeel (2011), rispetto ai panel tradizionali, i panel online offrono una ricchezza di dati superiore grazie alla possibilità di somministrare questionari più lunghi distribuendoli nel tempo, ai bassi costi di gestione di grandi campioni, all'accessibilità dei paradata e alla semplicità nell'implementazione di disegni sperimentali.

2.5 La progressiva affermazione degli OPP

L'origine degli OPP può essere fatta risalire alla seconda metà degli anni '80 con la progressiva affermazione delle tecniche di rilevazione dei dati informatizzate e autosomministrate (Callegaro *et al.* 2014). Anche se si rilevano già alcune sperimentazioni nel Regno Unito e in Francia, un punto di svolta avviene con la creazione del primo telepanel probabilistico olandese, che nel 1986 coinvolse ben 1.000 famiglie in una rilevazione settimanale, installando presso i singoli domicili un personal computer connesso a Internet, attraverso cui le risposte venivano periodicamente inviate a un *server* centrale (Saris 1998).

Una seconda fase di sviluppo degli OPP si è verificata a partire dalla metà degli anni '90 (Postoaca 2006), inizialmente negli Stati Uniti e, in misura minore, anche in alcuni Paesi europei. Come evidenziato da Callegaro e colleghi (2014), la principale espansione degli OPP si verifica successivamente, portando a un progressivo consolidamento soprattutto negli Stati Uniti e in Europa, ma via via anche in altre aree del mondo (Kocar e Kaczmirek 2023).

Il primo esempio strutturato può esser fatto risalire al *Knowledge Panel* realizzato da Ipsos negli Stati Uniti nel 1999. Negli anni successivi, si assiste alla nascita dell'*American Life Panel* (2003) e del *Gallup Panel* (2004). A partire dal 2007, l'espansione si intensifica: è in quell'anno che viene istituito il FFRISP (*Face-to-Face Recruited Internet Survey Panel*) in collaborazione tra l'Università di Stanford e l'istituto Abt SRBI, cui seguono numerosi altri OPP attualmente operativi. Il campione del FFRISP, solido, rappresentativo e reclutato faccia-a-faccia, confluisce poi nell'*American Life Panel*, rendendolo tutt'oggi una delle infrastrutture più affidabili e durature per l'analisi longitudinale basata su campionamento probabilistico negli Stati Uniti.

Tra il 2014 e il 2015, si registrano iniziative quali l'*AmeriSpeak Panel* (2014), l'*American Trends Panel* (2014) e lo *SSRS Opinion Panel* (2015), oltre a panel tematici come quello utilizzato per lo studio elettorale dalla Stanford University nel 2008 (*American National Election Study Panel*). La progettazione e l'implementazione di questi OPP è spesso riconducibile alla collaborazione tra istituzioni ac-

cademiche e centri di ricerca privati, che congiuntamente garantiscono sia solidità metodologica che risultati attendibili e stabilità nel tempo. La scelta del disegno probabilistico e di panel si accompagna a una varietà di modalità di reclutamento, tra cui spiccano il contatto telefonico, la posta ordinaria e, in alcuni casi, il reclutamento faccia-a-faccia. La presenza di incentivi per i partecipanti risulta ampiamente utilizzata, così come l'adozione di moduli aggiuntivi che si vanno ad innestare nell'indagine principale.

Il caso europeo presenta caratteristiche parzialmente differenti, sia sul piano temporale sia su quello organizzativo. Se si esclude l'esperienza precoce del *Sifo Panel* in Svezia (1995), la diffusione degli OPP nel continente si avvia in modo sistematico a partire dal 2005, con il *Demoskop Panel*. Un'accelerazione si registra a partire dal 2007, con l'avvio del LISS Panel (*Longitudinal Internet Studies for the Social Sciences*), gestito da Centerdata presso l'Università di Tilburg nei Paesi Bassi, che si è affermato come punto di riferimento per i panel europei. Seguono l'istituzione del *Novus Panel Sweden* (2008), dell'islandese *Social Science Research Institute – SSRI Panel* (2010), del *The Citizen Panel* svedese (2010), del GIP (*German Internet Panel*) dell'Università di Mannheim (2012), dell'ELIPSS Panel (*Longitudinal Internet Studies for Social Sciences*) gestito da Sciences Po di Parigi in collaborazione con il CDSP – *Centre de Données Socio-Politiques* (2012), del *GESIS Panel* anch'esso tedesco (2013), e del *Norwegian Citizen Panel* dell'Università di Bergen (2013), oltre che i panel britannici, come *NatCen Opinion Panel* (2015) e *Taking Part* (2016).

In particolare, è il *Longitudinal Internet Studies for the Social sciences* (LISS) ad essere diventato il punto di riferimento europeo e internazionale per gli OPP. Avviato nei Paesi Bassi nel 2007 da Statistics Netherlands, si basa su un campione probabilistico della popolazione residente, costruito a partire dall'anagrafe nazionale. Comprende ad oggi circa 5.000 famiglie per un totale di oltre 7.500 individui intervistati di 16 anni o più. Sebbene il processo di rilevazione avvenga principalmente online, si includono anche gli *offliners* fornendo la strumentazione necessaria alla compilazione e garantendo così la copertura totale della popolazione di riferimento. Inoltre, la rappresentatività è garantita da periodici *refreshment sample*. I que-

stionari – costituiti da moduli *core* somministrati annualmente e moduli addizionali – sono compilati mensilmente dai singoli panelisti che ricevono incentivi condizionati al tempo di partecipazione. La gestione dei dati è totalmente *open*, con un archivio accessibile gratuitamente online ed utilizzato da migliaia di studiosi a livello globale. Ad oggi, sono state prodotte a partire dai dati LISS oltre 1.600 pubblicazioni scientifiche (Das 2025). Sono state sperimentate anche numerose innovazioni metodologiche, come l'integrazione di dati provenienti da dispositivi smart (es. *balance*), *wearable* (es. accelerometri), smartphone (es. Uso del Tempo, dati GPS), e si stanno attualmente raccogliendo anche dati visuali e vocali, un'opportunità per rendere le indagini sempre più inclusive e meglio contestualizzate. Il LISS è infine un chiaro esempio di come sia possibile condurre esperimenti comportamentali tramite OPP e continua ad essere pioniere nell'innovazione metodologica. Recentemente si è aperta, ad esempio, la possibilità della *data donation*, ovvero di raccogliere dati digitali volontariamente forniti dai panelisti che apre molteplici orizzonti di ricerca (*ibidem*).

Un caso particolare è poi rappresentato da CRONOS (*CROss-National Online Survey*), attivato tra il 2016 e il 2017, nell'ambito dell'infrastruttura europea SERISS (*Synergies for Europe's Research Infrastructures in the Social Sciences*) e coordinato centralmente dal consorzio ESS ERIC (*European Social Survey*). Il panel ha coinvolto inizialmente tre Paesi – Regno Unito, Slovenia ed Estonia – per poi allargarsi a dodici, e costituisce uno dei primi esempi concreti di panel probabilistico online multinazionale in Europa. La sua progettazione è stata orientata sin dall'inizio alla creazione di un disegno metodologico armonizzato, utile a testare la fattibilità di indagini longitudinali comparabili in contesti nazionali differenti. Si sviluppa in sei *waves* e la struttura modulare dei questionari consente di includere tematiche comuni e specifiche per ciascun Paese.

Nel panorama tedesco è utile segnalare anche il progetto GLEN (*German Longitudinal Environmental Study*), il primo OPP non multiscopo e dedicato allo studio di tematiche ambientali coordinato dall'Università Ludwig Maximilian di Monaco di Kaiserslautern e dall'Università di Lipsia a partire dal 2024 e per una durata programmata di 12 anni.

In generale, gli OPP europei condividono l'adozione del campionamento probabilistico e impiegano modalità di reclutamento spesso ibride, mescolando tecniche CAPI (*Computer-Assisted Personal Interviewing*), CATI (*Computer-Assisted Telephone Interviewing*) e postali. Una peculiarità rilevante è l'attenzione alla copertura della popolazione *offline*: esclusi i Paesi scandinavi, nella maggioranza dei casi, sono previsti dispositivi digitali o connessione Internet forniti ai partecipanti. Mentre l'uso di incentivi è piuttosto diffuso, la pratica del *piggybacking* – ovvero la richiesta al rispondente di un'indagine di partecipare a ulteriori indagini risparmiando reclutamento e risorse – rimane meno estesa rispetto al contesto statunitense, ma risulta comunque applicata in alcuni contesti, tra cui il GESIS Panel e CRONOS.

Nel gruppo eterogeneo di Paesi esterni all'area euro-atlantica, l'introduzione degli OPP si può considerare un fenomeno più recente. In Canada, l'esperienza del *Probit*, attiva dal 2014, si configura come una trasposizione del modello nordamericano. In Australia, Kocar e Kaczmarek (2023) distinguono, in particolare, due OPP: l'*Australian Health and Social Science Survey* (2009), attualmente non attivo, e il *Life in Australia* (2016), ancora attivo. In Asia, il *KAMOS* della Corea del Sud e l'*IranPoll* iraniano sono stati entrambi avviati nel 2016 e testimoniano l'interesse crescente per gli OPP anche in contesti caratterizzati da diverse condizioni istituzionali e tecnologiche. In queste esperienze, il campionamento è sempre di tipo probabilistico e le modalità di reclutamento includono telefono, posta e, in alcuni casi, l'intervista faccia-a-faccia. L'uso di incentivi appare meno sistematico, mentre la presenza di *piggybacking* è attestata solo in alcune iniziative. È invece meno frequente, rispetto agli esempi europei e statunitensi, la previsione di dispositivi o connessione Internet per i partecipanti non digitalizzati.

Nel complesso, la sequenza cronologica e la distribuzione geografica degli OPP presi in considerazione mostrano un processo di diffusione differenziale: precoce e trainato da attori privati negli Stati Uniti, più istituzionalizzato e armonizzato in Europa, in fase di consolidamento in contesti extra-occidentali. A questa diversità temporale e geografica si affianca una pluralità di configurazioni operative: le modalità di reclutamento, la gestione della popolazione *offline*,

l'impiego di incentivi e l'adozione di pratiche accessorie, come il *piggybacking*, si rivelano fortemente influenzate dai vincoli strutturali, dalle infrastrutture tecnologiche e dalle culture metodologiche proprie di ciascun sistema nazionale. Si delinea così un quadro policentrico e in continua evoluzione, in cui la tensione tra standardizzazione e adattamento locale rappresenta una delle principali sfide della ricerca sociale contemporanea condotta attraverso OPP.

Tuttavia, nonostante le differenze, questi panel condividono alcuni aspetti fondamentali, tra cui la natura longitudinale e la finalità di rappresentare fedelmente la popolazione generale dei rispettivi Paesi, oltreché la necessità di andare verso la completa digitalizzazione dei questionari. Sebbene le scelte metodologiche siano differenti, le prospettive di sviluppo degli OPP in Europa sembrano convergere verso una maggiore armonizzazione metodologica e tendere verso l'*Open Science*, in particolare per facilitare gli studi e la comparazione internazionale.

2.6 Prospettive future

Gli OPP offrono una risposta metodologicamente solida alle sfide poste dalla complessità sociale contemporanea, costituendo uno strumento ormai imprescindibile per l'analisi empirica orientata alla comprensione delle trasformazioni strutturali e culturali in atto. Benché permangano alcune criticità, in particolare in termini di sostenibilità, mantenimento di un campione rappresentativo, e disuguaglianze digitali, il bilancio tra costi e benefici appare, nel complesso, largamente favorevole.

Negli ultimi anni si rileva una progressiva affermazione degli OPP anche in contesti nazionali finora marginalmente coinvolti in questo tipo di infrastrutture. Si assiste, da un lato, all'istituzione di nuovi panel costruiti *ex novo* secondo criteri probabilistici; dall'altro, alla trasformazione di panel preesistenti, inizialmente strutturati su campioni non probabilistici o di tipo misto, in direzione di disegni metodologici maggiormente rigorosi e orientati alla rappresentatività. Ne sono esempi recenti l'avvio di *The Social Study* in Belgio e dell'*Italian Online Probability Panel* (IOPP), di cui si occupa il pre-

sente volume, così come ulteriori iniziative emergenti di cui si ha contezza, ad esempio, in Croazia e Serbia. Questi sviluppi riflettono un più ampio processo di ristrutturazione delle pratiche di rilevazione campionaria, che coinvolge non solo l'adozione di tecniche di campionamento probabilistico, ma anche la progressiva digitalizzazione dell'intero ciclo di rilevazione, dalla somministrazione all'elaborazione e alla diffusione dei dati, secondo i principi FAIR e nell'ottica di una più ampia apertura del processo di ricerca e del coinvolgimento costante della comunità scientifica.

In un contesto segnato da mutamenti sociali rapidi e da crisi a carattere sistemico – come quella pandemica, i conflitti armati in corso o l'instabilità dell'ordine politico internazionale – gli OPP si confermano, dunque, come strumenti particolarmente adatti alla rilevazione di dinamiche complesse, sia per la loro capacità di cogliere variazioni individuali nel tempo, sia per la loro versatilità rispetto alle esigenze di monitoraggio continuo. La loro evoluzione, inoltre, si intreccia sempre più strettamente con i progressi tecnologici, che consentono oggi l'integrazione di componenti avanzate (intelligenza artificiale, *machine learning*, sensori, metadati comportamentali), e la possibilità di integrazione di diverse e variegate forme di dati (vocali, testuali, amministrativi, social, etc.). In tale prospettiva, gli OPP non si configurano soltanto come un'evoluzione tecnico-operativa di una metodologia d'indagine, ma piuttosto come una nuova infrastruttura di ricerca sociale a disposizione degli studi empirici, capace di coniugare rappresentatività, efficienza, rapidità di raccolta e di diffusione dei dati seguendo i principi FAIR e dell'*Open Science*.

2.7 Riferimenti bibliografici

- Agnoli, M.S., 2008, *Il disegno della ricerca sociale*, Roma, Carocci editore.
- Biolcati Rinaldi F., Vezzoni C., 2012, *L'analisi secondaria nella ricerca sociale*, Bologna, Il Mulino.
- Blom A. G., Bosnjak M., Cornilleau A., Cousteaux A. S., Das M., Douhou S., Krieger U., 2016, *A comparison of four probability-based online and mixed-mode panels in Europe*, in *Social Science Computer Review*, 34(1), 8-25.

- Callegaro M., Baker R., Bethlehem J., Göritz A. S., Krosnick J. A., Lavrakas P. J., 2014, *Online panel research: A data quality perspective*. United States, New York, Wiley.
- Cornesse C., Blom A. G., 2023, *Response quality in nonprobability and probability-based online panels*, in *Sociological Methods & Research*, 52(2), 879-908.
- Couper M. P., 2008, *Designing effective web surveys*, Cambridge, Cambridge University Press.
- Das M., 2025, *Building on the LISS Experience: Future Challenges and Opportunities for Online Probability-based Panels*, in Conference "European Online Probability Panels (EOPPs): opportunities for the social research", Milan 5-6 May 2025.
- Galtung J., 1967, *Theory and Methods of Social Research*, Oslo: Universitetsforlaget.
- Göritz A. S., 2007, *Using online panels in psychological research*, In A. N. Joinson, K. Y. A. McKenna, T. Postmes, and U. D. Reips (Eds.), *The Oxford handbook of Internet psychology*, 473–85. Oxford: Oxford University Press.
- Istat, 2017, *L'effetto tecnica nelle indagini mixed-mode. Aspetti teorici e sperimentazioni su indagini sociali che utilizzano il web*, Roma, Istituto Nazionale di Statistica.
- Kocar S., Kaczmirek L., 2023, *A meta-analysis of worldwide recruitment rates in 23 probability-based online panels, between 2007 and 2019*, in *International Journal of Social Research Methodology*, 27(5), pp. 589-604. doi: 10.1080/13645579.2023.2242202.
- Lynn P., 2012, *Longitudinal survey methods for the household finances and consumption survey*. Report prepared for the European Central Bank, https://www.ecb.europa.eu/home/pdf/research/hfcn/Lynn_LongitudinalSurveyMethodsfor_HFCS.pdf?a262e4f79441fcf06a02fb92bf317d27.
- Mauceri S., Faggiano M. P., Di Censi L., 2020, *Survey 2.0. L'indagine con questionario nell'era digitale*, in *Sociologia e Ricerca Sociale*, 121, pp. 25-48. doi: 10.3280/SR2020-121002.
- Maslovskaya O., Lugtig P., 2022, *Representativeness in Six Waves of Cross-National Online Survey (CRONOS) Panel*, in *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185(3), 851-871. doi: <https://doi.org/10.1111/rssa.12801>.
- Menard S., 2002, *Longitudinal Research. Quantitative Applications in the Social Sciences*, Thousand Oaks, CA: SAGE Publications, Inc.
- Postoaca A., 2006, *The anonymous elect: Market research through online access panels*, Berlin, Springer Science & Business Media.
- Ruspini E., 2002, *An Introduction to Longitudinal Research*, London and New York, Routledge.

- Saris W., 1998, *Ten Years of Interviewing without Interviewers: The Telepanel*, in M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, W. L. Nicholls II, J. M. O' Reilly (eds.), *Computer assisted survey information collection*, New York, Wiley & Sons, Inc.
- Scherpenzeel A., 2009, *Start of the LISS panel: Sample and recruitment of a probability-based Internet panel*, Tilburg, CentERdata.
- Scherpenzeel A., 2011, *Data Collection in a Probability-Based Internet Panel: How the LISS Panel Was Built and How It Can Be Used*, in *Bulletin of Sociological Methodology*, 109, 56-61.
- Taddei L., Diana P., 2024, *Web survey: esperienze LGBT+ nello sport in Europa*, in P. Diana, M. C. Catone, L. Taddei (a c. di), *Fare ricerca sociale sullo sport. Metodologia, tecniche e applicazioni*, Milano, FrancoAngeli.
- Veltri G. A., 2021, *La ricerca sociale digitale*, Milano, Mondadori.
- Yeager D. S., Krosnick J. A., Chang L. C., Javitz H. S., Levendusky M. S., Simpser A., Wang R., 2011, *Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples*, in *Public Opinion Quarterly*, 75, pp. 709-747.

3. L'Architettura di IOPP tra sfide scientifiche e operative

di Nicolò Marchesini, Francesco Piacentini, Ilaria Primerano e Luciana Taddei*

3.1 Introduzione

L'*Italian Online Probability Panel* (IOPP) è stato progettato, in linea con gli altri panel probabilistici online europei, seguendo un'architettura procedurale definita e prestrutturata.

Il funzionamento di un panel di ricerca sociale basato su campionamento probabilistico richiede un disegno attento non solo alla rappresentatività iniziale, ma anche alla capacità di mantenere nel tempo la qualità e la coerenza della rilevazione.

Questo capitolo entra nel merito degli aspetti strutturali e strumentali che rendono operativa l'architettura del panel, con particolare attenzione alla periodicità delle rilevazioni, alla gestione dei diversi segmenti di popolazione e all'organizzazione degli strumenti di raccolta dei dati. Lungi dal fornire una semplice infrastruttura tecnica, in IOPP ogni elemento – dal reclutamento iniziale alla costruzione condivisa dei moduli – è pensato per produrre dati che non siano solo “disponibili a tutti”, ma anche epistemologicamente fondati, metodologicamente robusti e socialmente utili.

La combinazione tra continuità e adattabilità, tra rigore metodologico e apertura all'innovazione, rappresenta un presupposto essenziale per la tenuta longitudinale del disegno e per la sua capacità di intercettare con tempestività i mutamenti sociali in corso. La struttura

* Il presente capitolo è il frutto di una riflessione comune da parte delle autrici e degli autori.

delle *wave*, i meccanismi di controllo e aggiornamento del campione, così come la configurazione e l'impiego degli strumenti di rilevazione, non costituiscono semplici accorgimenti tecnici, ma dispositivi centrali nella costruzione del dato. È attraverso queste scelte che si rende possibile una produzione informativa solida, significativa e in grado di offrire letture empiricamente fondate di una realtà sociale in continuo mutamento.

3.2 Struttura e gestione di IOPP

IOPP, in continuità con gli *Online Probability Panels* europei più all'avanguardia, si distingue nel panorama italiano per la sua architettura operativa: altamente strutturata, progettata per garantire qualità, tempestività e rappresentatività dei dati in un contesto sociale in rapido mutamento. L'articolazione in *multiple wave* annuali, la flessibilità offerta dalla struttura modulare dei questionari – che prevedono una parte *core*, somministrata con cadenza periodica e una parte *additional* somministrata *una tantum* per approfondire tematiche specifiche – l'aggiornamento del campione e la continua attenzione al *panel care* costituiscono i pilastri fondamentali della struttura di IOPP.

Nello specifico, l'attività di rilevazione di IOPP è suddivisa in cinque *wave* annuali che costituiscono la parte preponderante del panel, e vengono somministrate con cadenza regolare ogni 2-3 mesi. Brevi intervalli tra le rilevazioni permettono di migliorare la sensibilità del panel, aumentando la capacità di rilevare tempestivamente i mutamenti nelle dinamiche sociali ovvero di catturare variazioni rapide nei comportamenti, nelle percezioni e nelle condizioni socioeconomiche (Haas *et al.* 2021; Blom *et al.* 2020), riducendone al contempo gli errori e migliorando la precisione delle stime (Halpern-Manners e Warren 2012). Così facendo è possibile ottenere un monitoraggio costante dei fenomeni oggetto di studio e, dunque, della popolazione italiana in termini di strutture familiari, condizioni abitative, condizioni socioeconomiche, disuguaglianze, competenze digitali, stili di vita, comportamenti, opinioni, valori e atteggiamenti (cfr. cap. 6). Inoltre, al fine di ridurre la mancata risposta o le risposte le-

gate alla contingenza del periodo, le *wave* sono state pianificate in modo da evitare i periodi di ferie ed alta mobilità degli individui, quali i mesi estivi e la pausa natalizia (Dillman *et al.* 2014; Callegaro *et al.* 2015).

Nella pianificazione delle *wave* si è tenuto conto del *digital divide* che caratterizza il nostro Paese, dovuto sia a fattori geografici legati a questioni infrastrutturali (es. aree interne e isole minori) sia a fattori sociodemografici, quali l'elevata età media della popolazione e la bassa istruzione digitale che si intrecciano con le questioni di genere (Plan International e Università Bocconi 2021; European Commission 2024; Governo Italiano 2021). Infatti, al fine di includere nella rilevazione le persone che non hanno accesso regolare a Internet, hanno scarse o nulle competenze digitali o che preferiscono modalità di risposta alternative (*offliners*), è prevista la possibilità di somministrare il questionario *core* (cfr. paragrafo 3.3), opportunamente riadattato, in modalità PAPI (*Paper and Pencil Interviewing*), a seguito dell'invio postale. Questa doppia componente, in linea con le raccomandazioni metodologiche internazionali (Couper 2000; De Leeuw 2005), consente di ridurre il *bias* di copertura e garantire l'inclusione di sottogruppi della popolazione potenzialmente esclusi dai panel *web-based*, rafforzando la rappresentatività del campione complessivo.

Le *wave* dell'indagine seguono tempistiche precise, con date definite di apertura e chiusura per garantire coerenza e qualità nella raccolta dei dati che variano in funzione della modalità di somministrazione. In particolare, le rilevazioni online hanno una finestra di compilazione di durata predefinita. Un lasso temporale pensato per essere sufficiente a garantire un'alta qualità del dato e una raccolta rapida, riducendo il rischio di deterioramento della memoria dei rispondenti e mantenendo alta la coerenza interna tra le risposte (Dillman *et al.* 2014). Per le *wave* rivolte agli *offliners*, le tempistiche si estendono così da garantire che si compia l'intero *iter* del processo postale e si possa compilare il questionario nella sua interezza. Questa maggiore durata è giustificata dalla necessità di utilizzare modalità di contatto e risposta più tradizionali, come il postale e il PAPI, per le quali in fase di pianificazione è importante tenere conto: (a) delle tempistiche postali, (b) della compilazione cartacea da parte della o del rispondente, (c) del seguente re-invio del questionario. Tale strategia risulta

funzionale ad ottimizzare la risposta pur in presenza di barriere tecnologiche o culturali.

Sempre nell'ottica di facilitare la partecipazione, minimizzare il tasso di non risposta e consentire la personalizzazione delle somministrazioni, nella pianificazione di IOPP si prevede la messa in opera di differenti servizi integrativi a supporto del panelista, quali un *help desk* telefonico, un piano di solleciti telefonici e un sistema di *reminder* via app, attivi durante l'intero periodo di rilevazione del questionario. L'*help desk* è stato pensato per permettere alle e ai rispondenti di ricevere assistenza tecnica, chiarimenti sui contenuti del questionario e informazioni sulle modalità di accesso alla piattaforma. Il piano di solleciti periodici prevede il contatto della o del panelista via e-mail e/o lettera cartacea (per gli/le *offliners*) e nei casi di mancata risposta a questi primi solleciti, il contatto telefonico. Parallelamente, il sistema di *reminder* via app prevede l'invio di notifiche personalizzate e progressivamente più incisive, calibrate in base alla sensibilità e alla storicità del comportamento della e del rispondente.

A supporto di queste attività, ogni panelista registrato ha a disposizione una *web app* accessibile sia da smartphone che da *browser desktop*, attraverso la quale può partecipare alle diverse *wave* e aggiornare il proprio profilo utente. La web app è stata sviluppata utilizzando il *framework Laravel*, che garantisce un'architettura solida, flessibile e sicura per la gestione dell'interazione con l'utenza.

Inoltre, al fine di mantenere la rappresentatività nel tempo e di monitorare l'evoluzione di nuove generazioni e gruppi emergenti (Lynn 2009), nel disegno di IOPP la somministrazione ciclica delle *wave* è parallelamente accompagnata da un meccanismo di *refreshment* annuale del campione, ossia il reclutamento e l'inserimento programmato di nuove e nuovi partecipanti realizzato seguendo gli stessi standard metodologici del campione originario e finalizzato a compensare nel tempo le uscite fisiologiche dal campione (per rinuncia, per non partecipazione attiva, per raggiunta età massima o per decesso) e l'inserimento di nuove coorti nella popolazione target (ad esempio i neo-18enni).

Solo per il primo anno di avvio dell'indagine particolare attenzione sarà posta alla valutazione della qualità del campione e alla calibrazione dello strumento, finalizzata al miglioramento delle procedu-

re di somministrazione e sollecito, nonché delle domande del questionario *core*. A tal fine, è prevista una sequenza di tre *wave* consecutive rivolte al primo gruppo di panelisti e paneliste reclutati per l'indagine. Nello specifico, questa fase mira a testare lo strumento di rilevazione sotto diversi profili: contenuti, difficoltà, tempi di compilazione o mancate risposte, ma anche *user experience*, facilità di utilizzo attraverso diversi dispositivi, modalità di invito alla compilazione ed efficacia dei *reminder*, al fine di identificare eventuali problemi tecnici, contenutistici o organizzativi. I risultati di questa fase saranno utili per l'ottimizzazione del disegno del panel e delle procedure di somministrazione prima dell'avvio della rilevazione ordinaria. Inoltre, il ciclo di tre *wave* iniziali consente anche di valutare la qualità del campione, monitorare il tasso di *attrition*, oltre che intervenire su eventuali criticità nella progettazione dei questionari o nei protocolli di somministrazione.

Le prime tre *wave* fungono, dunque, anche da *benchmark* per la successiva implementazione di strategie atte a bilanciare i risultati (es. pesatura e post-stratificazione), fondamentali per correggere le distorsioni dovute a non-risposta selettiva e per garantire l'inferenza rappresentativa a livello nazionale e di ripartizione geografica (Groves *et al.* 2009).

3.3 Dal disegno del panel alla somministrazione dell'indagine

L'architettura di IOPP si fonda su un'articolazione sistematica funzionale di tre principali strumenti di rilevazione, simili ed integrabili tra loro. Nello specifico, ci si riferisce a questionari standardizzati, differenziati per contenuti, finalità e modalità di somministrazione. Una delle caratteristiche peculiari di tali strumenti è la prospettiva di flessibilità che li caratterizza, sia al loro interno, sia nella possibilità di interazione reciproca.

In una prima fase, al fine di reclutare i panelisti di IOPP è stato necessario predisporre e somministrare un primo breve questionario, il *questionario di reclutamento*, che si ponesse come strumento di

aggancio degli individui selezionati in fase di campionamento (cfr. cap. 4).

Il questionario di reclutamento offre ai e alle rispondenti la possibilità di comprendere la tipologia di domande e il possibile impegno operativo e cognitivo richiesto da IOPP, consentendo così una registrazione libera e consapevole al panel. Al tempo stesso, i dati rilevati consentono di ottenere informazioni iniziali sui e sulle rispondenti, utili per identificarne le caratteristiche e, se necessario, categorizzarli in gruppi omogenei, così da consentire l'eventuale estrazione di sotto-campioni per specifiche raccolte o analisi di dati (cfr. cap. 7).

Il questionario di reclutamento raccoglie dati sociodemografici e informazioni su tematiche ampie e varie, oltreché gli atteggiamenti verso l'indagine e lo strumento stesso. Viene somministrato in *mixed-mode*, modalità CAWI e CAPI, tramite un approccio *push-to-web* che prevede l'invio iniziale di una lettera di invito corredata da *link* e QR code personalizzati. I dati rilevati in questa prima fase di indagine consentono di ottenere variabili di base utili a definire possibili sottopopolazioni nelle successive fasi di raccolta o analisi dati, come filtri per indirizzare specifiche *wave* o moduli di domande a sottogruppi del campione complessivo. Al fine di rendere il panel flessibile alle esigenze della ricerca e dinamico rispetto all'evoluzione sociodemografica della popolazione, queste variabili-attributi che definiscono il campione potranno essere aggiornate e monitorate da ogni panelista tramite area riservata, accessibile sia da app che da *browser*.

I panelisti aderenti a IOPP risponderanno a due tipi di questionari: il *Core Questionnaire* e gli *Additional Questionnaires*.

Il *Core Questionnaire* (CQ), somministrato – a regime – con cadenza regolare suddivisa in tre soluzioni lungo le prime tre *wave* annuali previste, rappresenta lo strumento cardine per la raccolta di dati longitudinali. La sua struttura copre ambiti centrali della vita individuale e sociale (informazioni sociodemografiche; benessere e salute; situazione economica; famiglia e nucleo familiare; istruzione e formazione; integrazione sociale e caratteristiche del quartiere; uso dei media e tempo libero; religione; atteggiamenti individuali e verso la politica; scienza e tecnologia), ed è progettata per garantire la comparabilità temporale e la coerenza con i principali standard internazio-

nali (cfr. cap. 6). In particolare, nelle prime tre *wave* annuali il CQ coprirà il 70% dello spazio disponibile per il questionario, lasciando la parte restante alla somministrazione di eventuali moduli aggiuntivi. Il tempo di somministrazione è stimato in circa 60 minuti totali, ovvero 20 minuti circa di somministrazione per *wave*.

A complemento del CQ, il disegno prevede l'impiego di *moduli aggiuntivi* (altrimenti detti *Additional Questionnaires – AQ*), somministrati su sottocampioni o su tutta la popolazione di panelisti e paneliste, e distinguibili tra *rotanti*, ovvero somministrati con cadenza pluriennale a seconda del disegno di ricerca, e *non rotanti*, ossia somministrati una sola volta (*one shot*), al fine di indagare questioni attuali, specifiche o contingenti.

Una delle principali potenzialità offerte da un disegno di panel consiste infatti nel fornire un'infrastruttura stabile e prontamente attivabile sempre disponibile per l'implementazione di rilevazioni dirette, anche in situazioni emergenziali o di particolare interesse sociale. Inoltre, va rilevato come la natura *one-shot* di un modulo non implichi necessariamente la sua irripetibilità nel tempo: la possibilità di riproporre lo stesso contenuto, anche a distanza di anni, sul medesimo insieme di rispondenti, consente di trasformare *ex post* un'indagine trasversale in un'analisi diacronica, con evidenti vantaggi in termini di comparabilità e profondità interpretativa.

Inoltre, l'architettura immaginata per IOPP offre la possibilità di integrare i dati raccolti negli AQ con le informazioni longitudinali del CQ. Ciò consente di collegare efficacemente ambiti tematici diversi – come salute, personalità, valori, partecipazione sociale, condizioni economiche – e di analizzarli in modo congiunto. La struttura integrata del sistema permette così lo sviluppo di indagini complesse e multidimensionali, in cui le diverse dimensioni indagate possono essere studiate in relazione tra loro.

Gli *Additional Questionnaire* potranno essere strutturati dal Comitato Scientifico interno a IOPP o proposti da gruppi e/o enti di ricerca e università tramite la partecipazione ad *open call* cicliche. I moduli aggiuntivi consentono di esplorare in profondità tematiche specifiche o emergenti e, limitandosi nell'estensione, consentono di non gravare ulteriormente sul rispondente.

La somministrazione avverrà principalmente in maniera auto-somministrata online (CAWI), salvo l'inclusione di quelle parti di popolazione italiana che presentano difficoltà nell'accesso a internet o nell'utilizzo di strumenti digitali.

L'architettura modulare di IOPP consente di creare un'infrastruttura dinamica e collaborativa, di cui l'intera comunità scientifica si può avvalere, utilizzando sia i dati già disponibili, sia creando dati nuovi ed integrabili, a loro volta condivisibili.

Elemento qualificante del panel è, dunque, il ricorso sistematico a una serie di *open call*, rivolte alla comunità scientifica nazionale e internazionale.

L'*open call* si configura come uno strumento strategico di partecipazione e co-progettazione della ricerca, attraverso cui gruppi di ricerca possono proporre moduli tematici da integrare nel programma di rilevazione del panel. Le proposte dovranno essere presentate secondo una procedura formalizzata, che prevede la compilazione di un apposito modulo contenente: titolo e descrizione del modulo, obiettivi teorici ed empirici, inquadramento metodologico, indicazione della popolazione *target* (intero campione o sottocampione), stima del tempo di somministrazione e, se necessario, eventuali richieste di dati aggiuntivi o eventuali collegamenti con altre fonti.

Le domande sono valutate dal Comitato Scientifico di IOPP, secondo criteri che includono: coerenza con le finalità generali del panel, solidità teorico-metodologica, rilevanza scientifica e sociale del tema, sostenibilità operativa (durata del modulo, compatibilità con le *wave* disponibili, onere per il rispondente), e potenziale valore aggiunto per la comunità scientifica. In linea con esperienze internazionali come quelle sviluppate nei panel di ricerca europei, il processo di selezione è strutturato in più fasi: valutazione scientifica, verifica tecnica da parte del gruppo operativo, ed eventuale revisione condivisa con i proponenti per garantire l'ottimizzazione del disegno del modulo.

Questo approccio consente l'apertura in progettazione alla comunità scientifica, favorisce la flessibilità tematica e rafforza la natura collaborativa, pubblica e plurale del panel, in coerenza con i principi dell'*Open Science*. Inoltre, consente al panel di rispondere in modo tempestivo a esigenze conoscitive emergenti, promuovendo una ri-

cerca empirica aggiornata, comparabile e orientata all’impatto politico-sociale.

3.4 Riflessioni conclusive

La costruzione di un panel probabilistico come IOPP, soprattutto in un contesto come quello italiano, non è mai un esercizio puramente tecnico. Ogni scelta – dalla frequenza delle *wave* alla configurazione degli strumenti, dal trattamento degli *offliners* alla logica delle *open call* – riflette un’idea di ricerca e una visione della società. L’architettura descritta in questo capitolo non si limita dunque a garantire la raccolta sistematica e rappresentativa di dati, ma disegna uno spazio in cui la produzione empirica è messa costantemente alla prova delle sue condizioni di possibilità: tecniche, teoriche, ma anche sociali e politiche.

La sfida non sta solo nel garantire coerenza e qualità nel tempo, ma nel costruire un’infrastruttura di ricerca in grado di restituire complessità, tenere insieme continuità e mutamento, e riconoscere nel dato non un semplice esito oggettivo, ma il risultato di una relazione fra soggetti, strumenti e contesto. In questo senso, IOPP rappresenta più di un progetto: è un laboratorio aperto in cui la ricerca sociale italiana può misurarsi con le proprie ambizioni, i propri limiti e le proprie responsabilità. Non si tratta solo di raccogliere “buoni dati”, ma di contribuire a ridefinire – nel tempo – che cosa significhi oggi fare “buona ricerca”.

3.5 Riferimenti bibliografici

Blom A. G., Cornesse C., Friedel S., Krieger U., Fikel M., Rettig T., Reifenscheid M., 2020, *High Frequency and High Quality Survey Data Collection: The Mannheim Corona Study*, in *Survey Research Methods*, 14(2), 171-178.

Callegaro M., Manfreda K., Vehovar V., 2015, *Web survey methodology*, London, SAGE Publications Ltd.

Couper M. P., 2000, *Web Surveys: A Review of Issues and Approaches*, in *The Public Opinion Quarterly*, Vol. 64 (4), 464-494.

- De Leeuw E., 2005, *To Mix or Not to Mix Data Collection Modes*, in *Surveys*, in *Journal of Official Statistics*, 21, 233-255.
- Dillman D. A., Smyth J. D., Christian L. M., 2014, *Internet, phone, mail, and mixed mode surveys: The tailored design method* (4th ed.), Hoboken, New Jersey, John Wiley & Sons, Inc.
- European Commission, 2024, *2030 DIGITAL DECADE. Annex 1: COMPETITIVENESS AND SOVEREIGNTY, PEOPLE, SMART GREENING, POLICY COHERENCE AND SYNERGIES*, Luxembourg, Publications Office of the European Union.
- Governo Italiano, 2021, *Piano Nazionale di Ripresa e Resilienza (Missione 1: Digitalizzazione, innovazione, competitività, cultura e turismo)*.
- Groves R. M., Fowler F. J., Couper M. P., Lepkowski J. M., Singer E., Tourangeau R., 2009, *Survey methodology* (2nd ed.), Hoboken, New Jersey, Wiley.
- Haas G. C., Müller B., Osiander C., Schmidtke J., Trahms A., Volkert M., Zins S., 2021, *Development of a new COVID-19 panel survey: the IAB high-frequency online personal panel (HOPP)*, in *J. Labour Mark Res.*, 55(1), 16.
- Halpern-Manners A., Warren J. R., 2012, *Panel conditioning in longitudinal studies: evidence from labor force items in the Current Population Survey*, in *Demography*, 49(4), 1499-519.
- Lynn P., 2009, *Methodology of Longitudinal Surveys*, Hoboken, New Jersey, Wiley.
- Plan International e Università Bocconi, 2021, *Sfide attuali e future per la Parità di Genere in Italia: Il divario digitale di genere*, Milano, Università Bocconi.

4. La strategia di campionamento

di *Ilaria Primerano, Nicolò Marchesini e Michele Santurro**

4.1 Introduzione

Il presente capitolo illustra il disegno di campionamento adottato per la costituzione dell'*Italian Online Probability Panel* (IOPP), finalizzato a garantire la rappresentatività statistica della popolazione residente in Italia di età compresa tra i 18 e i 74 anni. Il disegno dell'indagine è di tipo probabilistico, con un piano di campionamento stratificato a più stadi, con selezione dei comuni come unità di primo stadio e degli individui come unità di secondo stadio. La stratificazione è stata realizzata secondo criteri geografici e demografici: per i comuni, in base alla regione, all'ampiezza della popolazione residente e al loro ruolo amministrativo; per gli individui, in base al sesso e alla classe di età. Il campione per IOPP è stato progettato bilanciando esigenze di precisione statistica, vincoli logistici e coerenza con la struttura della popolazione reale, in modo da garantire la possibilità di effettuare analisi inferenziali robuste e comparazioni affidabili, rendendo IOPP un'infrastruttura per la produzione di dati preziosa e affidabile per la ricerca sociale e demografica in Italia.

Il presente capitolo, dopo una breve introduzione al quadro teorico del campionamento nelle scienze sociali, illustra i criteri che hanno guidato la scelta della tipologia di campionamento adottata per IOPP e descrive in modo dettagliato il piano di campionamento implementato, evidenziandone la metodologia e le principali fasi operative.

* Il presente capitolo è il frutto di una riflessione comune da parte dell'autrice e degli autori.

4.2 Il campionamento nella ricerca sociale

Il campionamento rappresenta una delle fasi più decisive nella progettazione della ricerca sociale. Esso consiste nel processo attraverso cui, a partire da una popolazione di riferimento, si seleziona un numero limitato di unità di analisi, denominato campione.

Nello specifico, la popolazione statistica è l'insieme finito di unità (individui, gruppi, enti o eventi) che costituisce l'oggetto dell'indagine. Il campione, al contrario, è un sottoinsieme estratto secondo un procedimento preciso con lo scopo di rappresentare la popolazione rispetto alle variabili di interesse. La rappresentatività è la proprietà che consente di estendere i risultati del campione all'intera popolazione, garantendo che le statistiche campionarie siano stime non distorte e sufficientemente precise dei parametri della popolazione (Lohr 2022). Essa dipende strettamente dal metodo di selezione, dalla numerosità campionaria e dalla struttura della popolazione indagata. Determinare la numerosità ottimale del campione costituisce uno dei passaggi più delicati nella pianificazione di una ricerca. Tale numero deve essere sufficientemente grande da assicurare l'affidabilità dei risultati e abbastanza contenuto da rispettare vincoli di tempo e risorse. La precisione delle stime cresce proporzionalmente alla dimensione del campione, mentre i costi tendono a crescere in modo più che lineare, imponendo la ricerca di un equilibrio tra accuratezza e sostenibilità (Corbetta 2014). È importante sottolineare che la precisione non dipende dalla grandezza della popolazione, ma solo da quella del campione. Inoltre, essa è anche influenzata dalla struttura della popolazione di riferimento: una caratterizzata da un'elevata variabilità può essere rappresentata efficacemente solo attraverso un campione sufficientemente grande da includerne tutte le differenti sfumature. In altri termini, anche popolazioni molto vaste possono essere stimate con un numero relativamente ridotto di unità, purché la selezione sia effettuata in modo corretto (Agresti e Finlay 2009).

Ogni procedura di campionamento comporta inevitabilmente un certo margine di errore, che può essere di natura casuale o sistematica. L'errore casuale deriva dalle fluttuazioni inevitabili dovute al caso: due campioni estratti dalla medesima popolazione possono restituire risultati diversi semplicemente per effetto della variabilità natu-

rale, ovvero della casualità insita nel processo di selezione. Tale casualità genera variabili stocastiche che possono essere studiate nella loro forma matematica e consentono di stimare la probabilità dell'errore. Questo tipo di errore può essere stimato e ridotto sia aumentando la numerosità del campione, sia attraverso una corretta pianificazione del disegno dell'indagine. L'errore sistematico, invece, è prodotto da distorsioni strutturali nel piano di campionamento o nelle procedure di raccolta dati: ad esempio, *bias* di selezione, comportamenti distorti degli intervistatori o risposte tendenziose degli intervistati (Blalock 1989). Mentre l'errore casuale è controllabile mediante strumenti statistici, quello sistematico deve essere prevenuto attraverso un disegno di campionamento accurato e una rigorosa standardizzazione delle tecniche di rilevazione.

Le tecniche di campionamento si distinguono convenzionalmente in due grandi categorie: non probabilistiche, basate su criteri di convenienza, giudizio o disponibilità dei soggetti; oppure probabilistiche, basate su criteri che ne garantiscano la rappresentatività e permettono la generalizzazione dei risultati all'intera popolazione.

Nei campionamenti non probabilistici, quindi, la selezione delle unità avviene sulla base di criteri intenzionali o di convenienza. Sebbene tale approccio possa apparire più semplice e veloce, esso comporta il rischio di ottenere risultati distorti, non generalizzabili all'universo di riferimento, e validi unicamente per la descrizione del campione (Babbie 2021). Questo approccio risulta pertanto più adatto alla produzione di informazioni di tipo qualitativo, che consentono di esplorare i fenomeni indagati ma non ne restituiscono una misura statisticamente rappresentativa. Infatti, nei campioni non probabilistici le probabilità d'inclusione sono ignote: l'errore non è stimabile e non si può parlare, in senso forte, di rappresentatività né di generalizzazione alla popolazione (Baker *et al.* 2013; Elliott e Valliant 2017). Sono possibili interventi di aggiustamento finalizzati al bilanciamento del campione rispetto ad alcune caratteristiche osservabili, che non portano ad una vera rappresentatività statistica (quali, pesi di post-stratificazione, calibrazione, propensione, definibili a diversi livelli) che talvolta riducono la distorsione delle stime, ma poggiano su assunzioni forti e non sostituiscono un disegno probabilistico (Deville e Särndal 1992; Little 2004). In questi casi è più appropriato parla-

re di generalizzazione analitica/trasferibilità a contesti simili, piuttosto che di inferenza dei risultati all'universo sulla base del disegno dell'indagine (Lincoln e Guba 1985; Corbetta 2014; Yin 2018).

Le tecniche non probabilistiche comprendono: il campionamento a scelta ragionata, in cui le unità sono selezionate sulla base della loro rilevanza per l'oggetto di studio; il campionamento per quote, molto usato nei sondaggi di opinione, dove si stabilisce a priori il numero di casi da includere per ciascuna categoria sociale; il campionamento per testimoni privilegiati, che coinvolge esperti del fenomeno indagato; e il campionamento a valanga (*snowball*), utilizzato per studiare popolazioni difficilmente accessibili. Sebbene queste tecniche non permettano inferenze statistiche generalizzabili, esse sono di grande valore in ricerche esplorative o qualitative, dove la profondità dell'analisi prevale sull'estensione della rappresentatività.

Nei campionamenti probabilistici, invece, la rappresentatività del campione è definita dal fatto che ogni unità della popolazione ha una probabilità di inclusione nota e maggiore di 0; ciò consente di associare alle stime un errore campionario e degli intervalli di fiducia e, quindi, la conseguente possibilità di generalizzazione statistica all'intera popolazione (Kish 1965; Lohr 2022). In altri termini, il campione è una porzione osservabile di una realtà più ampia, scelta in modo tale che le osservazioni condotte su di essa consentano di trarre conclusioni affidabili sulla popolazione complessiva. La possibilità di generalizzare le conclusioni di un'indagine a una popolazione più ampia costituisce una delle condizioni fondamentali di scientificità nella ricerca quantitativa (Bryman 2016). Nelle scienze sociali, il campionamento probabilistico rappresenta un pilastro metodologico essenziale perché consente di fondare la generalizzazione dei risultati su basi solide e di stimare la precisione delle inferenze.

Tra i metodi probabilistici più noti, il campionamento casuale semplice rappresenta la forma più pura e teoricamente rigorosa: ogni unità ha la stessa probabilità di essere inclusa, garantendo la massima neutralità nella selezione. Tuttavia, nella pratica della ricerca sociale, la difficoltà di disporre di elenchi completi delle popolazioni rende questo metodo spesso impraticabile. Il campionamento stratificato, che prevede la suddivisione della popolazione in gruppi omogenei (strati) e l'estrazione casuale di unità da ciascuno di essi, migliora la

precisione e assicura una rappresentazione equilibrata anche dei sottogruppi minoritari. Il campionamento sistematico, invece, seleziona le unità a intervalli regolari dopo una prima estrazione casuale, offrendo un buon compromesso tra casualità e semplicità operativa (Lohr 2022). Esistono poi il campionamento a grappoli, in cui si estraggono interi gruppi di unità, generalmente preesistenti rispetto alla ricerca come scuole o altri luoghi di aggregazione, e il campionamento a più stadi, in cui la selezione avviene in fasi successive e la componente casuale può essere inclusa in più di uno stadio, spesso utilizzato nelle indagini statistiche ufficiali.

Particolarmente rilevante è anche il caso dei campionamenti per le indagini panel, che consentono di osservare nel tempo le stesse unità, rendendo possibile lo studio delle dinamiche sociali e dei processi di cambiamento nelle opinioni, nei comportamenti o nelle condizioni di vita. L'uso del campionamento probabilistico è fondamentale nelle indagini panel che di per sé sono affette da una debolezza metodologica dovuta alla caratteristica di intervistare lo stesso collettivo più volte, generando effetti indesiderati di cui si discuterà in seguito. Solo con un campione selezionato in modo probabilistico si possono compensare parzialmente tali effetti perché si può garantire che il panel rappresenti davvero la popolazione e che le variazioni osservate riflettano fenomeni reali e non errori di selezione. Tuttavia, la costruzione di un panel, come IOPP, oltre all'importanza della corretta selezione delle unità tramite il disegno di campionamento iniziale, presenta sfide metodologiche rilevanti legate alla gestione del panel nel tempo, come l'*attrition*, ossia la perdita progressiva dei partecipanti e la conseguente necessità di correggere le distorsioni mediante tecniche di ponderazione o di rotazione dei campioni (Lynn 2009, Saris e Gallhofer 2014) e la mancata o parziale risposta in differenti ondate di rilevazione (Kalton 1986). Nei panel probabilistici, è possibile adottare procedure di compensazione che preservino la continuità del panel senza comprometterne la rappresentatività, garantendo così la validità statistica dei risultati nel tempo, quali la sostituzione dei e delle paneliste persi con unità simili (la similarità in senso statistico deve comunque comprendere la componente probabilistica per la preservazione nel tempo dell'attendibilità delle stime), l'uso di pesi statistici per compensare eventuali squilibri, incentivi e prome-

moria (*panel care*) per favorire la partecipazione e, quando necessario, l'imputazione dei dati mancanti o l'aggiornamento periodico delle informazioni dei e delle paneliste.

Negli ultimi anni, con la diffusione delle tecnologie digitali, il miglioramento dell'alfabetizzazione informatica e l'allargamento dell'accesso alla digitalizzazione a tutti gli strati di popolazione, si è affermato l'uso della modalità online per le indagini, e ciò anche per il caso dei panel. I panel online rappresentano oggi una nuova frontiera per la ricerca empirica. Essi consistono in gruppi di individui che acconsentono a partecipare periodicamente a indagini online, permettendo di raccogliere dati rapidamente e con costi ridotti. Tuttavia, i panel online sollevano questioni cruciali legate alla rappresentatività e al processo di selezione del campione, poiché spesso si basano su partecipazioni spontanee che includono reclutamenti non probabilistici e sono affetti in modo amplificato da meccanismi di autoselezione. Tali modalità possono aumentare possibili distorsioni legate all'accesso alle tecnologie digitali o alla disponibilità individuale a partecipare alle indagini (Callegaro *et al.* 2014). Per superare questi limiti, si stanno sviluppando panel basati su disegni probabilistici o ibridi (*Online Probability Panel*, OPP, cfr. cap. 2), che integrano registri anagrafici, campionamenti misti e tecniche di riponderazione. Queste soluzioni mirano a garantire che anche nella dimensione online sia preservata la rappresentatività, mantenendo la coerenza con i principi del campionamento probabilistico tradizionale. Gli OPP, come dimostra l'esperienza pionieristica del LISS nei Paesi Bassi, possono garantire dati comparabili ai panel tradizionali, preservando la rappresentatività nel tempo e offrendo la possibilità di analisi longitudinale affidabili (Scherpenzeel e Das 2011).

Il disegno campionario di IOPP si colloca dunque nel solco delle esperienze europee degli OPP, come il LISS Panel nei Paesi Bassi, il GESIS Panel in Germania e l'ELIPSS Panel in Francia (Blom *et al.* 2016). Infatti, IOPP, creato nell'ambito del progetto FOSSR, rispetta i criteri di rappresentatività, combinando i vantaggi della raccolta dati online con la rigosità del campionamento probabilistico. I dettagli sulla sua costruzione e gestione saranno presentati nei paragrafi seguenti.

4.3 La strategia di campionamento per IOPP

L'obiettivo principale del piano di campionamento di IOPP è garantire la rappresentatività statistica della popolazione adulta residente in Italia, rispettando vincoli di efficienza, proporzionalità e copertura territoriale, necessari per la produzione di stime affidabili a livello nazionale e sub-nazionale.

Per la costituzione e messa a sistema di un panel, in generale, è poco realistico pensare di poter disegnare una strategia campionaria che assicuri prefissati livelli di precisione di tutte le stime prodotte. Uno degli elementi di maggiore complessità è legato all'esigenza di produrre stime di parametri riferiti a un numero elevato di domini di studio, sia territoriali che individuali. In tali circostanze, la ricerca di soluzioni ottime per ciascun dominio può contrastare con l'obiettivo di individuare una soluzione ottima generale. Ciò significa che, al fine di determinare la numerosità campionaria minima atta a consentire il calcolo di stime con predeterminati livelli di precisione, in alcuni casi è necessario adottare una soluzione metodologica di compromesso, che tenga conto contemporaneamente di una molteplicità di obiettivi e di vincoli.

Per affrontare questo problema, nel caso di IOPP si è fatto ricorso ad una strategia che pervenisse alla definizione della numerosità campionaria attraverso approssimazioni successive, adottando un'ottica mista basata sia su criteri di costo ed organizzativi, sia sulla possibilità di riferire le stime a livello nazionale e con riferimento a ciascuno dei domini territoriali di interesse.

I due principali vincoli di partenza hanno riguardato la definizione della dimensione complessiva del campione di individui che, si è stabilito, non poteva superare una dimensione di 30.000 unità e del numero di comuni campione coinvolti che, in base a criteri di costo e operativi, non poteva essere superiore a 600. Si è scelto tuttavia, per ragioni di efficienza campionaria, di diminuire il numero di comuni campione a 550, in modo da consentire un buon lavoro di controllo e supervisione. Un ulteriore criterio considerato è stato quello del numero minimo di interviste da effettuare in ciascun comune campione. Con riferimento alla scelta del numero minimo di interviste, si fa presente infatti che esiste una relazione diretta tra il numero minimo

di interviste e la dimensione degli strati, inversamente legata a sua volta al numero di comuni campione. Di conseguenza, il numero minimo di interviste non può essere fissato solamente in base a criteri di efficienza delle stime, ma è necessario tenere anche conto di questioni di costo. Il piano di campionamento prescelto ha previsto una allocazione di compromesso ed un numero minimo di interviste per comune pari a 15. Tale scelta ha consentito da un lato di garantire stime affidabili sia al livello di tutti gli strati e le ripartizioni geografiche che al livello dell'intero territorio nazionale, dall'altro di mantenere il numero totale di comuni campione pari a 550.

In particolare, ricalcando il disegno campionario di altre indagini condotte in Italia e assimilabili a IOPP rispetto agli obiettivi di ricerca e/o all'ambito di riferimento (si veda, ad esempio, il ciclo di indagini multiscopo *Famiglie e Soggetti Sociali* di Istat), la strategia di campionamento impiegata è stata di tipo probabilistico a più stadi, pur avendo a disposizione l'intera lista di individui componenti la popolazione di riferimento in quanto i dati sono stati estratti dall'Anagrafe Nazionale della Popolazione Residente (ANPR) che costituisce l'elenco aggiornato e unificato della popolazione residente in Italia. Tuttavia, le ragioni pratiche di reale applicabilità della strategia di campionamento hanno richiesto di optare per una concentrazione in alcune località della selezione degli intervistati, anziché procedere con un campionamento casuale semplice che, oltre a comportare costi elevati, avrebbe potuto non rappresentare tutte le caratteristiche di interesse che invece con la stratificazione vengono garantite. In particolare, il disegno di campionamento per IOPP si è articolato su due livelli: 1) la selezione dei comuni (unità primarie), e 2) la selezione finale degli individui (unità secondarie). La stratificazione ha interessato sia le unità primarie sia quelle secondarie, ed è stata realizzata secondo criteri geografici e demografici: per i comuni, in base alla regione, all'ampiezza della popolazione residente e al loro ruolo amministrativo (centro/periferia di un'area metropolitana piuttosto che capoluogo di regione); per gli individui, in base al sesso e alla classe di età (18-29, 30-39, 40-49, 50-59, 60-74).

Tab. 4.1 – Distribuzione per ripartizione geografica e dominio di studio dei comuni italiani e degli individui nella popolazione, nel campione teorico e nel campione effettivo dell'indagine IOPP. Popolazione residente 18-74 anni al 01/01/2023

Ripartizione	Dominio	Popolazione		Campione teorico		Campione effettivo	
		Comuni	Individui	Comuni	Individui	Comuni	Individui
Nord-ovest	A1	3	1.981.510	3	1.392	3	521
	A2	113	1.324.651	19	874	19	323
	B1	2.655	4.313.890	59	2.773	59	1.062
	B2	220	3.744.865	56	2.576	56	952
	B3	1	23.268	1	73	1	37
Nord-est	A1	2	457.960	2	321	2	120
	A2	47	560.690	9	414	9	153
	B1	1.109	2.759.026	38	1.786	38	684
	B2	229	4.228.591	64	2.944	64	1.088
	B3	3	299.663	3	384	3	144
Centro	A1	2	2.245.980	2	1.579	2	592
	A2	72	1.050.244	16	736	16	272
	B1	726	1.557.904	22	1.034	22	396
	B2	166	3.412.148	53	2.438	53	901
	B3	2	186.359	2	246	2	92
Sud	A1	2	898.423	2	630	2	236
	A2	52	1.162.526	20	920	20	340
	B1	1.475	2.770.364	41	1.927	41	738
	B2	250	4.817.718	73	3.358	73	1.241
	B3	4	194.951	4	363	4	152
Isole	A1	3	780.353	3	548	3	204
	A2	71	828.567	12	552	12	204
	B1	601	1.140.520	16	752	16	288
	B2	93	1.913.712	30	1.380	30	510
	B3	0	0	0	0	0	0

Nota: A1 = comune centro dell'area metropolitana; A2 = comune periferia dell'area metropolitana; B1 = comune non appartenente all'area metropolitana fino a 10.000 abitanti; B2 = comune non appartenente all'area metropolitana con oltre 10.000 abitanti; B3 = comune capoluogo di regione.

La popolazione di riferimento, proveniente dai dati Istat al 1° gennaio 2023, è stata quindi suddivisa in domini di studio comprendenti le cinque ripartizioni geografiche, le regioni (con separazione tra le province autonome di Bolzano e Trento) e la tipologia di ampiezza comunale articolata in cinque classi e derivante dalla combinazione dei criteri sopradescritti (centro di area metropolitana, periferia di

area metropolitana, comune fino a 10.000 abitanti, comune con oltre 10.000 abitanti, capoluogo di regione). Al fine di assicurare un'adeguata numerosità campionaria anche nei domini a bassa densità, è stato introdotto un sovra-campionamento dei capoluoghi di regione, che ha consentito di ottenere stime più stabili in contesti con rischi di sottorappresentazione. La distribuzione riportata nella tabella 4.1 evidenzia per l'appunto come il campione iniziale abbia permesso di assicurare un'adeguata rappresentatività delle diverse ripartizioni geografiche e dei domini di studio, tenendo conto della popolazione *target*, ovvero gli individui residenti in Italia di età compresa tra 18 e 74 anni.

La numerosità campionaria finale prevista per il panel è pari, come detto, a circa 7.500 individui, estratti da una base teorica di 30.000 nominativi, distribuiti in 550 comuni campione, secondo una selezione effettuata con metodo casuale sistematico per i comuni e casuale semplice per gli individui, nel rispetto di criteri di autoponderazione e proporzionalità interna agli strati. Tale schema assicura, in breve, una buona dispersione spaziale del campione, compatibile al tempo stesso con le esigenze di copertura statistica e contenimento dei costi operativi.

4.4 Riflessioni conclusive

Il campionamento rappresenta una fase fondamentale della ricerca sociale, poiché determina la qualità e l'affidabilità dei risultati di un'indagine, e la possibilità di generalizzarli alla popolazione di riferimento. La sua corretta progettazione permette di trasformare dati empirici in conoscenza scientifica, ponendo le basi per inferenze affidabili. Nella misura in cui la ricerca sociale aspira a formulare conclusioni generalizzabili, il campionamento probabilistico rimane il fondamento metodologico su cui costruire ogni processo di indagine.

Oggi, le trasformazioni digitali e l'evoluzione delle società contemporanee pongono sempre nuove sfide ai ricercatori, richiedendo aggiornamenti continui dei metodi di selezione del campione per la realizzazione di *survey* e, ancora di più, quando queste indagini

sono condotte tramite i panel. Tuttavia, tali cambiamenti non mettono in discussione il principio fondamentale della rappresentatività.

Il piano di campionamento adottato per IOPP è stato progettato proprio per rispettare questo principio. La selezione del campione garantisce la rappresentatività rispetto alla popolazione di riferimento, ossia la capacità del campione di riflettere fedelmente la composizione reale della popolazione in relazione alle variabili considerate nel disegno di campionamento per la costruzione del panel. L'approccio utilizzato consentirà di effettuare inferenze statistiche affidabili e di generalizzare i risultati dal campione all'intera popolazione. Infatti, grazie ai dati delle rilevazioni basate su disegni di tipo probabilistico, è possibile conoscere le reali caratteristiche della popolazione di riferimento, in termini di dimensione, distribuzione geografica e composizione demografica, e applicare, se necessario, aggiustamenti di ponderazione per correggere eventuali discrepanze tra campione e popolazione. In questo modo, IOPP assicura la correttezza e l'affidabilità delle stime, offrendo uno strumento solido per la ricerca sociale che unisce la flessibilità dei panel online alla rigorosità metodologica del campionamento probabilistico. In ogni caso si tratta di un campionamento incrementale, proprio per il fatto di avere la disponibilità di un panel di individui da intervistare più volte e per garantire la rappresentatività e l'affidabilità delle stime nel tempo. Come è accaduto nelle esperienze internazionali già riferite, la costruzione di un panel di tale rilevanza non può che essere graduale, ma nel tempo, monitorando sempre la caratteristica di panel probabilistico con opportuni strumenti statistici, si possono raggiungere obiettivi conoscitivi sempre più ambiziosi.

4.5 Riferimenti bibliografici

Agresti A., Finlay B., 2009, *Statistical Methods for the Social Sciences (4th ed.)*, London, Pearson.

Babbie E., 2021, *The practice of social research (15th ed.)*, Cengage Learning.

Baker R., Brick J. M., Bates N. A., Battaglia M., Couper M. P., Dever J. A., Tourangeau R., 2013, *Summary report of the AAPOR task force on non-*

probability sampling, in *Journal of survey statistics and methodology*, 1(2), 90-143.

Blalock H. M., 1989, *Social statistics (Rev. ed.)*, New York, McGraw-Hill.

Blom A. G., Bosnjak M., Cornilleau A., Cousteaux A. S., Das M., Douhou S., Krieger U., 2016, *A comparison of four probability-based online and mixed-mode panels in Europe*, in *Social Science Computer Review*, 34(1), 8-25.

Bryman A., 2016, *Social research methods (5th ed.)*, Oxford, University Press.

Callegaro M., Baker R., Bethlehem J., Göritz A. S., Krosnick J. A., Lavrakas P. J., (Eds.), 2014, *Online panel research: A data quality perspective*, New York, Wiley.

Corbetta P. G., 2014, *Metodologia e tecniche della ricerca sociale (2ª ed.)*, Bologna, Il Mulino.

Deville J. C., Särndal C. E., 1992, *Calibration estimators in survey sampling*, in *Journal of the American statistical Association*, 87(418), 376-382.

Elliott M. R., Valliant, R., 2017, *Inference for nonprobability samples*, in *Statistical Science*, 32 (2), 249-264, <https://doi.org/10.1214/16-STS598>.

Kalton G., 1986, *Handling wave nonresponse in panel surveys*, in *Journal of Official Statistics*, 2(3), 303-314.

Kish L., 1965, *Survey sampling*, New York, John Wiley & Sons.

Lincoln Y. S., Guba E. G., 1985, *Naturalistic inquiry*, Vol. 75, Sage.

Lohr S., 2022, *Sampling: Design and analysis (3rd ed.)*, London, Chapman & Hall/CRC.

Little R. J., 2004, *To model or not to model? Competing modes of inference for finite population sampling*, in *Journal of the American Statistical Association*, 99(466), 546-556.

Lynn P., (Ed.), 2009, *Methodology of longitudinal surveys*, New York, Wiley.

Saris W. E., Gallhofer I. N., 2014, *Design, evaluation, and analysis of questionnaires for survey research (2nd ed.)*, New York, Wiley.

Scherpenzeel A. C., Das M., 2011, *True Longitudinal and Probability-Based Internet Panels: Evidence from the Netherlands*, in M. Das, P. Ester, L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet (77-104)*, London, Routledge.

Yin R. K., 2018, *Case study research and applications (Vol. 6)*, Thousand Oaks, CA, Sage.

5. Reclutamento e Mantenimento

di Francesco Piacentini, Claudia Pennacchiotti e Michele Santurro*

5.1 Introduzione

I panel probabilistici costituiscono una risorsa fondamentale per la ricerca sociale, economica e politica contemporanea. Offrono una infrastruttura stabile per studi longitudinali e comparativi, fornendo dati di alta qualità, rappresentativi della popolazione di riferimento. Tuttavia, il loro successo metodologico dipende in modo cruciale da due fattori interdipendenti: un reclutamento accurato e rappresentativo, e una strategia efficace di mantenimento nel tempo.

Nel contesto italiano, l'*Italian Online Probability Panel* (IOPP) si propone come uno dei primi esperimenti strutturati di panel probabilistico con approccio misto (*online/offline*), ispirato alle migliori esperienze europee. Il disegno adottato integra la raccolta dati *web-based* con tecniche CAWI (*Computer-Assisted Web Interviewing*) e CAPI (*Computer-Assisted Personal Interviewing*), cercando di superare il *digital divide* e garantire l'inclusione anche degli individui con scarse competenze digitali o senza accesso a Internet, seppure applicando una policy di *push-to-web* in modo da massimizzare le rilevazioni autonome via dispositivo.

In questo capitolo vengono approfonditi i principali aspetti teorici e operativi legati alla costruzione e al mantenimento di un panel probabilistico: le modalità di reclutamento, il problema della non copertura, l'utilizzo di incentivi, le tecniche di fidelizzazione e il ruolo del-

* Il presente capitolo è il frutto di una riflessione comune da parte dell'autrice e degli autori.

le tecnologie di gestione contatti. Si presentano inoltre esempi internazionali e indicazioni tratte dalla letteratura scientifica che possono fungere da riferimento per il disegno e la conduzione di panel di alta qualità.

5.2 Il reclutamento nei panel probabilistici

Come già sottolineato nel capitolo precedente, un panel probabilistico, a differenza dei panel *opt-in* o basati su volontari, si fonda su un disegno campionario basato su probabilità note e positive di selezione. In poche parole, ogni membro dell'universo di riferimento deve avere le stesse possibilità di essere selezionato all'interno del campione probabilistico. Questo permette, almeno in teoria, di calcolare stime imparziali e di correggere i *bias* tramite l'uso di pesi campionari, prerequisiti essenziali per la validità inferenziale di qualsiasi indagine (Groves *et al.* 2011).

Tuttavia, la sola probabilità di selezione non è sufficiente a garantire la rappresentatività: essa deve essere accompagnata da alti tassi di risposta iniziale e da una strategia efficace di contatto. I panel più affidabili, come l'olandese LISS (*Longitudinal Internet Studies for the Social Sciences*) o il tedesco GESIS, si caratterizzano per investimenti considerevoli nella fase di reclutamento, sapendo che un buon inizio è cruciale per evitare distorsioni sistematiche a lungo termine.

Il reclutamento in un panel probabilistico richiede in genere l'invio di un invito formale personalizzato, spesso per posta, contenente: una lettera di presentazione del progetto, un *QR code* o codice personale per accedere al questionario online, materiale informativo che spiega i benefici della partecipazione e le garanzie etiche. L'utilizzo del *push-to-web*, una tecnica ampiamente sperimentata (Dillman *et al.* 2014), ha lo scopo di orientare i e le rispondenti verso la modalità online, più economica e flessibile. Tuttavia, la sola modalità web rischia di generare una copertura diseguale: per questo motivo si ricorre a disegni sequenziali e adattivi, in cui la mancata risposta online attiva tentativi successivi in modalità alternative, come interviste faccia-a-faccia tramite il lavoro di intervistatori e intervista-

trici professioniste, interviste telefoniche o autocompilazione di questionari cartacei inviati via posta. Il percorso di reclutamento messo a punto per l'IOPP aderisce alle pratiche standard implementate dai principali panel europei. Una volta conclusa la prassi di campionamento e quindi ottenuta una lista di contatti, essa viene caricata sul *software* di gestione. Un numero personale identificativo (PID) viene creato per ogni nominativo da un responsabile esterno, che procederà successivamente al contatto dei singoli nominativi del campione estratto.

La prassi prevede poi la spedizione di una lettera via posta all'indirizzo di residenza del singolo contatto. Essa contiene tutte le informazioni e gli strumenti per un corretto percorso di adesione al panel. Per quanto riguarda la parte informativa, ogni contatto riceve la descrizione del progetto e una *brochure* che ne illustra la natura, il trattamento dei dati e le modalità di partecipazione. Inoltre, per l'adesione, ogni lettera contiene *link* per accedere al questionario di reclutamento, *QR code* e PID personale, in modo che, una volta ricevuta la lettera, il nominativo possa accedere con il proprio PID all'area riservata della pagina dedicata IOPP per procedere alla registrazione all'indagine.

Nella modalità CAWI, il o la intervistata inserirà dunque l'identificativo univoco (PID) fornito e dovrà indicare di aver letto l'informativa sul trattamento dei dati personali e di acconsentire al trattamento degli stessi per procedere alla partecipazione all'indagine.

Una volta completato il percorso di reclutamento con questa modalità, il contatto diventa ufficialmente membro del panel in modalità online. È però ben noto sia in letteratura che nell'esperienza diretta di reclutamento dei principali panel probabilistici europei che questa modalità di adesione autonoma viene portata a termine da una percentuale più o meno grande a seconda del contesto, ma mai dall'intero campione.

È dunque necessario elaborare una strategia per integrare anche coloro che per i motivi più disparati, che vanno dalla mera dimenticanza all'esistenza di barriere tecnologiche o personali, non fanno parte del gruppo degli *onliners*.

5.3 Strategie per l'inclusione degli *offliners*

Nel contesto attuale, il *digital divide* è uno dei principali ostacoli alla costruzione di panel online veramente rappresentativi. In Italia, secondo i dati Istat (2023), circa il 17% delle famiglie non ha accesso a Internet da casa, con picchi maggiori nel Sud e tra gli anziani. Il divario digitale non riguarda solo l'accesso materiale alla tecnologia, ma anche le competenze digitali, spesso scarse tra le persone con basso titolo di studio.

Questa disuguaglianza produce, nei campioni con sola compilazione via web, un *bias* di copertura, con una sovra-rappresentazione di persone giovani, istruite e urbanizzate, tendenzialmente aventi più dimestichezza con la tecnologia. Come sottolineato in letteratura, ciò non solo rende le stime meno generalizzabili e dunque meno affidabili, ma può compromettere l'analisi di interi fenomeni sociali, dato lo squilibrio del campione statistico (Couper 2000). Ecco quindi che, nonostante le compilazioni CAWI portino con sé una serie di vantaggi (uno su tutti il costo inferiore rispetto a modalità differenti, ma anche la minore incidenza di problematiche quali per esempio il cosiddetto *bias* di desiderabilità sociale, più alto in interviste faccia-a-faccia), l'inclusione di un protocollo di inclusione degli *offliners* diventa essenziale per la buona riuscita di un panel probabilistico, seppur online.

Diversi panel probabilistici hanno elaborato soluzioni per garantire l'inclusione degli *offliners*. Il LISS, per esempio, fornisce dispositivi e connessione Internet gratuita ai e alle partecipanti che ne sono sprovvisti, abbattendo la barriera tecnologica (Scherpenzeel 2011). Simile la strategia adottata dal panel francese ELIPSS, che ha distribuito tablet preconfigurati con SIM dati annessa in modo da consentire una partecipazione online anche in assenza di Wi-Fi domestico o in zone non ancora raggiunte o raggiunte solo in parte da una connessione stabile e di buona qualità (Cornilleau *et al.* 2014). Il GESIS Panel tedesco ha invece mantenuto l'opzione di risposta cartacea per i meno digitalizzati, aumentando sensibilmente i tassi di risposta tra gli e le over 65 (Bosnjak *et al.* 2018). Queste strategie, sebbene abbiano costi logistici importanti, risultano essenziali per ridurre la di-

storsione sistematica del campione e per rispettare i principi di equità e inclusione.

Per quanto riguarda l'implementazione della strategia per l'inclusione degli e delle *offliners*, IOPP prevede che coloro i quali non accederanno entro il periodo di tempo definito alla pagina online dedicata per procedere con l'iscrizione all'indagine saranno raggiunti da un intervistatore o un'intervistatrice per svolgere l'intervista in modalità CAPI. In tale modalità, una volta stabilito il contatto con l'intervistato o l'intervistata, e solo dopo la lettura dell'informativa messa a disposizione del medesimo e la registrazione del suo consenso, si procederà alla somministrazione del questionario inserendo il PID dell'intervistato. Alla fine del questionario, verrà richiesto un ulteriore consenso informato ad entrare nel panel IOPP. I dati delle e degli intervistati che non acconsentano a entrare nel panel, così come i nominativi non rintracciabili e di chi rifiuta l'intervista di reclutamento, sono cancellati dalla lista di campionamento. I tentativi di ricontatto previsti si limitano a un massimo di sei, dopo i quali il contatto viene considerato perso e ugualmente cancellato dalla lista di campionamento.

Va però sottolineato come, anche a fronte di un reclutamento da *offliner*, i e le partecipanti al panel non siano in alcun modo vincolati a tale posizione. Questo per due ordini di motivi: da una parte, IOPP si prefissa di essere un panel probabilistico online, di conseguenza predilige la massimizzazione della percentuale di *onliners* sul totale del campione; dall'altra, i costi di gestione di un *offliner* non sono paragonabili a quelli di un o una rispondente che compila le indagini in modalità CAWI. Per questo, IOPP prevede, dal lato del reclutamento, una politica di *push-to-web* da parte degli intervistatori e delle intervistatrici stesse che favorisca l'adesione come *onliner* anche degli individui che hanno usufruito della modalità di reclutamento CAPI.

Inoltre, in qualsiasi momento del proprio percorso, ogni panelista *offliner* può richiedere di passare dalla modalità PAPI (ovvero il completamento di questionari cartacei inviati via posta) alla modalità di risposta CAWI.

5.4 Incentivi e motivazione alla partecipazione

Gli incentivi rappresentano una leva fondamentale per stimolare la partecipazione a indagini *survey*. Siano essi monetari o simbolici, la letteratura specializzata ha attribuito loro un ruolo centrale, sottolineando non solo l'effetto diretto sul tasso di risposta, ma anche la complessa relazione tra modalità, tempistica, entità e percezione dell'incentivo. Due assi fondamentali distinguono le tipologie di incentivi: condizionalità (offerta prima o dopo la partecipazione) e natura (monetaria o simbolica).

Una distinzione cruciale riguarda il momento dell'erogazione: gli incentivi incondizionati sono forniti prima della compilazione del questionario, mentre quelli condizionati sono erogati solamente dopo il completamento dell'intervista. Secondo Singer e Ye (2013), gli incentivi incondizionati tendono a essere più efficaci nel promuovere la partecipazione iniziale, in quanto attivano un meccanismo di reciprocità psicologica: chi riceve un dono è più incline a restituire il favore, anche solo per motivi morali o sociali. Questa forma di innesco comportamentale è particolarmente utile nelle prime fasi di reclutamento, quando la fiducia nel committente della ricerca non è ancora consolidata. Al contrario, gli incentivi condizionati agiscono secondo una logica di scambio esplicito: si riceve una ricompensa solo se si porta a termine il compito richiesto. Questo approccio è spesso preferito per contenere i costi e prevenire casi di *drop-out* dopo la ricezione del premio. Tuttavia, diversi studi mostrano che gli incentivi post-pagati, pur essendo meno efficaci nel promuovere la risposta iniziale, possono comunque mantenere l'*engagement* in contesti di panel già consolidati, dove il o la partecipante ha sviluppato un certo livello di fiducia e abitudine alla partecipazione (Church 1993).

Dal punto di vista della natura dell'incentivo, si distinguono quelli monetari (denaro diretto, buoni regalo, *voucher* elettronici) da quelli simbolici (*gadget*, ringraziamenti personalizzati, *badge* digitali, accesso a informazioni esclusive). I primi hanno un impatto generalmente più forte sul comportamento del rispondente, soprattutto nei gruppi a bassa motivazione o a basso reddito. Tuttavia, la percezione del valore gioca un ruolo importante anche per gli incentivi simbolici che, se ben progettati, possono rinforzare il senso di appartenenza e

la motivazione intrinseca. Siano *newsletter* periodiche, inviti a eventi pubblici o l'invio di piccoli *gadget*, anche gli incentivi simbolici recano un impatto positivo in termini di *engagement*.

Un ulteriore aspetto centrale riguarda la quantità dell'incentivo monetario. Diverse meta-analisi (una tra tante: Göritz 2006) mostrano che esiste una relazione positiva ma decrescente tra entità dell'incentivo e tasso di risposta: incentivi più alti producono maggiori risposte, ma con rendimento marginale decrescente. Ad esempio, passare da 0 a 5 euro produce un aumento molto marcato della partecipazione, mentre passare da 20 a 25 euro genera un incremento marginale quasi trascurabile. Generalmente, le principali strategie sono due: optare per una somma fissa relativamente contenuta per ogni intervista, calibrata sul *budget* disponibile, oppure la combinazione tra incentivi fissi per ogni *wave* e premi fedeltà cumulativi, che rafforzano la partecipazione nel lungo periodo. Tale strategia ha effetti positivi sia sulla continuità di partecipazione che sulla qualità delle risposte, in quanto incentiva la risposta regolare senza generare aspettative inflazionate (Scherpenzeel e Toepoel 2012).

È stato inoltre dimostrato che, soprattutto in contesti con alta motivazione intrinseca, anche incentivi moderati possono risultare sostenibili e sufficientemente adeguati. In ogni caso, l'efficacia di ogni incentivo si basa su una buona capacità di comunicazione sia in sede di reclutamento che di panel *care*, unita a una solida sensazione di sicurezza da parte dell'intervistato soprattutto per quanto riguarda anonimato e sicurezza dei dati.

Alla luce di questa lunga serie di risultati in letteratura, nel caso di IOPP si è scelto di implementare una politica di incentivazione mista. Ogni contatto che sceglie di iscriversi al panel riceve un incentivo non condizionale di 15 euro, a cui vengono infine aggiunti ulteriori 10 euro questa volta condizionali al completamento dei questionari richiesti. Dunque, il o la panelista che, nel corso del primo anno di somministrazione, abbia svolto adeguatamente il completamento di ogni *wave* somministratagli riceverebbe un totale di 25 euro di incentivo. Come già intuibile alla lettura delle cifre, l'incentivo scelto è di tipo economico, in quanto ampiamente considerato il più efficace. Poiché la legislazione italiana non permette l'elargizione diretta di

somme di denaro ai rispondenti, la somma verrà fornita in forma di buono spesa.

5.5 Fidelizzazione e strategie *anti-attrition*

La panel *attrition* è uno dei problemi metodologici più rilevanti nei panel longitudinali: si tratta della perdita progressiva dei e delle partecipanti nel tempo, ovvero dell'interruzione volontaria o involontaria della partecipazione alle ondate successive di indagine. A differenza della semplice non-risposta occasionale, l'*attrition* implica una uscita sistematica dal panel, e dunque rappresenta una minaccia diretta alla rappresentatività longitudinale del campione, alla validità inferenziale e alla possibilità di analisi dinamiche su base individuale.

Le motivazioni dietro all'*attrition* sono molteplici e si va dalle cause individuali come disinteresse o perdita di motivazione, alle cause strutturali e organizzative, come l'inefficacia degli incentivi, difficoltà tecniche o problemi legati ai questionari. Va però sottolineato che l'*attrition* raramente avviene in modo casuale. Infatti, il fenomeno colpisce più frequentemente alcuni gruppi sociali, in particolare persone con basso interesse civico o politico, minor scolarizzazione, redditi più bassi, livelli inferiori di competenza digitale o *background* migratorio (Lynn 2009). Si tratta, non a caso, delle stesse categorie che risultano già sottorappresentate nei campioni iniziali e che tendono a disconnettersi più facilmente se non adeguatamente seguite.

Questa *attrition* non randomica può portare nel tempo a *bias* selettivi sistematici, compromettendo la validità delle inferenze e rendendo necessario il ricorso a pesature complesse o a strategie di riattivazione mirata. La ricerca metodologica ha prodotto numerose indicazioni per prevenire l'abbandono e mantenere alta la partecipazione nel tempo. Le strategie più efficaci non si limitano a incentivare economicamente, ma mirano a costruire una relazione duratura e personalizzata tra il o la panelista e il progetto, fatta di solleciti mirati, comunicazione puntuale e bidirezionale, sistemi di *gamification* e rafforzamento del senso di appartenenza.

5.6 Etica, trasparenza e sostenibilità

La partecipazione continuativa a un panel probabilistico non è un semplice atto tecnico o logistico, ma implica un rapporto fiduciario profondo tra il o la partecipante e l'ente di ricerca. A differenza delle indagini *una tantum*, la relazione tra panelista e ricercatore si sviluppa nel tempo, e il grado di fiducia instaurato può influenzare significativamente la qualità e la completezza delle risposte. In tal senso, la fiducia epistemica diventa una componente cruciale del successo a lungo termine di ogni panel.

Per costruire e mantenere questo rapporto fiduciario, è essenziale adottare pratiche di comunicazione chiare, coerenti e trasparenti, fin dalla fase di reclutamento. Vanno comunicati con chiarezza gli obiettivi della ricerca, spiegando non solo il tema del questionario, ma anche l'impatto sociale e scientifico atteso. I e le partecipanti devono poter comprendere il "perché" e il "per chi" del loro contributo. Va garantita la protezione e la riservatezza dei dati personali, in conformità con il GDPR e altre normative nazionali.

È fondamentale chiarire quali informazioni vengono raccolte, come vengono conservate e chi potrà accedervi. Come già sostenuto in questo capitolo, la percezione di sicurezza è un fattore determinante per la fiducia, così come la percezione di libertà e volontarietà di partecipazione. Infatti, offrire un meccanismo semplice e trasparente per il recesso è essenziale affinché il o la panelista non si senta vincolato a tal punto da perdere interesse e intaccare la qualità delle risposte fornite. Il rapporto di fiducia viene inoltre cementato dalla divulgazione di *report* infografici e aggiornamenti regolari sull'utilizzo dei dati, in modo da aumentare la consapevolezza dell'utilità dello sforzo del o della singola panelista.

Oltre all'aspetto etico e relazionale, un panel probabilistico deve affrontare anche sfide legate alla sostenibilità economica. Il mantenimento di un panel comporta costi ricorrenti e non trascurabili, che includono gli incentivi monetari o simbolici per i e le partecipanti, le spese di contatto (invii postali, SMS, supporto telefonico), la gestione tecnica delle piattaforme di raccolta dati, il monitoraggio della qualità, le pesature, *data backup*, assistenza tecnica e formazione del personale. Per questo motivo, molti panel pubblici o accademici

adottano modelli di finanziamento misto, combinando risorse da università, enti pubblici, ONG e agenzie europee.

Per garantire la sostenibilità economica nel medio-lungo periodo, molti panel probabilistici, incluso IOPP, adottano il modello delle *call for modules*: inviti pubblici rivolti a ricercatori esterni per proporre moduli di indagine da somministrare al panel, spesso a fronte di un contributo economico che aiuta a coprire i costi operativi e di incentivazione. Questa strategia consente di ampliare l'uso scientifico del panel, mantenendo al contempo attiva e finanziariamente solida l'infrastruttura.

Infine, forse l'aspetto più rilevante è la capacità di produrre valore sociale e scientifico visibile, che consenta al panel di attrarre nuovi investimenti e collaborazioni nel tempo. La sostenibilità, dunque, non è solo una questione contabile, ma si basa anche sulla credibilità, la visibilità e l'impatto del progetto sulla comunità scientifica e sulla società in generale.

5.7 Riflessioni conclusive

La costruzione e il mantenimento di un panel probabilistico di alta qualità come IOPP richiedono un equilibrio complesso tra rigore scientifico, innovazione metodologica e sostenibilità operativa. L'adozione di strategie miste di reclutamento, come l'approccio *push-to-web* integrato con interviste faccia-a-faccia, consente di raggiungere in modo efficace ampie fasce della popolazione, compresi gli individui meno digitalizzati, affrontando così il problema strutturale della sottorappresentazione degli e delle *offliners*. L'esperienza dei principali panel europei conferma che l'inclusione di queste categorie è essenziale per garantire la rappresentatività e la validità inferenziale dei dati raccolti.

Fondamentale è anche la progettazione di un sistema di incentivi efficace, che combini componenti condizionali e non-condizionali, pre e post-pagate, calibrate in base agli obiettivi e al pubblico. La letteratura evidenzia come anche incentivi modesti, se ben strutturati, possano generare significativi miglioramenti nei tassi di risposta e nella fidelizzazione. Tuttavia, nessuna strategia può prescindere dalla

gestione della *panel attrition*, che va contrastata con azioni mirate: comunicazioni personalizzate, *reminders*, riconoscimenti simbolici e meccanismi flessibili di riattivazione. Segmentare la comunicazione e curare la relazione con ciascun panelista in modo proattivo si dimostra cruciale per il mantenimento nel tempo.

Al centro di tutto ciò vi è la dimensione etica del panel: il rispetto per i e le partecipanti, la chiarezza degli scopi, la trasparenza nella gestione dei dati e la possibilità di recesso devono essere garantiti con continuità. In quest'ottica, anche la sostenibilità economica va intesa non solo come questione contabile, ma come condizione per il rispetto di standard elevati: l'adozione di *call for modules* rivolte alla comunità scientifica rappresenta uno strumento strategico per mantenere vivo il panel e al contempo promuoverne l'uso condiviso. In sintesi, la qualità di un panel probabilistico si misura nella capacità di integrare in modo coerente reclutamento, inclusione, incentivazione, fidelizzazione ed etica, in una visione sistemica, trasparente e scientificamente solida, capace di evolversi insieme alla società che intende rappresentare.

5.8 Riferimenti bibliografici

- Bosnjak M., Dannwolf T., Enderle T., Schaurer I., Struminskaya B., Tanner A., Weyandt K. W., 2018, *Establishing an open probability-based mixed-mode panel of the general population in Germany: The GESIS panel*, in Social Science Computer Review, 36(1), 103-115.
- Church A. H., 1993, *Estimating the effect of incentives on mail survey response rates: A meta-analysis*, in Public opinion quarterly, 57(1), 62-79.
- Cornilleau A., Cousteaux A. S., Legleye S., Razakamanana N., 2014, *Le recrutement du panel ELIPSS*, Doctoral dissertation, Centre de données socio-politiques.
- Couper M. P., 2000, *Web surveys: A review of issues and approaches*, in The public opinion quarterly, 64(4), 464-494.
- Dillman D. A., Smyth J. D., Christian L. M., 2014, *Internet, phone, mail, and mixed-mode surveys: The tailored design method*, Indianapolis, Indiana.
- Görizt A. S., 2006, *Incentives in web studies: Methodological issues and a review*, in International Journal of Internet Science, 1(1), 58-70.
- Groves R. M., Fowler jr F. J., Couper M. P., Lepkowski J. M., Singer E., Tourangeau R., 2011, *Survey methodology*, New York. John Wiley & Sons.
- Lynn P., 2009, *Methodology of Longitudinal Surveys*, New York, Wiley.

Scherpenzeel A., 2011, *Data collection in a probability-based internet panel: how the LISS panel was built and how it can be used*, in *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 109(1), 56-61.

Scherpenzeel A., Toepoel V., 2012, *Recruiting a probability sample for an online panel: Effects of contact mode, incentives, and information*, in *Public opinion quarterly*, 76(3), 470-490.

Singer E., Ye C., 2013, *The use and effects of incentives in surveys*, in *The ANNALS of the American Academy of Political and Social*.

6. Core questionnaire

di Michele Santurro, Angela Paparusso, Frank Heins, Caterina Ambrosio e Valentina Tudisca*

6.1 Introduzione

Il presente capitolo illustra il processo di sviluppo del *Core Questionnaire* (CQ) dell'*Italian Online Probability Panel* (IOPP), con particolare attenzione alle scelte metodologiche che ne hanno guidato la progettazione.

Il CQ è uno strumento utile alla raccolta di dati quantitativi su ambiti centrali della vita degli individui, con l'obiettivo di rilevarne caratteristiche, opinioni, valori, atteggiamenti e comportamenti. Esso offre una modalità strutturata e standardizzata di collezione dei dati, che consente di acquisire informazioni su un ampio campione di individui, permettendo di identificare modelli ricorrenti e formulare generalizzazioni riferibili a una determinata popolazione.

Il CQ rappresenta l'elemento cardine della costruzione di una infrastruttura di ricerca fondata su dati panel longitudinali, i quali si basano sul principio per cui le stesse persone vengono intervistate regolarmente nel tempo e le medesime informazioni vengono raccolte periodicamente, condizione imprescindibile per l'analisi sia dei fenomeni persistenti sia del cambiamento individuale e sociale (Lynn 2009; White e Arzi 2005).

* Il presente capitolo è il frutto di una riflessione comune da parte delle autrici e degli autori. In ogni caso, i paragrafi 6.1 e 6.5 sono da attribuire a Caterina Ambrosio e Angela Paparusso, il paragrafo 6.2 a Frank Heins e Angela Paparusso, il paragrafo 6.3 a Michele Santurro, mentre il paragrafo 6.4 a Michele Santurro e Valentina Tudisca.

Il CQ costituisce anche l'ossatura attorno alla quale possono essere aggiunti moduli tematici su specifici ambiti di indagine, che possono essere proposti a intervalli regolari o *una tantum* da gruppi di ricerca esterni attraverso delle procedure competitive.

Inoltre, il CQ rappresenta la componente dell'indagine tendenzialmente non soggetta a modifiche nel corso del tempo, in virtù della sua funzione di garantire la coerenza e la continuità della rilevazione longitudinale. Tuttavia, non si può prescindere dalla possibilità che eventi di portata eccezionale rendano necessarie delle modifiche o la rilevazione di aspetti ulteriori. Un caso esemplare in tal senso è documentato nella *General Social Survey* (GSS) (Davern *et al.* 2021), in cui si evidenzia come dal 1972 il CQ abbia mantenuto una notevole stabilità sia nei temi affrontati che nella formulazione delle domande. Soltanto con l'avvento della pandemia di Covid-19 nel 2020 si è resa opportuna l'introduzione di alcune modifiche, al fine di cogliere le nuove dinamiche sociali in atto. Tale considerazione risulta cruciale per chiunque operi nell'ambito delle indagini panel, in quanto sottolinea la necessità di bilanciare la stabilità dello strumento con la flessibilità dovuta all'evoluzione del contesto sociale di riferimento.

La definizione esatta dei contenuti di un CQ costituisce un'operazione metodologicamente complessa. Non è guidata da scelte teoriche ed empiriche aprioristiche; al contrario, il contesto socioculturale di riferimento gioca un ruolo fondamentale nella sua formulazione. Un esempio è rappresentato dall'*European Social Survey* (ESS), il cui obiettivo principale è analizzare l'evoluzione delle strutture demografiche, sociali, politiche e valoriali all'interno delle società europee (ESS 2025). In coerenza con tali finalità, il CQ dell'ESS include sezioni dedicate alla struttura sociale di base, agli orientamenti valoriali e ideologici degli individui, nonché alle identità culturali e nazionali (ESS 2025). Analogamente, tra gli obiettivi principali del panel gestito dal *GESIS – Leibniz Institute for the Social Sciences*, si annovera la rilevazione nel tempo di fenomeni di interesse duraturo per le scienze sociali (GESIS 2025). In tale ottica, le sezioni del CQ sono progettate per rispondere alle esigenze analitiche tipiche della ricerca longitudinale: la continuità tematica, che consente l'osservazione sistematica dei processi sociali nel lungo periodo, e la possi-

bilità di legare i cambiamenti individuali al contesto sociale di riferimento (GESIS 2025).

In sintesi, alla luce delle considerazioni fin qui esposte, si ritiene che un CQ destinato a un'indagine panel longitudinale debba includere, come struttura minima essenziale, i seguenti aspetti: le caratteristiche sociodemografiche e socioeconomiche dell'individuo, i principali tratti della personalità, gli orientamenti valoriali e politici, nonché le percezioni, le opinioni e i comportamenti individuali.

A partire da questa premessa iniziale, nel paragrafo successivo verrà illustrato il processo attraverso cui è stato progettato il CQ di IOPP. Nel terzo paragrafo si presenterà la struttura corrente del CQ, mentre nel quarto verranno analizzate le principali criticità emerse in fase di sviluppo, unitamente alle sfide previste per il futuro. Infine, si offriranno alcune considerazioni conclusive volte a sintetizzare i principali risultati raggiunti.

6.2 Come abbiamo progettato il Core Questionnaire di IOPP

Come si è detto, la costruzione di un CQ rappresenta una fase cruciale nella progettazione di un'indagine panel. In mancanza di una letteratura specifica sull'argomento, il nostro punto di partenza è stato considerare gli aspetti rilevanti per l'individuo sulla base delle principali evidenze scientifiche presenti nella letteratura demografica, sociologica e psicologica (per esempio, Aassve *et al.* 2013; Ajzen 2001; Davis 1985; Nazirova e Borbala 2024). Da tali risultati, sono emersi i temi che sono stati inclusi nel questionario, adottando un approccio il più possibile longitudinale e comparativo, anche nell'ottica di favorire un *Life Course Observatory* a livello italiano (cfr. Introduzione) e di permettere eventuali confronti internazionali.

In particolare, nell'ambito della progettazione del CQ dell'indagine IOPP, abbiamo condotto un'analisi comparativa dei questionari di alcuni panel online per le scienze sociali attualmente attivi in Europa. Tale operazione ha avuto come obiettivo principale quello di verificare che tutti gli aspetti fondamentali della vita degli individui, già in parte considerati nel questionario di reclutamento dell'indagi-

ne, fossero stati adeguatamente considerati. Tale analisi è stata utile anche per assicurare la confrontabilità internazionale di IOPP. Tale obiettivo è stato perseguito non solo per conferire maggiore validità scientifica al nostro strumento, ma anche per aumentarne l'utilizzabilità futura da parte della comunità scientifica nazionale e internazionale, rendendolo più competitivo e spendibile nel contesto della ricerca sociale.

In particolare, l'analisi ha riguardato il *LISS Panel* (Paesi Bassi), il *Belgian Social Study* (Belgio), il *GESIS Panel* (Germania), l'*UK Household Longitudinal Study* (Regno Unito) e l'*ELIPSS* (Francia). Il confronto si è concentrato sui nuclei fondamentali di domande del CQ di ciascun panel, con un'attenzione specifica ai moduli sociodemografici, socioeconomici e a quelli relativi agli orientamenti politici e valoriali degli individui e delle loro famiglie.

Sempre in quest'ottica, alcune batterie di domande sono state scelte appositamente per consentire l'allineamento con altre infrastrutture di ricerca europee, come ad esempio *Generations and Gender Survey* (GGS) per gli *item* relativi alle intenzioni riproduttive e alla divisione del lavoro domestico, *Survey of Health, Ageing and Retirement in Europe* (SHARE) per quelli relativi alla salute e alle malattie croniche, e *Growing Up In Digital Europe: Euro Cohort* (GUIDE) per il benessere soggettivo e le condizioni di vita delle famiglie con bambini. Nello specifico, abbiamo inserito dei quesiti presenti in GGS per agganciarne le informazioni a IOPP. Al contempo, sono stati posti limiti precisi all'integrazione di strumenti esterni, per evitare sovrapposizioni eccessive, come nel caso di SHARE, per le sezioni dedicate alla popolazione *over 55*.

Si è fatto, inoltre, riferimento a indagini trasversali sia internazionali come l'*European Social Survey* (ESS), l'*European Values Study* (EVS), la *European Quality of Life Survey* (EQLS), che nazionali, come *Aspetti della Vita Quotidiana e Famiglie e Soggetti Sociali* dell'Istat e l'*Indagine Bilanci delle famiglie italiane* della Banca d'Italia, nonché l'*Eurobarometro*. Infine, non sono mancati casi in cui sono stati inseriti dei quesiti originali, appositamente sviluppati per IOPP e per il contesto italiano. In tutti i casi, comunque, si è cercato di garantire un livello minimo di approfondimento, lasciando

l'affondo su alcuni temi di interesse specifico a eventuali moduli aggiuntivi.

Il risultato di questo processo empirico è un CQ che include *item* caratterizzati da una o più delle seguenti proprietà: diffusamente utilizzati nelle principali indagini internazionali e riconducibili a tematiche consolidate nella letteratura scientifica; non ridondanti rispetto a strumenti già esistenti o paralleli; utilizzabili come variabili indipendenti o di controllo nelle future analisi statistiche; utili a operare stratificazioni sociodemografiche e socioeconomiche.

6.3 La struttura attuale del Core Questionnaire

In questo paragrafo presentiamo e descriviamo sinteticamente le dimensioni principali del CQ di IOPP. La tabella 6.1 riporta i moduli e il numero di domande previsti per il primo anno di somministrazione. Ovviamente, l'ordine delle dimensioni rispecchia unicamente la loro sequenza progressiva all'interno del CQ e non si riferisce in alcun modo al loro eventuale grado di importanza.

Tab. 6.1 – Moduli e rispettivo numero di domande nel Core Questionnaire di IOPP

Modulo	N. domande
Informazioni sociodemografiche	33
Benessere e salute	18
Situazione economica	24
Famiglia e nucleo familiare	65
Istruzione e formazione	14
Integrazione sociale e caratteristiche del quartiere	28
Uso dei media e tempo libero	30
Religione	6
Atteggiamenti individuali e verso la politica	63
Scienza e tecnologia	19
Totale	300

Il primo modulo riguarda l'aggiornamento e l'integrazione delle informazioni sociodemografiche raccolte attraverso il questionario di reclutamento. Esso si suddivide al suo interno nei seguenti ambiti tematici: caratteristiche sociodemografiche (3 domande), luogo di residenza (14 domande), abitazione (9 domande), cittadinanza/*background* migratorio (7 domande).

Il secondo modulo indaga la sfera del benessere soggettivo e della salute. Queste due dimensioni costituiscono infatti i due principali nuclei attorno a cui si articola la seconda sezione del CQ (per un totale, rispettivamente, di 6 e 12 domande).

Il terzo modulo fa riferimento alla situazione economica personale e familiare, considerata sia sotto il profilo dell'attività – o inattività – lavorativa (14 domande), sia dal punto di vista del reddito (10 domande).

Segue il modulo sulla famiglia e il nucleo familiare, che, in maniera simile rispetto alla sezione iniziale, arricchisce e dettaglia ulteriormente il profilo tracciato attraverso il questionario di reclutamento. Nel dettaglio, il quarto modulo è così composto: dimensione e composizione del nucleo familiare (6 domande), relazioni di coppia e intenzioni di fecondità (18 domande), compiti domestici (13 domande), relazioni con i genitori (13 domande), benessere familiare (15 domande).

Il quinto modulo copre la sfera dell'istruzione e della formazione, intesa sia nella sua dimensione oggettiva, cioè come istruzione e formazione professionale completata (7 domande), che in termini soggettivi, come opinioni e atteggiamenti verso il sistema educativo-scolastico (7 domande).

Il sesto modulo mira a catturare il contesto sociale di appartenenza attraverso alcune domande sulle relazioni e attività sociali (13 domande) e sulle caratteristiche del quartiere di residenza (15 domande).

Il settimo modulo riguarda l'uso dei mezzi di comunicazione e dei *social media* (19 domande), nonché l'uso del tempo libero (11 domande).

Mentre l'ottavo modulo verte sulla religione. Si tratta del modulo che, tra tutti, prevede il minor numero di domande (6).

Quanto al nono modulo, esso fa riferimento alla sfera degli atteggiamenti individuali e verso la politica. Come si può vedere dalla tabella 6.1, esso si colloca al secondo posto per numero di *item* e si ripartisce nel seguente modo: atteggiamenti e valori individuali (11 domande), partecipazione politica (9 domande), voto (12 domande), fiducia (18 domande), soddisfazione per la democrazia e per le istituzioni (12 domande), auto-collocazione sull'asse sinistra-destra (1 domanda).

Il decimo e ultimo modulo si concentra sugli atteggiamenti verso la scienza e la tecnologia. Tale modulo prevede un nucleo di domande che riguardano le relazioni tra scienza, tecnologia e società – dall’impatto sulla vita quotidiana alle istanze di partecipazione pubblica alle decisioni su temi tecno-scientifici – da porre con cadenza biennale (9 domande) e alcuni focus su questioni di stretta attualità quali cambiamento climatico, Intelligenza Artificiale e vaccini, da monitorare con cadenza annuale (10 domande).

Le informazioni rilevate dal CQ sono dunque numerose e caratterizzate da livelli di dettaglio differenti in funzione delle caratteristiche dei panelisti. Inevitabilmente, data la complessità e la molteplicità degli ambiti esplorabili, il numero di *item* definito per ciascuna sezione costituisce il risultato di una serie di compromessi tra esigenze contrapposte: da un lato, la necessità di contenere la lunghezza complessiva del questionario e il conseguente rischio di affaticamento del rispondente (*respondent burden*) che, come noto, può tradursi in risposte meno accurate o in mancate risposte, a scapito di una apparente maggiore analiticità (Gummer e Daikeler 2018); dall’altro, l’esigenza di garantire un adeguato livello di approfondimento dei temi affrontati e di raccogliere informazioni potenzialmente applicabili all’intero campione, limitando il ricorso a filtri estremamente selettivi (cosiddetto *routing*, cfr. Krosnick 2018). A questi elementi si sono aggiunte considerazioni relative alla coerenza tematica interna dei moduli (Marradi 2019) e alla comparabilità con altre indagini nazionali e internazionali (Vandenplas e Lipps 2014).

Il confronto con le altre indagini nazionali e internazionali è stato condotto su un doppio livello al fine di garantire la coerenza del CQ rispetto agli standard in uso negli altri *panel studies* e la sua capacità di cogliere le dimensioni chiave sviluppate dalla ricerca internazionale: in primo luogo, rispetto ai singoli *item* da utilizzare, valutandone formulazione e modalità di risposta, in secondo luogo, considerando l’eventuale sotto-copertura o sovra-copertura dei vari nuclei tematici.

Dato l’ampio numero di *item* selezionati, si poneva la necessità di suddividere in *wave* il CQ, al fine di avere dei questionari che non richiedessero mediamente più di 15 o 20 minuti per essere completati. La letteratura, infatti, evidenzia come nelle indagini CAWI che hanno una durata superiore a tale soglia la qualità delle risposte tenda

progressivamente a diminuire (Crawford *et al.* 2001; Galesic e Bosnjak 2009; Revilla e Höhne 2020). Per calcolare la durata di un singolo questionario, tenuto conto del numero massimo di *wave* a disposizione per il CQ pari a 3 (cfr. cap. 3), sulla base delle evidenze raccolte tramite il questionario di reclutamento e delle indicazioni provenienti dagli altri panel (GESIS 2025), è stata adottata la regola dei 5 *item* al minuto, stimando cioè la possibilità di somministrare in un minuto di tempo circa 5 brevi domande a scelta singola – per un totale, appunto, di un massimo di circa 300 domande in 60 minuti. Ovviamente, è importante notare che tale regola costituisce una stima solo approssimativa per calcolare la durata di un questionario e che la complessità, nonché la lunghezza delle domande possono influenzare significativamente il suo tempo di completamento.

Per un buon bilanciamento rispetto al tipo di domande all'interno di ciascun questionario, il CQ è stato suddiviso in tre macro-aree tematiche:

- caratteristiche individuali (es. età, genere, livello di istruzione);
- caratteristiche socio-strutturali (es. composizione familiare, situazione lavorativa);
- atteggiamenti e comportamenti (es. opinioni politiche, abitudini di consumo).

Aggregando le prime due macro-aree in un unico gruppo, è stata ottenuta quindi la seguente ripartizione:

1. caratteristiche sociografiche: informazioni sociodemografiche, condizioni di salute, situazione economica, famiglia e nucleo familiare, istruzione e formazione, integrazione sociale e caratteristiche del quartiere;
2. atteggiamenti e comportamenti: uso dei media e tempo libero, religione, atteggiamenti individuali e verso la politica, scienza e tecnologia.

Infine, questi due blocchi sono stati distribuiti in maniera trasversale tra le *wave* dedicate al CQ, in modo da avere una presenza equilibra-

ta all'interno di ciascun questionario tra le domande sulle proprietà sociografiche di base dei partecipanti e quelle sui loro atteggiamenti e comportamenti.

Inoltre, per rendere il CQ uno strumento di monitoraggio delle tendenze e del cambiamento sociale, sono state previste sia domande riferite ad aspetti tendenzialmente stabili, destinati a cambiare solo nel medio-lungo periodo (*time-invariant*), sia domande su atteggiamenti e comportamenti legati più o meno strettamente al momento contingente (*time-varying*). Ciò ha implicato l'adozione di differenti periodicità di somministrazione, a seconda della probabile frequenza di cambiamento delle proprietà rilevate (vedi anche Callegaro *et al.* 2014). In particolare, sono stati previsti quattro tipi di somministrazione, anch'essi trasversali rispetto alle *wave*: aggiornamento periodico: caratteristiche di profilazione; somministrazione annuale: atteggiamenti o comportamenti; somministrazione biennale: caratteristiche di status o prospettive; somministrazione unica/condizionata: caratteristiche di status o retrospettive.

6.4 Le principali criticità incontrate e le sfide future da affrontare

La definizione del CQ di IOPP ha comportato il confronto con una serie di criticità metodologiche, legate alla necessità di soddisfare contemporaneamente requisiti di validità scientifica, comparabilità e sostenibilità operativa. Pur facendo ampio riferimento all'esperienza dei principali panel europei, per poter procedere nella selezione degli *item*, è stato necessario innanzitutto definire una serie di criteri guida.

In primo luogo, sono stati privilegiati *item* suscettibili di essere utilizzati come variabili indipendenti o di controllo in un ampio spettro di analisi, anche al di fuori degli obiettivi specifici di IOPP. In una virtuale gerarchia tra le domande, è stata data priorità a quelle largamente diffuse e consolidate nelle principali indagini internazionali, tenendo conto anche delle variabili maggiormente utilizzate nella letteratura scientifica che fa riferimento a tali indagini. Al contrario, sono stati esclusi o quantomeno limitati, come accennato nel paragrafo precedente, *item* eccessivamente specifici o discriminanti, ri-

tenuti più adatti a moduli tematici aggiuntivi. In questa prospettiva, sono stati spesso preferiti ad esempio *item* sugli atteggiamenti rispetto a quelli sui comportamenti, in quanto potenzialmente applicabili a una platea più ampia di rispondenti.

Un ulteriore criterio rilevante ha riguardato la natura degli atteggiamenti rilevati. Ove possibile, si è scelto di includere *item* riferiti ad atteggiamenti di primo livello, diretti verso oggetti o temi specifici (ad esempio l'immigrazione), piuttosto che atteggiamenti di secondo livello, relativi ad aspettative o valutazioni su altri attori (come il ruolo dello Stato in ambiti specifici di *policy*), ritenuti più complessi da interpretare e maggiormente sensibili al contesto.

Dal punto di vista formale, la selezione degli *item* ha dovuto rispondere anche a criteri di armonizzazione delle modalità di risposta. Nello specifico, si è cercato di utilizzare scale omogenee (ad esempio scale di tipo Cantril o Likert semanticamente affini o aventi le medesime modalità di risposta) e di preservare la logica delle domande originali quando organizzate in batterie, evitando suddivisioni in sotto-batterie o l'unione di più batterie di diversa provenienza. Alcuni *item* sono stati inoltre pensati per consentire una valutazione indiretta dell'affidabilità del rispondente (come nel caso delle domande, previste con cadenza annuale, sul peso e l'altezza del rispondente o sul suo comportamento di voto alle ultime elezioni politiche).

Una criticità specifica ha riguardato il trattamento dei quesiti sensibili e delle mancate risposte. In questo ambito, sono stati previsti tre principali approcci: l'utilizzo della modalità "non saprei" come trattamento base; l'introduzione congiunta delle due modalità "non saprei" e "preferisco non rispondere" per le domande eticamente o socialmente sensibili; e, in alcuni casi, l'uso mirato o l'esclusione di tali modalità per garantire la comparabilità con gli *item* originali provenienti da altre indagini. Il trattamento delle non risposte è spesso un aspetto trascurato nella redazione di un questionario. In alcune *survey* capita infatti che venga assegnato alle risposte "non so" e modalità simili un codice intermedio fra quelli assegnati alle risposte positive e quelli assegnati alle risposte negative. Questo uso è frequente nelle scale Likert dove si utilizza l'opzione "non so" come modalità intermedia fra le risposte con cui si esprime un accordo e le risposte con cui si esprime disaccordo, per perdere meno dati possibi-

li. Tale procedura tuttavia risulta criticabile, poiché chi dichiara di non avere un'opinione su un tema non ha in realtà alcuno stato rispetto alla proprietà che si intende rilevare (Marradi 2019).

Un ulteriore nodo problematico ha riguardato il bilanciamento tra l'adattamento degli *item* al contesto italiano e il mantenimento della formulazione originale per i quesiti provenienti da altri panel internazionali (come il LISS e il GESIS), al fine di assicurare la loro comparabilità a livello internazionale. La scelta tra riferimenti nazionali e internazionali è stata valutata caso per caso, privilegiando i primi per temi più specifici del contesto italiano, come le caratteristiche socio-demografiche e familiari, e i secondi rispetto ad ambiti consolidati nella letteratura comparata, quali la fiducia o gli orientamenti valoriali.

La traduzione degli *item* provenienti da indagini internazionali ha rappresentato un ulteriore elemento di attenzione rispetto alla comparabilità di IOPP. Per la definizione della versione finale del questionario, ove disponibili, sono state considerate traduzioni italiane già validate di questionari internazionali. Mentre per i quesiti disponibili unicamente in lingua straniera, si è fatto ricorso ai protocolli adottati nelle principali indagini comparative, in particolare il metodo TRAPD (*Translation, Review, Adjudication, Pretest, Documentation*) utilizzato in ESS e EVS, che prevede il confronto tra traduzioni indipendenti e una successiva fase di revisione condivisa.

Nel loro insieme, i criteri discussi delineano un impianto di selezione degli *item* orientato non tanto a definire uno strumento di indagine esaustivo o definitivo, quanto piuttosto a costruire una base comune sufficientemente stabile e condivisa, consapevole dei compromessi impliciti che ogni progettazione di questo tipo comporta. Le scelte compiute rappresentano una cornice di riferimento e non esauriscono la valutazione della qualità complessiva del CQ, che resta inevitabilmente aperta e suscettibile di revisione.

In un'ottica longitudinale, la validità delle scelte compiute potrà infatti essere pienamente valutata solo attraverso la loro verifica empirica nel tempo. L'osservazione dei comportamenti di risposta dei panelisti e delle paneliste consentirà di analizzare la qualità dei dati prodotti, i tassi di risposta e la stabilità degli indicatori, offrendo elementi utili sia per una valutazione della tenuta dei criteri metodo-

logici adottati sia per il monitoraggio dei risultati in chiave comparata rispetto alle altre indagini prese a riferimento.

6.5 Conclusioni

Il presente capitolo ha ricostruito in modo dettagliato il processo di sviluppo del CQ di IOPP, mettendo in evidenza le scelte metodologiche che ne hanno guidato la progettazione e sottolineando il ruolo cruciale che tale strumento riveste all'interno di un'infrastruttura di ricerca longitudinale.

È stato chiarito come il CQ rappresenti non solo il perno attorno al quale si struttura l'intera rilevazione, ma anche uno strumento centrale per l'osservazione dei processi di cambiamento individuale e sociale nel contesto italiano e in una prospettiva comparata internazionale.

La progettazione ha preso avvio da un'attenta selezione di temi ritenuti centrali per la vita degli individui sulla base della letteratura scientifica internazionale e si è arricchita attraverso un confronto con le esperienze di altri panel europei consolidati, assicurando così un elevato grado di comparabilità e validità scientifica.

Il questionario è stato strutturato in moduli tematici che coprono le principali dimensioni della vita degli individui – dalle caratteristiche sociodemografiche al benessere soggettivo, dalla situazione economica alle relazioni familiari, fino agli atteggiamenti politici e alla fiducia nelle istituzioni. Le domande sono organizzate in tre macro-aree: caratteristiche individuali; caratteristiche socio-strutturali; atteggiamenti e comportamenti. Particolare attenzione è stata dedicata alla sostenibilità del carico di risposta per i partecipanti, mediante una suddivisione del questionario in *wave* e una pianificazione attenta della periodicità di somministrazione degli *item*, a seconda della loro variabilità nel tempo. Questo approccio consente di rilevare, con il giusto grado di profondità e continuità, tanto gli aspetti strutturali e stabili, quanto quelli più contingenti e suscettibili di rapido cambiamento.

Il capitolo ha, inoltre, evidenziato come il CQ, oltre a costituire uno strumento di osservazione longitudinale, sia stato concepito anche come base flessibile per l'integrazione futura di moduli tematici aggiuntivi, proposti da gruppi di ricerca esterni, favorendo così

l'espansione dell'indagine e il coinvolgimento della comunità scientifica. Tale flessibilità, però, è stata pensata in equilibrio con l'esigenza di mantenere una struttura stabile nel tempo, in modo da non compromettere la confrontabilità dei dati raccolti. Questo rientra nel progetto di IOPP come bene collettivo, dagli standard elevati, che incoraggi la partecipazione e la fruizione più ampia possibile della comunità scientifica appartenente alle scienze sociali.

Dal punto di vista operativo, abbiamo visto come la definizione del CQ di IOPP sia stata guidata dalla necessità di bilanciare rigore metodologico, comparabilità internazionale e sostenibilità operativa, affrontando una pluralità di criticità nella selezione degli *item*. Le scelte si sono orientate verso domande consolidate nella letteratura e nelle principali indagini comparative, capaci di fungere da variabili trasversali e applicabili a un'ampia platea di rispondenti, privilegiando atteggiamenti di primo livello e limitando l'utilizzo di *item* troppo specifici o sensibili al contesto. Particolare attenzione è stata dedicata all'armonizzazione delle scale di risposta, al trattamento delle non risposte e alla traduzione degli *item*, al fine di garantire affidabilità e comparabilità.

Il risultato è uno strumento solido, scientificamente fondato e orientato al lungo periodo, che si propone di diventare un punto di riferimento per lo studio della società italiana contemporanea e della sua evoluzione, contribuendo a colmare un'importante lacuna nel panorama nazionale delle indagini panel basate su campioni probabilistici.

6.6 Riferimenti bibliografici

Aassve A., Sironi M., Bassi V., 2013, *Explaining Attitudes Towards Demographic Behaviour*, in *European Sociological Review*, 29, 316-333.

Ajzen I., 2001, *Nature and Operation of Attitudes*, in *Annual Review of Psychology*, 52, 27-58.

Callegaro M., Baker R., Bethlehem J., Göritz A. S., Krosnick J. A., Lavrakas P. J., 2014, *Online panel research: History, concepts, applications and a look at the future*, in Callegaro M., Baker R., Bethlehem J., Göritz A. S., Krosnick J. A., Lavrakas P. J. (a c. di), *Online Panel Research. A Data Quality Perspective*, Chichester, Palgrave Macmillan, 1-22.

Crawford S. D., Couper M. P., Lamias M. J., 2001, *Web Surveys: Perceptions of Burden*, in *Social Science Computer Review*, 19(2), 146-162.

- Davern M., Bautista R., Freese J., Morgan S. L., Smith T. W., 2021, *General Social Survey 1972-2021*, Chicago, NORC at the University of Chicago.
- Davis R. A., 1985, *Social Structure, Belief, Attitude, Intention, and Behavior: A Partial Test of Liska's Revisions*, in *Social Psychology Quarterly*, 48, 89-93.
- European Social Survey, 2025, *Source Questionnaire Development*, <https://www.europeansocialsurvey.org/methodology/source-questionnaire/source-questionnaire-development>, consultato il 5 maggio 2025.
- Galesic M., Bosnjak M., 2009, *Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey*, in *Public Opinion Quarterly*, 73(2), 349-360.
- GESIS – Leibniz Institute for the Social Sciences, 2025, *Documentation GESIS Panel.pop*, <https://www.gesis.org/en/gesis-panel/documentation/documentation-gesis-panelpop>, consultato il 5 maggio 2025.
- Gummer T., Daikeler J., 2018, *A Note on How Prior Survey Experience With Self-Administered Panel Surveys Affects Attrition in Different Modes*, in *Social Science Computer Review*, 38(4), 490-498.
- Krosnick J. A., 2018, *Questionnaire Design*, in Vannette D. L. Krosnick J. A. (a c. di), *The Palgrave Handbook of Survey Research*, Cham, Palgrave Macmillan, 439-455.
- Lynn P., 2009. *Methodology of longitudinal surveys*, Chichester, John Wiley & Sons.
- Marradi A., 2019, *Tutti redigono questionari. Ma è davvero così facile?*, Milano, FrancoAngeli.
- Nazirova Z., Borbala S., 2024, *Values, Attitudes and the Behaviour Paradigm: A Systematic Literature Review*, in *Journal of Human Values*, 30, 214-239.
- Revilla M., Höhne J. K., 2020, *How long do respondents think online surveys should be? New evidence from two online panels in Germany*, in *International Journal of Market Research*, 62(5), 538-545.
- Vandenplas C., Lipps O., 2014, *Robustness of items within and across surveys*, in *FORS Working Paper Series*, 3, 1-26.
- White R. T., Arzi, H. J., 2005, *Longitudinal studies: Designs, validity, practicality, and value*, in *Research in Science Education*, 35, 137-149.

7. Archiviazione e disseminazione

di *Domingo Scisci, Giovanna De Santis e Francesco Piacentini**

7.1 Il ciclo di vita dei dati

Archiviazione e disseminazione dei dati sono spesso considerate come attività finali di uno studio empirico, da svolgere solo a conclusione delle fasi analitiche, talvolta dettate più da vincoli istituzionali che da un'autentica volontà di condivisione. Questo approccio, ancora diffuso nelle scienze sociali, non riconosce pienamente la centralità del Ciclo di vita dei dati (*Data Life Cycle*), un modello che descrive l'intero arco di vita del dato, dalla pianificazione iniziale al possibile riutilizzo.

Il ciclo di vita dei dati non si limita alla conservazione e alla condivisione, ma include anche la progettazione della raccolta, il controllo di qualità, la documentazione e l'interazione con la comunità scientifica. Come affermato da Ball (2012), queste fasi concorrono a definire un approccio sistemico alla gestione dei dati, fondato su qualità, accessibilità e sostenibilità. Vari modelli operativi sono stati sviluppati negli ultimi anni e, pur differendo nei dettagli, tutti convergono sull'obiettivo di garantire dati ben documentati, sicuri, accessibili e riutilizzabili secondo i principi FAIR (*Findable, Accessible, Interoperable, Reusable*; cfr. Wilkinson *et al.* 2016).

* Il presente capitolo è il frutto di una riflessione comune da parte dell'autrice e degli autori.

In termini di redazione, i paragrafi sono stati così ripartiti: il paragrafo 7.1 e 7.5 sono da attribuire congiuntamente a tutti gli autori; i paragrafi 7.2, 7.3.3 e 7.3.5 sono stati redatti da Domingo Scisci; i paragrafi 7.3.1, 7.3.2 e 7.3.4 da Giovanna De Santis; il paragrafo 7.3.6 da Francesco Piacentini. Il paragrafo 7.4 è stato curato congiuntamente da Domingo Scisci e Giovanna De Santis.

Nel contesto delle infrastrutture di ricerca basate su disegni longitudinali e panel, come IOPP, l'adozione di un approccio sistemico al ciclo di vita dei dati risulta particolarmente rilevante. La natura probabilistica e longitudinale del panel rende infatti necessario integrare archiviazione e disseminazione fin dalle fasi iniziali della progettazione dell'indagine, poiché la possibilità di analisi diacroniche e di riuso dei dati dipende strettamente dalla qualità delle scelte effettuate lungo l'intero ciclo di vita. In questa prospettiva, la progettazione degli strumenti di rilevazione, la definizione delle *wave* e la strutturazione dei questionari *core* e dei moduli aggiuntivi sono orientate a produrre dati non solo analizzabili nel breve periodo, ma anche adeguatamente documentati e riutilizzabili nel tempo da una pluralità di soggetti e per finalità di ricerca differenti.

In termini operativi, il ciclo di vita dei dati ha inizio con la fase di pianificazione, che prevede la definizione delle strategie di gestione, inclusa la redazione di un *Data Management Plan* (DMP), documento essenziale per garantire la conformità etica e legale della ricerca. Nel caso di IOPP, questa fase e le successive sono declinate tenendo conto della struttura a *wave* del panel e della necessità di garantire coerenza e comparabilità nel tempo. La pianificazione e il DMP sono definiti in relazione a un flusso continuo di raccolta dati; la fase di elaborazione include controlli di qualità e procedure di armonizzazione che consentono di integrare informazioni provenienti da *wave* successive; l'archiviazione e la condivisione sono progettate per rendere esplicite le condizioni di accesso e di riuso, nel rispetto delle normative sulla protezione dei dati personali.

7.2 Gli Archivi Dati per le Scienze Sociali

A partire dalla prospettiva del ciclo di vita dei dati, risulta evidente il ruolo strategico svolto dagli archivi nel garantire infrastrutture e competenze a supporto della produzione, conservazione, condivisione e riutilizzo delle informazioni empiriche. Scheuch (2003) sottolinea come la nascita degli archivi dati per le scienze sociali non è stata il semplice esito di un'evoluzione tecnica, ma una scelta metodologica e politica finalizzata a consolidare la ricerca empirica compa-

rata e a preservare nel tempo la conoscenza prodotta. Gli archivi non si limitano a custodire dati: li rendono accessibili e riutilizzabili attraverso attività di documentazione, anonimizzazione, armonizzazione e preservazione a lungo termine, agendo come mediatori tra chi li produce e chi li riutilizza. Questa funzione di mediazione risulta particolarmente rilevante nel caso di infrastrutture di ricerca longitudinali come IOPP, in cui la produzione dei dati è distribuita nel tempo e coinvolge una pluralità di attori. In tali contesti, gli archivi contribuiscono a garantire la continuità informativa tra le diverse fasi del progetto e a rendere utilizzabili nel lungo periodo dati che nascono all'interno di un disegno complesso e dinamico.

Gli archivi, come mostrano Gaspani *et al.* (2019), non solo accompagnano i ricercatori nella predisposizione del *Data Management Plan* e nel deposito dei dati, ma sono anche fondamentali per garantire che siano condivisi secondo i principi FAIR. In Europa, il riferimento è CESSDA (*Consortium of European Social Science Data Archives*), una rete di archivi nazionali che opera per fornire un accesso aperto, sostenibile e interoperabile ai dati delle scienze sociali. In Italia tale ruolo è svolto da DASSI (*Data Archive for Social Sciences in Italy*), che contribuisce all'infrastruttura CESSDA fornendo servizi specialistici per la documentazione, la metadatazione e la gestione dei diritti di accesso.

Un elemento distintivo del lavoro di DASSI e degli altri archivi affiliati a CESSDA è l'adozione dello standard DDI (*Data Documentation Initiative*), che consente una descrizione strutturata e *machine-readable* dei dati. Questo standard assicura una documentazione scientifica strutturata e interoperabile, supportata da vocabolari controllati come quelli forniti da CESSDA e dalla DDI *Alliance*, che ne rafforzano la coerenza semantica (Marker e Fink 2018). Tali strumenti sono cruciali per facilitare il riutilizzo dei dati, sia da parte dei ricercatori e delle ricercatrici sia da parte di sistemi automatizzati. Per un panel come IOPP, caratterizzato da rilevazioni ripetute e da una struttura modulare dei questionari, l'adozione del DDI assume un ruolo strategico anche per la documentazione delle variazioni tra le *wave*, delle modifiche agli strumenti di rilevazione e delle procedure di armonizzazione adottate. Questo livello di dettaglio documentale è essenziale per consentire analisi longitudinali corrette e

confronti nel tempo. L'ancoraggio di IOPP a infrastrutture archivistiche consolidate consente inoltre di separare in modo chiaro le responsabilità legate alla produzione dei dati da quelle relative alla loro conservazione, diffusione e riuso, rafforzando la trasparenza del processo e la fiducia degli utenti finali.

Infine, gli archivi assumono un ruolo attivo anche nella promozione della cultura del *data sharing*, attraverso programmi di formazione e supporto all'uso dei dati (cfr., ad es., Dekker 2020). In questo senso, l'integrazione di IOPP nelle infrastrutture archivistiche non risponde solo a esigenze tecniche, ma contribuisce anche alla diffusione di pratiche di utilizzo e riutilizzo dei dati orientate alla trasparenza, alla responsabilità scientifica e alla collaborazione all'interno della comunità delle scienze sociali.

7.3 La Data Curation

Nelle scienze sociali, la *data curation* rappresenta l'insieme coordinato di attività che garantiscono la qualità e la sostenibilità dei dati nel tempo. Essa non si esaurisce con le operazioni tecniche di archiviazione o documentazione, ma include una visione complessiva del ciclo di vita del dato, dalla sua produzione al riutilizzo. In particolare, nelle indagini panel, dove la dimensione temporale introduce ulteriori complessità metodologiche e gestionali, la curatela dei dati assume una rilevanza strategica. In questo paragrafo verranno analizzate le principali attività di *data curation* alla luce degli standard internazionali e dei principi discussi nei precedenti paragrafi.

7.3.1 Pulizia dei dati

Il processo di *data cleaning* costituisce una fase cruciale e trasversale all'intero ciclo di vita dei dati, assumendo un ruolo centrale nelle indagini panel. In un contesto longitudinale, infatti, i dati sono soggetti a una molteplicità di rischi, tra cui errori di inserimento, incongruenze logiche e tassi variabili di non risposta, che possono compromettere l'affidabilità delle analisi se non correttamente gestiti. Il *data*

cleaning interviene quindi per assicurare la coerenza interna, la qualità complessiva e la tracciabilità del dato. In un'indagine panel come l'*Italian Online Probability Panel* (IOPP), tali obiettivi assumono un rilievo strutturale, poiché la qualità delle analisi longitudinali dipende dalla possibilità di applicare procedure di controllo coerenti e replicabili nel tempo.

Tra le procedure fondamentali rientra la verifica della coerenza logica delle risposte, con l'obiettivo di identificare valori implausibili o fuori intervallo, derivanti da errori di codifica, di digitazione o da malfunzionamenti nella logica dei filtri. Tali incongruenze possono compromettere l'affidabilità dell'intero blocco informativo e devono essere individuate e corrette attraverso regole formali di validazione. Nella progettazione di indagini complesse, e in particolare in contesti panel, gli errori nei filtri condizionali sono una delle principali fonti di distorsione sistematica. Come sottolinea Fowler (2013), anche piccoli malfunzionamenti nei percorsi logici di un questionario possono introdurre errori non casuali difficili da rilevare nelle fasi successive dell'analisi, compromettendo la qualità delle inferenze prodotte. Attenzione deve essere dedicata al controllo dei valori mancanti, in particolare nella costruzione di un sistema coerente di codifica e documentazione. Ogni valore *missing* deve essere descritto e classificato secondo uno schema esplicito e condiviso – ad esempio rifiuto, non applicabilità, assenza di risposta – evitando l'uso di mancanti di sistema privi di significato semantico. Questo approccio permette non solo una migliore leggibilità del dato, ma anche una maggiore interoperabilità tra le *wave*, evitando ambiguità che potrebbero ostacolare il riutilizzo.

All'interno di IOPP questa e altre procedure di *data cleaning* sono state adottate e costituiscono un elemento essenziale per garantire la coerenza longitudinale dei dati. Le regole di controllo e validazione sono progettate per essere applicabili in modo sistematico nel tempo, consentendo di identificare non solo errori puntuali, ma anche incoerenze emergenti tra rilevazioni successive.

7.3.2 Architettura dei dati

L'architettura dei dati progettata per un'indagine panel riflette l'esigenza di organizzare in modo ordinato, coerente e scalabile le informazioni raccolte nel tempo, consentendo una lettura sincronica e diacronica dei fenomeni studiati. In un'indagine longitudinale, la gestione efficace dei dati richiede una struttura modulare, capace di integrare progressivamente le diverse ondate di rilevazione senza pregiudicare la coerenza interna del sistema informativo.

In questo quadro, l'organizzazione dei dati si struttura generalmente attorno a un *dataset* anagrafico, contenente le informazioni identificative e strutturali degli individui e delle famiglie, affiancato da una serie di *dataset* tematici collegati, ciascuno relativo a un dominio specifico dell'indagine (Lucchini *et al.* 2023, 51-56). Il *dataset* anagrafico svolge la funzione di nodo centrale dell'architettura: include identificativi univoci e persistenti, progettati per permettere la tracciabilità longitudinale dei soggetti, mantenere la consistenza delle unità nel tempo e assicurare la corretta associazione tra individui, famiglie e informazioni raccolte nelle diverse ondate.

Questa organizzazione modulare, fondata su identificativi univoci, consente di mantenere la coerenza del sistema informativo e di supportarne l'evoluzione nel tempo. Questa impostazione architetturale consente inoltre di rendere esplicite le relazioni tra i diversi livelli informativi del panel, facilitando la documentazione, l'archiviazione e il riuso dei dati. In un progetto come IOPP, l'architettura dei dati non è quindi solo una scelta tecnica, ma una componente centrale della strategia di *data curation* e di sostenibilità dell'infrastruttura.

La presenza di strutture dati ben documentate, modulari e interoperabili rappresenta una condizione chiave per la qualità della ricerca e la replicabilità dei risultati (Singh *et al.* 2024). La progettazione dell'architettura dei dati assume così un significato strategico: non solo supporta le procedure analitiche, ma incide direttamente sulla trasparenza metodologica e sulla sostenibilità delle risorse informative prodotte. In uno studio panel, dove il tempo è una dimensione strutturante dell'informazione, l'architettura dei dati diventa un elemento centrale di validazione scientifica.

L'architettura di IOPP, nello specifico, è stata progettata seguendo proprio una logica modulare. Da una parte, il *dataset* anagrafico contiene le informazioni persistenti e gli identificativi univoci dei panelisti; dall'altra, i *dataset* tematici organizzano le informazioni raccolte durante le varie *wave* attraverso i *core questionnaire* e i moduli rotanti e aggiuntivi. Questa progettazione dinamica è in grado, dunque, di facilitare l'aggiunta nel tempo di ulteriori moduli rotanti o *one-shot* proposti dalla comunità scientifica che vorrà utilizzare IOPP come strumento di raccolta di dati di alta qualità.

7.3.3 Documentazione dei dati

La documentazione dei dati rappresenta una componente essenziale perché permette di mantenere traccia del contesto epistemologico, metodologico e tecnico in cui i dati sono stati prodotti, favorendone l'interpretazione e il riutilizzo anche a distanza di tempo. In particolare, nei contesti longitudinali, documentare in modo accurato i cambiamenti intervenuti nelle strategie di rilevazione, nella composizione campionaria o nella codifica delle variabili è decisivo per permettere la comparabilità tra le *wave* e la coerenza analitica dei dati nel tempo.

Una documentazione efficace include la descrizione del disegno di ricerca, delle modalità di campionamento e degli strumenti di rilevazione. Deve inoltre riportare i processi di pulizia e trasformazione del dato, le scelte adottate per l'anonimizzazione, la codifica e la gestione dei valori mancanti. Come indicano Van den Eynden *et al.* (2011), tale pratica si estende lungo l'intero ciclo di vita dei dati e può attingere sia a documenti già esistenti (progetti di ricerca, protocolli etici, note metodologiche, istruzioni o guide operative per gli intervistatori e le intervistatrici) sia a materiali creati *ad hoc* in fase di archiviazione. Nel caso di IOPP, questa documentazione accompagna in modo sistematico l'intero ciclo di produzione dei dati, integrando materiali metodologici, scelte operative e aggiornamenti successivi in una struttura coerente e orientata al riuso longitudinale. La documentazione incrementale rappresenta uno strumento chiave per preservare la comparabilità tra le *wave* e rendere trasparenti le deci-

sioni metodologiche adottate nel tempo. La registrazione sistematica delle modifiche ai questionari, alle codifiche e alle procedure di raccolta consente di supportare analisi longitudinali robuste e di facilitare il riuso dei dati da parte di ricercatori esterni al progetto.

Un altro aspetto importante riguarda l'utilizzo di strumenti semantici e automatizzati per la generazione dei metadati. CESSDA, ad esempio, promuove l'utilizzo di strumenti basati su vocabolari controllati per arricchire la descrizione dei dati con metadati interconnessi e *machine-actionable*. Questo tipo di documentazione prevede non solo la registrazione delle diverse versioni dei dati, ma anche il collegamento esplicito con risorse correlate come articoli scientifici, *script* di analisi o strumenti di raccolta. Tali pratiche migliorano la tracciabilità e la reperibilità dei dati, facilitandone il collegamento con altri contenuti scientifici e strumenti digitali (CESSDA Training Team 2022).

In questo scenario, la documentazione non è più un'operazione finale o secondaria, ma una componente strutturale del disegno di ricerca: contribuisce alla costruzione di una memoria condivisa del dato, sostenendo il riuso, interoperabilità e trasparenza che sono centrali per la ricerca empirica su base longitudinale.

7.3.4 Armonizzazione e standardizzazione

Nei contesti longitudinali e comparativi, la possibilità di combinare dati raccolti in tempi e luoghi diversi richiede processi rigorosi di armonizzazione e standardizzazione. Questo implica l'utilizzo di classificazioni condivise, come ISCO (*International Standard Classification of Occupations*) per le professioni, ISCED (*International Standard Classification of Education*) per i livelli di istruzione e NACE (*Nomenclature statistique des activités économiques dans la Communauté européenne*) per i settori economici, che consentono di allineare semanticamente variabili omologhe raccolte in contesti eterogenei. Tali strumenti favoriscono la comparabilità tra ondate successive all'interno di uno stesso studio e tra indagini condotte in paesi diversi, potenziando il valore cumulativo dei dati raccolti.

Nel panorama europeo, l'*European Social Survey* (ESS) rappresenta un esempio consolidato di armonizzazione, grazie a un approccio misto che combina standardizzazione in fase di *input* (progettazione del questionario) e in fase di *output* (documentazione e codifica). Il progetto adotta classificazioni internazionali accreditate, preservando allo stesso tempo le specificità nazionali attraverso procedure di collegamento e documentazione strutturata (Kolsrud e Skjak 2005).

Su scala globale, un'esperienza di particolare rilievo è quella del *Cross-National Equivalent File* (CNEF), che armonizza dati longitudinali provenienti da panel familiari di diversi paesi, tra cui Stati Uniti, Germania, Regno Unito e Australia. Il CNEF applica regole di armonizzazione trasparenti e dettagliate, documentando in modo sistematico l'origine delle variabili, le trasformazioni applicate e le differenze rispetto ai *dataset* nazionali, allo scopo di garantire la coerenza concettuale delle misure nel tempo e nello spazio (Frick *et al.* 2007).

Queste esperienze costituiscono un riferimento metodologico anche per IOPP, che adotta classificazioni e standard internazionali con l'obiettivo di garantire la comparabilità sia longitudinale sia *cross*-nazionale. L'adozione di questi standard è parte integrante del disegno del questionario, con l'obiettivo di rendere i dati di IOPP immediatamente integrabili con quelli di altri panel probabilistici europei e internazionali.

7.3.5 Archiviazione dei dati

La conservazione dei dati non può essere ridotta alla semplice archiviazione di file in un *repository* digitale, ma implica un insieme di pratiche finalizzate a garantirne la leggibilità, la comprensibilità e il riutilizzo nel tempo. In questa prospettiva, l'archiviazione rappresenta una componente strutturale del ciclo di vita dei dati e richiede l'adozione di infrastrutture adeguate, standard aperti e procedure coordinate che integrino aspetti tecnici, documentali e giuridici.

Un riferimento consolidato per la preservazione digitale è il modello OAIS (*Open Archival Information System*) che fornisce un

quadro concettuale per organizzare le attività di acquisizione, conservazione e disseminazione dei dati nel lungo periodo (*Consultative Committee for Space Data Systems*, 2024; Michetti 2007). Più che come schema operativo, OAIS è qui richiamato come modello di riferimento per evidenziare come la conservazione richieda un lavoro di cura che accompagna il dato fin dalla sua produzione, includendo la strutturazione dei metadati, la documentazione dei processi di trasformazione e la definizione delle condizioni d'uso.

In infrastrutture di ricerca complesse e di lungo periodo come l'*Italian Online Probability Panel* (IOPP), tali aspetti assumono una rilevanza particolare. La natura longitudinale del panel richiede infatti che i dati siano preservati in modo coerente e continuo nel tempo, assicurando l'integrità delle informazioni, la tracciabilità delle versioni e l'accessibilità a distanza di anni dalla loro raccolta.

La sostenibilità della conservazione passa inoltre attraverso scelte tecniche orientate alla durabilità, come l'utilizzo di formati aperti, il monitoraggio dell'integrità dei *file* e l'adozione di strategie di migrazione volte a contrastare l'obsolescenza tecnologica. In questo senso, la preservazione a lungo termine costituisce una dimensione essenziale della qualità dei dati e rappresenta una condizione necessaria per garantirne l'effettivo accesso e utilizzo nel tempo, in coerenza con i principi della scienza aperta (Dekker 2020; Pennock 2007).

7.3.6 *Riuso, disseminazione e divulgazione dei dati*

Il riuso dei dati rappresenta uno degli obiettivi centrali della scienza aperta e delle infrastrutture di *data curation*. Affinché i dati siano effettivamente riutilizzabili, tuttavia, non è sufficiente renderli disponibili: è necessario che siano accompagnati da una documentazione accurata, organizzati secondo strutture coerenti e resi accessibili attraverso licenze che ne regolino in modo esplicito le condizioni d'uso. In questa prospettiva, la dimensione *reusable* dei dati, così come definita dai principi FAIR, dipende dalla combinazione di scelte tecniche, documentali e giuridiche che incidono direttamente sul loro potenziale analitico (Kush *et al.* 2020).

Le piattaforme di pubblicazione dei dati svolgono un ruolo fondamentale nel supportare il riuso, offrendo ambienti trasparenti e tracciabili per la disseminazione delle risorse informative. Soluzioni *open source* come *Dataverse* forniscono meccanismi consolidati per l'attribuzione di identificativi persistenti, la gestione delle versioni e l'associazione dei *dataset* con materiali complementari, come *script* di analisi o documentazione metodologica, favorendo così la replicabilità delle ricerche (King 2007). L'adozione di standard di metadato condivisi, come il DDI nel contesto CESSDA, contribuisce inoltre a garantire l'interoperabilità tra archivi e la reperibilità dei dati all'interno di cataloghi di ricerca sovranazionali (Accordino *et al.* 2025).

L'accessibilità ai dati è regolata attraverso sistemi flessibili di licenza e di controllo degli accessi, che consentono di bilanciare l'esigenza di apertura con la tutela della riservatezza dei partecipanti. L'utilizzo di licenze standardizzate, come le *Creative Commons*, e di meccanismi di autenticazione o valutazione delle richieste di accesso rappresentano una strategia diffusa per gestire dati sensibili senza comprometterne la riusabilità scientifica (Chapoy *et al.* 2020). In questo quadro, anche la previsione di periodi di embargo può costituire uno strumento utile per conciliare la priorità scientifica dei ricercatori con la successiva apertura dei dati.

Nel caso di IOPP, le strategie di riuso e disseminazione sono progettate in coerenza con l'infrastruttura di archiviazione e con le pratiche di *data curation* adottate lungo l'intero ciclo di vita dei dati. I *dataset* prodotti dal panel saranno resi accessibili attraverso modalità differenziate, in funzione della tipologia delle informazioni e dei vincoli di tutela della privacy, e accompagnati da metadati e strumenti che ne facilitino l'esplorazione e l'utilizzo anche da parte di utenti non specialisti. L'obiettivo è dare la massima visibilità ai dati, facilitare il riutilizzo al fine di ampliare la portata dei risultati, rafforzare la trasparenza scientifica e promuovere una cultura del riuso responsabile e condiviso dei dati nelle scienze sociali.

7.4 La *Data Curation* e il paradigma *Data as Code*

Le attività di curatela e preservazione dei dati, fin qui discusse, stanno convergendo verso modelli sempre più dinamici e integrati, che valorizzano non solo l'accesso e la documentazione, ma anche la gestione attiva e iterativa delle informazioni nel tempo. In questo contesto si inserisce un approccio emergente e sempre più rilevante: quello del *Data as Code* (DaC). In un'infrastruttura di ricerca longitudinale come IOPP, il paradigma *Data as Code* fornisce una cornice metodologica particolarmente adatta per gestire l'evoluzione dei dati, delle procedure di pulizia e della documentazione nel tempo, garantendo tracciabilità e coerenza tra le diverse *wave*.

Il paradigma *Data as Code* propone di gestire i dati scientifici adottando strumenti e metodi propri dello sviluppo *software*: *versioning*, *testing*, automazione, controllo delle dipendenze. L'obiettivo non è solo tecnico, ma metodologico: rendere i dati trasparenti, tracciabili e collaborativi, al pari del codice. Come sottolineano Gentzkow e Shapiro (2014), i ricercatori nelle scienze sociali producono codice tanto quanto dati, e affrontano sfide analoghe a quelle dei programmatori: tracciabilità delle modifiche, riproduzione dei risultati, gestione dell'evoluzione nel tempo. Millman e Pérez (2014), in un contributo fondamentale sull'*open source* scientifico, evidenziano come l'adozione di ambienti versionati e pratiche condivise favorisca la replicabilità e la sostenibilità del lavoro scientifico. Wilson *et al.* (2017) aggiungono che pratiche semplici ma strutturate – come l'uso di Git, l'organizzazione dei *file* e la documentazione continua – possono migliorare sensibilmente la qualità della ricerca, anche in gruppi privi di competenze informatiche specialistiche.

Concretamente, trattare i dati come codice implica l'adozione di *workflow* tracciabili e formalizzati, in grado di sostituire approcci manuali e non documentati. L'uso di sistemi di *version control* consente di registrare in modo incrementale e reversibile le modifiche apportate ai dati e alle procedure di trattamento, rendendo esplicite le trasformazioni applicate nel tempo e facilitando la ricostruzione delle scelte metodologiche. All'interno di questi *workflow*, anche la gestione strutturata delle segnalazioni di problemi (*issue tracking*) diventa centrale: ogni intervento sui dati è accompagnato da una di-

scussione contestuale che ne motiva la necessità e ne registra l'impatto. Questa modalità facilita sia la redazione della documentazione da allegare allo studio, sia la ricostruzione puntuale delle scelte metodologiche in caso di revisioni successive.

In ambienti di ricerca longitudinali, l'approccio *Data as Code* permette di superare la frammentazione operativa e assicurare continuità tra fasi di lavoro diverse, gruppi di ricerca e *wave* successive. Il *dataset* non è più un prodotto statico, ma un'entità in evoluzione, accompagnata da una documentazione computazionale che ne assicura la tracciabilità e la riusabilità nel tempo. Nel caso di IOPP, questo approccio risulta particolarmente coerente con la struttura longitudinale del panel e con la necessità di gestire nel tempo versioni successive dei *dataset*, delle procedure di pulizia e dei materiali documentali. L'adozione di *workflow* tracciabili e di strumenti di *versioning* consente di rendere esplicite le trasformazioni applicate ai dati tra una *wave* e l'altra, facilitando sia la ricostruzione delle scelte metodologiche sia il riuso dei dati da parte della comunità scientifica. In questo senso, *Data as Code* si configura come un modello avanzato di *data management* pienamente coerente con i principi FAIR e con le esigenze di trasparenza, interoperabilità e apertura proprie dell'*Open Science*.

7.5 Conclusioni

Questo capitolo ha analizzato in modo sistematico le principali pratiche di *data management* e *data curation* applicabili agli studi longitudinali, mettendo in luce il ruolo centrale svolto dall'architettura dei dati, dalla documentazione, dall'armonizzazione e dalle strategie di archiviazione e disseminazione nel garantire la qualità, la sostenibilità e il riuso delle informazioni empiriche nel tempo. L'analisi ha mostrato come la gestione dei dati non possa essere concepita come una fase finale o accessoria del processo di ricerca, ma come una componente strutturale del disegno metodologico, che accompagna il dato lungo l'intero ciclo di vita.

In questo quadro, la *data curation* emerge come un insieme integrato di pratiche che richiedono competenze tecniche, attenzione do-

cumentale e responsabilità metodologica, in grado di supportare non solo la validazione scientifica dei risultati, ma anche la costruzione di infrastrutture condivise di conoscenza. L'adozione di standard internazionali, modelli di riferimento per la preservazione digitale e strategie orientate all'interoperabilità costituisce una condizione essenziale per rendere i dati effettivamente riutilizzabili e comparabili, in particolare nei contesti longitudinali.

L'esperienza dell'*Italian Online Probability Panel* (IOPP) rappresenta un caso concreto di applicazione coerente dei principi e delle pratiche discusse nel capitolo. L'integrazione del panel in un'infrastruttura di ricerca orientata alla preservazione, alla documentazione strutturata e alla disseminazione regolata dei dati consente di affrontare in modo sistematico le sfide poste dalla gestione di dati longitudinali complessi, favorendone la trasparenza, la tracciabilità e il riuso nel tempo.

Infine, l'introduzione di approcci come il paradigma *Data as Code* segnala una trasformazione in atto nelle modalità di gestione dei dati scientifici, che tende a valorizzare pratiche iterative, tracciabili e collaborative. In questa prospettiva, il *data management* non si limita a garantire la conformità a standard o requisiti formali, ma contribuisce a ridefinire il rapporto tra dati, strumenti e pratiche di ricerca, rafforzando le condizioni per una scienza sociale più aperta, cumulativa e sostenibile.

7.6 Riferimenti bibliografici

Consortium of European Social Science Data Archives (CESSDA) - <https://www.cessda.eu>.

Data Archive for Social Sciences in Italy (DASSI) - <https://www.dassi-archive.it>.

Data Documentation Alliance (DDI) - <https://ddialliance.org>.

Accordino F., Pecoraro F., Luzi D., 2025, *CESSDA data catalogue: An opportunity to enhance data in social sciences*, in *International Journal on Digital Libraries*, 26(1), 8. <https://doi.org/10.1007/s00799-025-00416-w>.

Ball A., 2012, *Review of Data Management Lifecycle Models*, University of Bath.

CESSDA Training Team, 2022, *CESSDA Data Management Expert Guide*. CESSDA ERIC. <https://dmeg.cessda.eu/>.

Chapoy E., Lesnard L., Gaultier-Voituriez O., Groshens E., Pedroja C., Beldiman-Moore A., 2020, *Sciences Po. Une histoire de données* (p. 91), [Research Report], Sciences Po. <https://sciencespo.hal.science/hal0361-2928>.

Consultative Committee for Space Data Systems, (2024, *Reference model for an Open Archival Information System (OAIS)* (Recommended Practice No. CCSDS 650.0-M-3, Magenta Book, NASA. <https://ccsds.org/Pubs/-650x0m3.pdf>).

Dekker R., 2020, *Social Data: CESSDA Best Practices*, in *Data Intelligence*, 2(1-2), 220-229. https://doi.org/10.1162/dint_a_00044.

Fowler F. J., 2013, *Survey Research Methods*, London, SAGE Publications, Inc.

Frick J. R., Jenkins S. P., Lillard D. R., Lipps O., Wooden M., 2007, *The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies*, in *Journal of Contextual Economics – SchmollersJahrbuch*, 127(4), 627-654. <https://doi.org/10.3790/schm.127.4.627>.

Gaspani F., Pisano C., Scisci D., 2019, *I dati oltre la ricerca: L'archiviazione nelle scienze sociali (Data beyond research: archival practices in the social sciences)*, in F. Gaspani, C. Pisano, D. Scisci, *AIS*, 14, 71-88. https://doi.org/10.1485/AIS_2019/14_3443545.

Gentzkow M., Shapiro J. M., 2014, *Code and Data for the Social Sciences: A Practitioner's Guide*, University of Chicago mimeo. <http://faculty.chicagobooth.edu/matthew.gentzkow/research/CodeAndData.pdf>

King G., 2007, *An Introduction to the Dataverse Network as an Infrastructure for Data Sharing*, in *Sociological Methods & Research*, 36(2), 173-199. <https://doi.org/10.1177/0049124107306660>.

Kolsrud K., Skjak K. K., 2005, *Harmonising background variables in the European Social Survey*, in J. H. P. Hoffmeyer-Zlotnik, J. Harkness, (a c. di), *Methodological aspects in crossnational research* (Vol. 11, pp. 163–182).

Kush R. D., Warzel D., Kush M. A., Sherman A., Navarro E. A., Fitzmartin R., Pétavy F., Galvez J., Becnel L. B., Zhou F. L., Harmon N., Jauregui B., Jackson T., Hudson L., 2020, *FAIR data sharing: The roles of common data elements and harmonization*, in *Journal of Biomedical Informatics*, 107, 103-421. <https://doi.org/10.1016/j.jbi.2020.103421>.

Lucchini M., Argentin G., Bussi D., Consolazio D., De Santis G., Gerosa T., Guidi G., Negrelli S., Piazzoni C., Pisati M., Respi C., Riva E., Sala E., Scisci D., Terraneo M., 2023, *Quality Profile. Questionnaires, Fieldwork, and Data Preparation*, in *IASSC TECHNICAL REPORTS*, 1, 1-68.

Marker H. J., Fink A. S., 2018, *CESSDA – a History of Research Data Management for Social Science Data*, in J. B. Thestrup, F. Kruse (a c. di), *Research Data Management—A European Perspective* (pp. 25–42). De Gruyter. <https://doi.org/10.1515/9783110365634-003>.

Michetti G., (a c. di), 2007, *OAIS - sistema informativo aperto per L'archiviazione*, ICCU.

Millman K. J., Pérez F., 2014, *Developing open-source scientific practice*, in *Implementing reproducible research* (149–183), Chapman and Hall/CRC.

Pennock M., 2007, *Digital Curation: A Life-Cycle Approach to Managing and Preserving Usable Digital Information*, in *Library & Archives*, 1(1), 1-3.

Scheuch E. K., 2003, *History and visions in the development of data services for the social sciences*, in *International Social Science Journal*, 55(177), 385-399. <https://doi.org/10.1111/j.1468-2451.2003.05503004.x>.

Singh R. K., Nagpal A., Jadhav A., Madalli D. P., 2024, *Assessment of FAIR (Findability, Accessibility, Interoperability, and Reusability) data implementation frameworks: A parametric approach* (No. arXiv:2504.06268). arXiv. <https://doi.org/10.48550/arXiv.2504.06268>.

Van den Eynden V., Corti L., Woollard M., Bishop L., Horton L., 2011, *Managing and Sharing Data*. UK Data Archive.

Wilkinson M. D., Dumontier M., Aalbersberg Ij. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.-W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., Mons B., 2016, *The FAIR Guiding Principles for scientific data management and stewardship*, in *Scientific Data*, 3, 160018.

Wilson G., Bryan J., Cranston K., Kitzes J., Nederbragt L., Teal T. K., 2017, *Good enough practices in scientific computing*, in *PLOS Computational Biology*, 13(6), e1005510. <https://doi.org/10.1371/journal.pcbi.1005510>.

8. Verso un approccio Integrato alla Sostenibilità per le Infrastrutture di Ricerca nelle Scienze Sociali: il caso dell'*Italian Online Probability Panel (IOPP)* nell'infrastruttura FOSSR

di *Claudia Pennacchiotti, Loredana Cerbara, Ferruccio Biolcati Rinaldi e Cristiano Vezzoni**

8.1 Introduzione

Se la progettazione e realizzazione di un panel probabilistico comporta una sfida di notevole impegno, altrettanto impegnativa, ma proprio per questo particolarmente importante, è la riflessione sulla sua sostenibilità nel tempo. Il capitolo prende in considerazione gli elementi che caratterizzano la sostenibilità di medio e lungo periodo del caso studio *Italian Online Probability Panel (IOPP)* (Santurro *et al.* 2025).

IOPP è un panel probabilistico (nel senso chiarito da Das 2025) inserito nel contesto più ampio di un'infrastruttura di ricerca per le scienze sociali (Hallonsten 2020), l'infrastruttura FOSSR (D'Ambrosio *et al.* 2025; Taddei e Paolucci 2026). Questa caratteristica, pur nella complessità della gestione in concomitanza con altre risorse di FOSSR, pone IOPP in una condizione privilegiata e unica, perché inserito in un contesto caratterizzato da ricchezza informativa e infrastrutturale. Per questo è necessario innanzi tutto delineare i diversi elementi strutturali, di sistema e gli aspetti specifici che possono favorire od ostacolarne la sopravvivenza e il pieno sviluppo.

È opportuno chiarire fin da subito che il tema della sostenibilità in questo caso sarà affrontato non tanto e non solo dal punto di vista della validità scientifica del panel (struttura, rappresentatività, livelli di *attrition*, strategie di *panel care*), aspetti già ampiamente trattati

* Il presente capitolo è il frutto di una riflessione comune da parte delle autrici e degli autori.

nei capitoli precedenti; si affronterà invece il tema della sostenibilità del panel in quanto parte integrante del sistema delle infrastrutture di ricerca, adottando un approccio quanto più possibile multidimensionale che possa sfruttare al meglio la condizione di partenza del panel, non solo come parte integrante di FOSSR, ma anche come realtà nazionale italiana capace di rientrare nel panorama internazionale. Accanto ai concetti di sostenibilità economica e *governance*, saranno infatti inclusi concetti come sostenibilità sociale, eccellenza scientifica, competenze e comunità professionale in modo da iniziare a delineare il possibile impatto sulla società civile e sulla comunità scientifica nazionale e internazionale.

Nonostante queste premesse che sembrano già individuare una prospettiva di lavoro, parlare di sostenibilità ha richiesto anzitutto l'identificazione di una definizione condivisa che guidasse il lavoro dei ricercatori e delle ricercatrici, consentendo di operativizzare il concetto in categorie e dimensioni, con attenzione a possibili associazioni e intersezioni.

Si è partiti dunque da un'analisi sistematica dei principali documenti e linee guida sul tema, con l'obiettivo di individuare ricorrenze e diversità tra le prospettive proposte, giungendo così ad una sintesi che fosse applicabile al caso studio IOPP. Nel lavoro si è scelto di privilegiare l'analisi critica dei *framework* concettuali proposti da istituzioni come la Commissione Europea, ESFRI (*European Strategy Forum on Research Infrastructures*) ed OECD (Organizzazione per la cooperazione e lo sviluppo economico), in quanto istituzioni coinvolte nella promozione dello sviluppo di un sistema integrato di infrastrutture all'interno e oltre lo Spazio Europeo per la Ricerca.

Categorie e dimensioni adottate sono successivamente state utilizzate per costruire un piano di sostenibilità di IOPP che comprendesse i diversi elementi in gioco e consentisse, da subito, di valorizzare le potenzialità offerte dalla struttura stessa del panel IOPP e dal contesto in cui esso si inserisce, e al contempo di tenere in considerazione i potenziali rischi di tipo endogeno ma anche strutturale e di sistema.

8.2 Per una definizione di sostenibilità delle Infrastrutture di Ricerca nelle scienze sociali

Prima di entrare nel merito della riflessione sul tema della sostenibilità, è opportuno richiamare la definizione operativa di Infrastruttura di Ricerca già promossa a livello europeo e a cui abbiamo fatto riferimento in IOPP:

“Le Infrastrutture di Ricerca sono strutture, risorse e servizi che sono usati dalle comunità di ricerca per condurre ricerca e promuovere l’innovazione nei rispettivi settori. Se del caso, esse possono essere utilizzate al di là della ricerca, ad esempio per scopi educativi o di servizio pubblico. Esse comprendono: attrezzature scientifiche di primaria importanza o serie di strumenti, risorse basate sulla conoscenza quali collezioni, archivi o dati scientifici, infrastrutture in rete quali sistemi di dati e calcolo e reti di comunicazione e qualsiasi altra infrastruttura di natura unica essenziale per raggiungere l’eccellenza nella ricerca e nell’innovazione.” (Parlamento Europeo 2013).

Le Infrastrutture di Ricerca (IR) possono dunque essere realizzate secondo diversi modelli organizzativi, in relazione ai loro obiettivi e ai rispettivi ambiti scientifici. Nel caso specifico (Santurro *et al.* 2025; Pennacchiotti *et al.* 2025), IOPP, all’interno dell’infrastruttura di ricerca per le scienze sociali FOSSR, si configura come un:

- *sistema di raccolta dei dati*: consente di raccogliere dati da uno stesso gruppo di partecipanti nel tempo, facilitando la comparabilità tra le diverse *wave* di rilevazione e costruendo nel tempo un monitoraggio longitudinale dei fenomeni oggetto di studio;
- *repository di dati*: è un archivio centralizzato che conserva e cura insieme di dati, garantendo l’accessibilità e la qualità dei dati;
- *spazio collaborativo*: è un ambiente virtuale che facilita lo scambio di conoscenze tra ricercatori e la co-costruzione di nuova conoscenza.

Data la complessità e le diverse sfaccettature dell’oggetto di studio, la sua sostenibilità non può che presentarsi come un concetto intrinsecamente multidimensionale, in cui si intrecciano e si influenzano reciprocamente aspetti di diversa natura.

In questa direzione, in effetti, si muovono i principali documenti e linee guida promossi da istituzioni e organismi internazionali coinvolti nella costruzione del sistema europeo delle infrastrutture di ricerca, a partire dai quali, attraverso un processo di analisi qualitativa del contenuto condotta con un approccio *grounded*, abbiamo ricostruito una definizione “integrata” da applicare al caso specifico di studio.

Per la nostra analisi abbiamo preso in considerazione sei tra linee guida e documenti di lavoro, indicati in tabella 8.1.

Tab. 8.1 – Linea guida utilizzate nell’analisi di contenuto

ESFRI	2017, Long-Term Sustainability of Research Infrastructures, ESFRI Scripta Volume II ISBN PDF: 978-88-901562-8-1
ESFRI	Hoffert, E. e alt., 2025, ESFRI Report: Funding of Research Infrastructures, DOI 10.5281/zenodo.14770890
European Commission	2025, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy on Research and Technology Infrastructures - COM(2025) 497 final/2
European Commission	2017, Sustainable European research infrastructures: a call for action: Commission staff working document: long-term sustainability of research infrastructures. Publications Office. https://data.europa.eu/doi/10.2777/76269 .
OECD	2017, Strengthening the effectiveness and sustainability of international Research Infrastructures, OECD Science, Technology and Industry Policy Papers, No. 48, OECD Publishing
EIRO Forum	2015, discussion paper: Long-term sustainability of Research Infrastructures

L’analisi del contenuto dei documenti è stata condotta, con il supporto del software NVivo, secondo un approccio qualitativo teso a identificare categorie tematiche e significati rilevanti e ricorrenti (de Lillo 2010; Zhang e Wildemuth 2009). I testi sono stati sottoposti a codifica attraverso ripetute letture. Le categorie di codifica sono state ottenute con un approccio *grounded* mediante induzione dai testi

esaminati (Charmaz 2014; Corbin e Strauss 2015). L'utilizzo della *Matrix Coding Query* di NVivo ha consentito poi di rilevare le relazioni tra i codici e il contenuto testuale, così da individuare, nell'eterogeneità delle prospettive proposte, schemi e associazioni non immediatamente evidenti. L'analisi ha evidenziato sette categorie ricorrenti e le relative dimensioni, come mostra la tabella 8.2.

Tab. 8.2 – Categorie della sostenibilità emerse dall'analisi di contenuto

CATEGORIA	DESCRIZIONE	DIMENSIONI
C1. Eccellenza Scientifica e Rilevanza	Questa categoria si articola intorno alla qualità della ricerca abilitata dall'IR e alla sua capacità di rimanere all'avanguardia	Qualità e Prestazione della Ricerca
		Valutazione e <i>Peer-Review</i>
		Rilevanza e Aggiornamento Tecnologico
		Accesso e Apertura (<i>Openness</i>)
		Monitoraggio dell'Utilizzo e degli <i>Output</i>
C2. Modello di Governance	Questa categoria si articola sulla struttura decisionale e operativa necessaria per la stabilità a lungo termine	Pianificazione del Ciclo di Vita (<i>Lifecycle Management</i>)
		Struttura Organizzativa e di <i>Governance</i>
		Gestione del Rischio (<i>Risk Management</i>)
C3. Modello di Finanziamento Sostenibile	Questa categoria si concentra sulla stabilità, l'adeguatezza e la trasparenza delle fonti di finanziamento lungo l'intero ciclo di vita	Stabilità e Impegno a Lungo Termine
		Trasparenza e Contabilizzazione dei Costi
		Diversificazione e Sinergia dei Fondi
C4. Risorse Umane e Competenze	Questa categoria si articola intorno alla disponibilità e al mantenimento di personale scientifico, tecnico e gestionale qualificato	Attrazione e Mantenimento del Talento
		Carriera e Mobilità
		Formazione e Sviluppo delle Competenze
C5. Impatto Socio-Economico e Potenziale di Innovazione	Questa categoria si articola sulla capacità dell'IR di generare benefici per la società, al di là dell'eccellenza scientifica	Misurazione e Valutazione dell'impatto sociale
		Innovazione e Trasferimento Tecnologico
		Comunicazione e Dialogo
C6. Gestione dei Dati e Infrastrutture Digitali (e-Infrastructure)	Questa categoria si concentra sulla gestione dei dati scientifici prodotti e sulla loro interconnessione con le infrastrutture digitali	Principi di Gestione dei Dati (FAIR e Open Science)
		Integrazione con <i>e-Infrastructures</i> (EOSC)
		Resilienza e Sovranità dei Dati
		Durata Temporale (e-IR)
C7. Dimensione Internazionale e Strategica / Resilienza	Questa categoria si articola sul posizionamento dell'IR nel contesto globale e sulla sua capacità di affrontare minacce esterne e strategiche	Cooperazione e <i>Diplomazia Scientifica</i>
		Sinergia IR/TI (<i>Technology Infrastructures</i>)

Elaborazione degli autori

Le sette categorie possono essere così specificate:

- C1. *Eccellenza Scientifica e Rilevanza*:** intesa come la capacità dell'infrastruttura di rimanere all'avanguardia grazie all'eccellenza nella ricerca condotta, allo sviluppo di nuove tecnologie e all'articolazione di robusti processi di valutazione.
- C2. *Modello di Governance*:** definisce la struttura decisionale e i processi operativi alla base dell'infrastruttura. Un modello di *governance* efficace necessita di una chiara assegnazione di ruoli e responsabilità, una definizione dei processi di lavoro e una chiara pianificazione del ciclo di vita.
- C3. *Modello di Finanziamento Sostenibile*:** strettamente collegato alla *governance*, è basato su fondi strutturati (ricorrenti) e adeguati, e supportato da un buon livello di diversificazione. Esso rappresenta uno dei pilastri fondamentali per la stabilità dell'infrastruttura e la possibilità di effettuare una programmazione di medio-lungo termine.
- C4. *Risorse Umane e Competenze*:** rappresentano il cuore dell'infrastruttura. La disponibilità e il mantenimento di personale scientifico, tecnico e gestionale qualificato, con percorsi di carriera chiari, opportunità di mobilità e formazione continua di alta qualità, forniscono all'infrastruttura solidità e consentono di aspirare all'eccellenza.
- C5. *Impatto Socio-Economico e Potenziale di Innovazione*:** si focalizza sulla capacità di generare benefici per la società e il sistema economico oltre che risultati scientifici. Soprattutto nell'ambito delle scienze sociali, le IR oltre ad essere un *hub* di innovazione e trasferimento tecnologico, rappresentano un punto di accesso/produzione di dati chiave per comprendere le sfide sociali complesse e per supportare la presa di decisioni politiche a partire da una reale conoscenza della realtà (OECD 2017).
- C6. *Gestione dei dati scientifici e loro interconnessione digitale*:** le scelte effettuate in quest'ambito determinano la capacità dell'infrastruttura (in quanto *data repository*, progettato specificamente per curare, conservare e rendere i dati disponibili per il riutilizzo) di integrarsi in sistemi come EOSC (*Euro-*

pean *Open Science Cloud*), affrontando la complessa questione della diversa durata temporale dei dati scientifici e degli elementi tecnici e tecnologici (*software, hardware*, sistemi informatici). Questo aspetto della sostenibilità trova un prezioso alleato nella adesione rigorosa ai principi FAIR (*Findable, Accessible, Interoperable, and Re-usable*) finalizzati a garantire che i dati siano per sé accessibili e riutilizzabili.

C7. Dimensione Internazionale e Strategica/Resilienza: questa dimensione definisce il posizionamento dell'IR nel contesto nazionale/internazionale della cooperazione scientifica, inserendola di fatto nei processi di diplomazia scientifica. Nel caso di un'infrastruttura per le scienze sociali che produce dati sulla popolazione, la diplomazia scientifica non riguarda solo la scienza "come oggetto" della politica estera, ma anche la scienza come linguaggio comune e strumento di cooperazione per la costruzione di relazioni basate sulla fiducia, sulla trasparenza dei dati e sulla condivisione di conoscenze per affrontare sfide sociali globali.

Queste sette categorie non operano in isolamento, ma formano un ecosistema complesso in cui la performance di una è strettamente correlata all'efficacia delle altre. Per questa ragione, con il supporto dei *software* NVivo e *Notebook LM*, è stata effettuata un'analisi delle possibili interazioni e associazioni fra le categorie. I risultati, centrali per lo sviluppo del piano di sostenibilità di IOPP, fanno emergere i principali legami esistenti tra categorie e le possibili dinamiche inter-categoriali. L'Eccellenza Scientifica (C1) costituisce un prerequisito imprescindibile, richiede standard elevati nella ricerca e un costante aggiornamento tecnologico. Tuttavia, tale eccellenza risulta intrinsecamente condizionata dalle decisioni strategiche relative alla Gestione dei Dati (C6), in particolare per quanto concerne il grado di apertura e accessibilità in conformità con i principi FAIR e le politiche di *Open Science*. Analogamente, le strategie implementate nell'ambito delle Risorse umane e delle competenze (C4) – quali i percorsi di carriera, le politiche di ritenzione e attrazione dei talenti e le opportunità di accrescimento delle competenze – esercitano un'influenza determinante sul conseguimento dell'eccellenza scientifica.

L'operatività dell'intera struttura necessita di un robusto Modello di *Governance* (C2), capace di presidiare l'intero ciclo di vita dell'infrastruttura, monitorarne le performance e gestire i rischi sistemici. Tale *governance* deve inoltre essere in grado di elaborare un Modello di Finanziamento Sostenibile (C3), fondato su risorse stabili e non dipendente esclusivamente da finanziamenti progettuali episodici. Questi due elementi costituiscono fattori determinanti per l'attrazione e il mantenimento del personale scientifico, tecnico e manageriale qualificato (C4), elementi essenziali per garantire il successo sia operativo che scientifico dell'infrastruttura.

L'Impatto Socio-Economico e il Potenziale di Innovazione (C5) rappresentano il valore generato dall'investimento in termini di avanzamento della conoscenza scientifica e di ricadute positive sulla società, la capacità dell'infrastruttura di supportare la comunità scientifica, facilitando scoperte e collaborazioni che producono benefici tangibili per il progresso sociale. Attraverso modelli di gestione strategici (C3) orientati alla sostenibilità e all'accessibilità dei servizi per i ricercatori e le ricercatrici, l'infrastruttura amplia il proprio impatto sociale, trasformando l'eccellenza scientifica in contributi concreti al benessere collettivo.

Infine, l'insieme di queste dimensioni si colloca all'interno della Dimensione Internazionale e Strategica (C7) dell'Infrastruttura, che presiede all'integrazione della stessa nel contesto globale della ricerca, promuovendo collaborazioni internazionali e rispondendo alle sfide scientifiche transnazionali. Questa dimensione assicura che tutte le operazioni – dalla gestione dei dati (C6) alla *governance* (C2) – siano caratterizzate da resilienza e sicurezza, garantendo continuità nell'erogazione dei servizi alla comunità scientifica e massimizzando l'impatto sociale della ricerca condotta, contribuendo così al progresso della conoscenza e al miglioramento delle condizioni di vita della società.

Le relazioni e le dipendenze descritte evidenziano la caratterizzazione delle infrastrutture di ricerca come sistemi complessi (Hallonsten, 2020) e la natura sistemica della sostenibilità, aspetti che richiedono una visione integrata e coordinata delle diverse dimensioni che le compongono.

8.3 Il caso dell'*Italian Online Probability Panel* nel contesto dell'infrastruttura di ricerca FOSSR

Le sette categorie individuate possono ora essere oggetto di riflessione nello specifico caso dell'infrastruttura di ricerca FOSSR declinandole opportunamente per il panel IOPP.

C1. Eccellenza scientifica e rilevanza

L'eccellenza scientifica e l'elevata qualità dei dati resi disponibili sono tratti caratteristici di IOPP. Sulla scia di esperienze internazionali come il panel olandese *Longitudinal Internet Studies for the Social sciences* (LISS), il tedesco GESIS Panel, il belga *Social Study* e lo *Swedish citizens Panel*, IOPP combina la solida tradizione delle indagini statistiche nazionali con l'innovazione dei panel probabilistici online, oggi considerati uno standard di riferimento per la produzione di dati affidabili, tempestivi e longitudinali nelle scienze sociali (Kocar e Kaczmirek 2023; Scherpenzeel 2011).

Come descritto nei capitoli precedenti, IOPP prevede cinque *wave* ogni anno, nel corso delle quali sono somministrati in più soluzioni un *Core Questionnaire*, concepito per una rilevazione longitudinale su base annuale di aspetti stabili e ricorrenti della vita sociale, e degli *Additional Questionnaire*, somministrati *una tantum* e proposti dalla comunità scientifica in risposta a *call* dedicate. Questa configurazione modulare, garantendo una continuità longitudinale, consente ai ricercatori e alle ricercatrici di analizzare i cambiamenti nel tempo di aspetti strutturali e salienti nella popolazione come la composizione demografica e familiare, l'istruzione, l'occupazione, il reddito, la salute, la partecipazione sociale, gli atteggiamenti politici e i valori. Tali variabili di base sono strutturate per garantire la comparabilità dei dati con infrastrutture internazionali quali l'*European Social Survey* (ESS), l'*European Values Study* (EVS), il *Generation & Gender Program* (GGP), la *Survey of Health, Ageing and Retirement in Europe* (SHARE-ERIC) e le più importanti indagini dell'Istituto Nazionale di Statistica (Istat; cfr. cap. 6). Al contempo la struttura modulare consente di integrare approfondimenti *una tantum* o ricorrenti su argomenti specifici e proposti dalla comunità scientifica, rafforzando

così la dimensione partecipativa della ricerca sociale e valorizzando la diversità dei punti di vista (cfr. cap. 3).

Questa struttura offre quindi una doppia opportunità di utilizzo di IOPP: usare i dati già raccolti in precedenza, sfruttando le informazioni già rilevate per integrarle in diverse prospettive di ricerca; proporre l'introduzione di nuovi quesiti utili alle ricerche di diversi gruppi di ricerca, arricchiti dalla profondità dei dati longitudinali a cui essi sono sempre relazionabili.

Va detto che a livello nazionale, nel contesto della produzione di statistiche ufficiali e accademiche, già esistono infrastrutture per la realizzazione di indagini probabilistiche longitudinali: ad esempio, l'Indagine sulle Forze di Lavoro, le Indagini Multiscopo e l'Indagine su Famiglia e Soggetti Sociali dell'Istat, anch'esse metodologicamente armonizzate con gli standard internazionali, che garantiscono al contempo validità inferenziale e comparabilità temporale.

L'innovazione introdotta da IOPP è rendere possibile ai ricercatori la realizzazione di indagini su un panel probabilistico (Lynn 2012; Menard 2002) che garantiscono un elevato rigore scientifico e permettono, al tempo stesso, rapidità d'implementazione, con risultati disponibili in un periodo di tempo molto ridotto rispetto alle tradizionali rilevazioni. In questo modo si cerca di colmare il *gap* attualmente esistente tra la realtà italiana e quella europea, dove alcuni Stati (Paesi Bassi, Svezia, Germania, Belgio) hanno già implementato questo strumento. Tale agilità si rivela particolarmente preziosa in periodi di crisi o di rapidi mutamenti sociali, in cui il *feedback* in tempo reale è essenziale, ad esempio, per fornire indicazioni ai decisori politici, facilitando risposte politiche più informate e tempestive (Blom *et al.* 2020).

IOPP è stato progettato non solo per generare nuovi dati, ma anche per renderli immediatamente utilizzabili, replicabili e interoperabili, seguendo i principi FAIR e la logica della *Open Science*, che promuove trasparenza, riusabilità e collaborazione all'interno e oltre i confini nazionali. Per questa ragione IOPP adotta i più elevati standard di archiviazione e conservazione dei dati raccolti, garantendone così persistenza e disponibilità, a prescindere dall'obsolescenza tecnologica dell'infrastruttura che li conserva. I *dataset* sono codificati secondo le classificazioni internazionali per permettere la comparabi-

lità con le altre principali indagini europee, mentre i metadati offrono una descrizione dettagliata della produzione e dell'interpretazione dei dati così da renderli effettivamente fruibili a chiunque voglia utilizzarli. Allo stesso modo tutta la documentazione relativa al panel e agli studi condotti (note sulla metodologia di campionamento, processo di costruzione, calibrazione e mantenimento del panel, studi longitudinali, *codebook*, questionari core utilizzati etc.) è liberamente accessibile ed esplorabile in linea con i principi dell'*Open Science* secondo cui “*It is not just the findings of science that must be transparent, subject to scrutiny and verification. So must the practices, social structures and methods of science be transparent, subject to scrutiny and critique.*” (CODATA 2020). È però importante precisare che l'adesione ai principi dell'*Open Science* non significa rilanciare l'idea di un'apertura indiscriminata dei processi scientifici e dei dati prodotti. Al contrario, IOPP è sostenuto da una governance responsabile, basata su principi e valori fondamentali che tengono in costante considerazione implicazioni etiche, epistemologiche, economiche, legali, politiche, sociali e tecnologiche. Ad esempio, tutti i questionari somministrati al panel vengono sottoposti ad una rigorosa valutazione preliminare da parte della Commissione per l'integrità e l'etica della ricerca del CNR. Inoltre, l'accesso aperto ai dati è consentito solo per fini scientifici e l'utilizzatore o l'utilizzatrice sottoscrivono un accordo di buon utilizzo, in linea con quanto previsto dai più importanti produttori di dati, nazionali e internazionali.

C2. Modello di governance

IOPP è il risultato di una collaborazione tra il CNR e l'Università degli Studi di Milano La Statale, disciplinata, nel corso della fase di avvio finanziata dal PNRR, da una Convenzione *ex art.* 15 L. 241/1990 in cui l'Università Statale è impegnata a fornire supporto scientifico alla costruzione e calibrazione del panel, al suo sviluppo strategico in collegamento con gli altri panel probabilistici europei, e alla disseminazione dei risultati. In prospettiva questa collaborazione verrà resa strutturale, attraverso una formalizzazione legale, ad esempio una *Joint Research Unit*, che ne disciplini ruoli e responsabilità.

IOPP si inserisce, come già noto, nella più ampia infrastruttura FOSSR, che ne indirizza il modello di *governance* (seppure con ampi spazi di specificità); un modello orientato al mantenimento operativo di lungo termine, con particolare attenzione al passaggio dalla fase di costruzione e calibrazione finanziata dal PNRR, alla fase “a regime”.

La *governance* di IOPP è assicurata da uno *Scientific Board*, costituito, nella fase di avvio e per l'intera durata del finanziamento PNRR, da rappresentanti del CNR e dell'Università degli Studi di Milano La Statale. Nella successiva fase di consolidamento e integrazione a sistema, il Board sarà ampliato per includere esponenti di rilievo provenienti dalle università già coinvolte nell'iniziativa FOSSR.

Lo *Scientific Board* ha la responsabilità di indirizzare e supervisionare le dimensioni scientifica e strategica dell'infrastruttura, garantendone l'eccellenza, la rilevanza rispetto alle esigenze della comunità degli utilizzatori e delle utilizzatrici e la sostenibilità a lungo termine.

La dimensione operativa è invece affidata a una figura di coordinamento operativo, incaricata di attuare le linee strategiche definite dallo *Scientific Board* e di coordinare il lavoro dei gruppi di ricerca e del personale assegnato a IOPP.

Questa struttura di *governance*, volutamente snella, è concepita per assicurare rapidità ed efficienza nei processi decisionali, in coerenza con le caratteristiche di innovazione precedentemente descritte, senza tuttavia compromettere la rigurosità metodologica e l'eccellenza scientifica del panel e dei dati da esso generati.

Particolare cura viene dedicata al disegno del ciclo di vita di IOPP che ne deve garantire sostenibilità, efficienza e impatto scientifico. Una pianificazione strutturata consente, infatti, di prevedere risorse, modelli di finanziamento e strategie di evoluzione, riducendo rischi di frammentazione od obsolescenza. Inoltre, il ciclo di vita assicura coerenza tra obiettivi scientifici, esigenze degli e delle utenti e capacità operative, favorendo una gestione trasparente e adattabile nel tempo.

Una volta terminato il finanziamento PNRR, nel passaggio tra fase di creazione e mantenimento, dal punto di vista del quadro giuridico, IOPP si inserirà nella *legal entity* adottata dall'Infrastruttura

FOSSR, mantenendo un certo livello di autonomia in ragione delle proprie specificità.

Nel quadro del ciclo di vita del panel, alla fase di reclutamento dei panelisti e delle paneliste seguirà una fase di calibrazione e di primo *refreshment*. Da quel momento in poi sarà centrale, per la continuità del panel, la strategia di *panel care* attivata, finalizzata a contenere i livelli di *attrition* (perdita progressiva dei partecipanti nel tempo). Da questo punto di vista, il gruppo di ricercatori e ricercatrici del *team* CNR - Università di Milano, a seguito di una riflessione a partire dalla letteratura di riferimento e di una fase di analisi delle pratiche adottate da alcuni tra i maggiori *benchmark* europei (LISS, GESIS panel, *Social Study*, *Swedish citizens panel*) ha delineato una strategia per la *retention* dei panelisti di tipo misto, puntando su un sistema di incentivazione annuale, calibrato progressivamente in base al numero di *wave* a cui il panelista ha contribuito, accompagnato da interventi volti a costruire una relazione duratura e personalizzata tra il panelista e il progetto, fatta di solleciti mirati, comunicazione puntuale e bidirezionale, e rafforzamento del senso di appartenenza e cura della dimensione di autorevolezza non solo in ambito scientifico (cfr. cap. 5). Il rapporto fiduciario tra il o la panelista e l'ente di ricerca è una delle dimensioni capaci di incidere non solo sulla fidelizzazione dei panelisti e delle paneliste nel tempo, ma anche sulla qualità e la completezza delle risposte ottenute.

Per questo, parte significativa delle strategie di *panel care* attivate in IOPP fin dalla fase di reclutamento include l'attenzione costante agli aspetti etici, alla trasparenza e chiarezza della comunicazione rispetto ad obiettivi della ricerca, impatto sociale e scientifico atteso, alle misure attivate per la protezione e la riservatezza dei dati personali.

C3. Modello di finanziamento sostenibile

La sostenibilità economica rappresenta una delle principali sfide affrontate fin dalle fasi iniziali dal *team* di ricerca di IOPP. Sebbene l'ingente investimento iniziale sia stato reso possibile grazie ai fondi del PNRR, a partire dalla metà del 2026 l'infrastruttura dovrà essere in grado di garantire la propria autonomia finanziaria. Questa,

nell'attuale contesto italiano, rappresenta una delle maggiori sfide da affrontare.

A tal fine, il *Business Plan* di IOPP delinea il quadro strategico per i primi cinque anni, con l'obiettivo di ridurre progressivamente la dipendenza dai finanziamenti temporanei del PNRR e di sviluppare un modello di finanziamento diversificato (*Cross-Funding*). Tale modello prevede, tra le fonti di sostentamento, la generazione di entrate grazie ai questionari aggiuntivi, resi disponibili nell'ambito di due delle cinque *wave* annuali con un sistema di accesso a pagamento (*Pay-Per-Use Fees*; cfr. cap. 3). Questa impostazione consente, secondo le previsioni, di coprire i costi operativi minimi – inclusi una quota del personale, gli incentivi ai panelisti e alle paneliste, la manutenzione dell'infrastruttura informatica, il *refreshment* del panel e la comunicazione e disseminazione – garantendo così una parziale continuità dell'infrastruttura. In questo modello un fattore dirimente è la costruzione di una solida reputazione scientifica basata non solo sulla qualità dei dati resi disponibili ma anche sulla serietà e trasparenza dei processi di valutazione e selezione delle richieste provenienti da ricercatori, ricercatrici ed enti di ricerca, che dovranno in ogni caso garantire elevati standard etici e scientifici.

Per contro, un elemento critico permane nella vulnerabilità alle fluttuazioni del mercato, che potrebbero incidere sulla stabilità economica di lungo periodo e sulla indipendenza scientifica del panel stesso. Le esperienze già realizzate in altri paesi dimostrano, infatti, quanto sia importante beneficiare di una quota di finanziamenti strutturali correnti che non solo forniscono una ragionevole garanzia di continuità e capacità di programmazione all'infrastruttura, ma consentono anche di attrarre personale scientifico, tecnico e gestionale qualificato, investire sulla sua formazione e l'acquisizione di nuove competenze. Nel contesto italiano attuale la “conquista” di un finanziamento strutturale è un fatto complesso. Lo strumento principale previsto per questo tipo di interventi è il FOE (Fondo Ordinario destinato agli Enti di ricerca) che tuttavia richiede un livello di maturità più avanzato da parte dell'Infrastruttura e soprattutto eroga importi che non sono in grado di coprire per intero il fabbisogno economico di base stimato per IOPP.

Per tutte queste ragioni, il modello di finanziamento implementato prevede una diversificazione che affianca alla quota parte di investimento pubblico a regime (ancora tutta da verificare), una quota proveniente dal mercato e una parte di *fund raising* rivolto a programmi di finanziamento Europei e nazionali per attività specifiche e mirate.

C4. Risorse umane e competenze

Come già sottolineato più volte, le risorse umane (ricercatori e ricercatrici, tecnologi e tecnologhe, tecnici e tecniche, amministrativi e amministrative) rappresentano il cuore dell'infrastruttura e sono parte integrante dei processi volti al raggiungimento e mantenimento dell'eccellenza scientifica.

Nella fase di costruzione, in cui si trova IOPP attualmente, attrarre e mantenere personale scientifico, tecnico e gestionale qualificato, offrire percorsi di carriera chiari, opportunità di mobilità e formazione continua di alta qualità, rappresenta una sfida. Il rischio di operare con un gruppo di lavoro sottodimensionato, non adeguatamente formato o privo della stabilità necessaria ad una progettazione di medio periodo è elevato. Per questo diventa necessaria una programmazione attenta fin dal primo momento.

A partire dalle esperienze già realizzate nell'ambito di panel come il GESIS e il LISS, è stata ipotizzata la composizione ideale del gruppo di lavoro IOPP che dovrebbe includere almeno: un coordinatore scientifico, un manager, un ricercatore esperto di campionamento, due ricercatori esperti in metodologia della ricerca sociale o statistica sociale, due ricercatori esperti di *data curation*, un ricercatore/tecnologo in *Computational Social Science*, due tecnici informatici, un esperto di GDPR e etica, un esperto di comunicazione, un amministrativo. Tuttavia, l'attuale carenza di finanziamenti strutturali rende difficile l'acquisizione e la fidelizzazione del gruppo di lavoro nella sua interezza.

Al momento, grazie ai fondi PNRR, sono infatti impegnati a tempo pieno in IOPP: come personale strutturato, un coordinatore scientifico e due ricercatori esperti in metodologia delle scienze sociali; come personale a tempo determinato, quattro tecnologi (di cui uno con competenze gestionali, uno con competenze statistiche, due con

competenze metodologiche), due ricercatori (uno con competenze statistiche e uno con competenze metodologiche). Le competenze informatiche sono invece esternalizzate al fornitore che sta costruendo il sistema integrato di *open software* IOPP.

Con la fine del sostegno PNRR sarà difficile garantire la continuità all'intero gruppo di lavoro e sarà inevitabilmente necessario ridimensionare le attività sostenibili.

Per far fronte a questo rischio, comune a tutta l'infrastruttura FOSSR, il gruppo di coordinamento sta attuando un piano di *fund raising* diversificato, cercando di intercettare fondi nazionali ed europei. È chiaro che si tratta di soluzioni ponte, che non garantiscono in alcun modo la piena funzionalità dell'infrastruttura né possono rappresentare una prospettiva sostenibile per il piano di sviluppo di IOPP.

Per quanto riguarda i piani formativi finalizzati al rafforzamento delle competenze del personale coinvolto in IOPP, oltre all'offerta formativa resa disponibile dall'infrastruttura FOSSR (*research data management and methodological courses*) il gruppo di lavoro IOPP potrà in parte accedere a programmi di *short mobility* messi a disposizione dal CNR per il suo personale. Grazie a quest'opportunità si sta pensando di programmare dei percorsi di *job shadowing* presso la rete dei panel europei, per studiare più da vicino le diverse esperienze e dare seguito ai processi di *networking* già attivati. L'opportunità sarà ovviamente limitata a quanti siano in servizio al CNR con una prospettiva almeno semestrale, lasciando quindi esclusa buona parte del gruppo di lavoro attuale.

C5. *Impatto socioeconomico*

L'impatto di IOPP può essere osservato a due livelli: nell'ambito della comunità scientifica e sulla società.

Impatto sulla comunità scientifica. Tra gli elementi caratteristici di IOPP è importante sottolineare la sua struttura aperta non solo "in uscita" (i dati vengono rilasciati tramite l'infrastruttura FOSSR in formato *open* e conforme ai principi FAIR) ma anche "in entrata" (tramite il sistema di *call* periodiche per l'acquisto degli *Additional-Questionnaires*; cfr. cap. 3). Questa interpretazione a due vie del

concetto di apertura, pone IOPP in linea con i principi della Raccomandazione UNESCO sull'*Open Science* del 2021 che dà centralità ad aspetti come il rispetto della libertà accademica, il sostegno ad una ricerca di alta qualità che accolga i diversi punti di vista e renda metodi e risultati della ricerca ampiamente disponibili per una revisione e un esame rigorosi e per processi di valutazione trasparenti.

In questo modo, IOPP contribuisce allo sviluppo dei processi di democratizzazione in campo scientifico, grazie a una equa distribuzione di risorse e opportunità di conoscenza (Fricker 2007). Per rendere i propri strumenti di conoscenza scientifica (*input* e *output*), accessibili a tutti e a tutte e co-costruibili congiuntamente, IOPP prevede la possibilità per ricercatori e ricercatrici ad inizio carriera o provenienti da contesti particolarmente complessi, di accedere alle opportunità offerte dagli *Additional Questionnaires* a titolo non oneroso, dedicando *call* specifiche a questo particolare *target* di scienziati. Quest'approccio favorisce non solo una democratizzazione della conoscenza scientifica, ma anche la possibilità per individui e comunità marginalizzate di affrontare problematiche locali – o globali con impatti locali – attraverso dati e strumenti condivisi (Bezuidenhout *et al.* 2020).

Infine, è importante evidenziare anche l'impatto economico che uno strumento come IOPP può generare per la comunità scientifica. Esso, infatti, consente l'accesso ad un panel probabilistico, uno strumento robusto e supportato da dati longitudinali di alta qualità relativi alle principali caratteristiche socio-demografiche degli individui, ad un costo pari, se non addirittura inferiore, agli attuali costi di mercato per indagini analoghe.

Impatto sulla società - Contributo ai Processi Decisionali. Il principale impatto sociale generato da IOPP è legato alla promozione di processi decisionali informati ed *evidence-based* grazie alla conoscenza che esso consente di acquisire rispetto alla società, alle sue caratteristiche, ai cambiamenti che si manifestano al suo interno, anche in ottica predittiva.

Inserito nel più ampio contesto dell'infrastruttura FOSSR che si propone di rafforzare il legame tra la ricerca scientifica e l'elaborazione delle politiche attraverso lo sviluppo di strumenti in grado di tradurre i risultati della ricerca in raccomandazioni politiche

concretamente attuabili (*evidence-based policy making*), IOPP rende disponibili i dati raccolti, anche in forme e attraverso canali accessibili ai non addetti. Al termine di ogni ciclo di rilevazione, e comunque secondo necessità, è prevista la predisposizione, sempre in formato *open*, di *report* (anche in forma sintetica e grafica per una capacità divulgativa ad ampio spettro) e *policy brief* appositamente strutturati per supportare i decisori politici. Su richiesta possono essere ulteriormente realizzati *policy brief* o *report customizzati* per territori e/o politiche rientranti tra gli ambiti monitorati e indagati tramite il panel.

C6. Gestione dei dati

Aspetti specifici sui sistemi di archiviazione e conservazione dei dati adottati sono stati già stati ampiamente trattati nel precedente capitolo 7 a cui si rimanda. In questa sede è opportuno ribadire solo alcuni degli aspetti di carattere generale, direttamente riconducibili alla persistenza e consultabilità dei dati raccolti.

Grazie all'applicazione dei principi FAIR, la struttura di archiviazione e disseminazione dei dati raccolti da IOPP consente che essi rimangano solidi, utilizzabili e accessibili a un pubblico più ampio di ricercatori, ricercatrici e decisori politici.

La dotazione informatica di IOPP, per la raccolta e archiviazione dei dati, risiede all'interno dell'infrastruttura FOSSR che è dotata di una rete di quattro *data center* per la conservazione dei dati, distribuiti a Napoli, Palermo, Torino e Catania.

In linea con i principi dell'*Open Science*, IOPP utilizza *software open source* per la raccolta e l'archiviazione dei dati. Questa scelta, volta a supportare la rilevanza e l'aggiornamento tecnologico dell'infrastruttura garantisce anche:

- trasparenza e accessibilità del codice;
- indipendenza dai produttori;
- interoperabilità, standard aperti e riduzione del rischio di “*lock-in tecnologico*”.

Sul fronte dei dati raccolti, particolare attenzione viene prestata alla *data curation* anche in considerazione delle specificità legate alla caratteristica di longitudinalità, dove la dimensione temporale introduce complessità metodologiche e gestionali. A seguito della fase di pulizia, i dati vengono organizzati all'interno di una architettura relazionale ordinata, coerente e scalabile che ne consente una lettura sincronica e diacronica. Inoltre, l'utilizzo, per specifiche variabili, di classificazioni condivise rende possibile la comparabilità sia tra *wave* successive, sia con indagini condotte in contesti differenti potenziando il valore cumulativo dei dati raccolti da IOPP.

A supporto delle singole *wave* realizzate, è fornita tutta la documentazione necessaria a mantenere traccia del contesto metodologico e tecnico di produzione, dei cambiamenti intervenuti rispetto alle caratteristiche del panel, alla codifica delle variabili o nelle strategie di rilevazione così da consentire la comparabilità tra le *wave* e la coerenza analitica dei dati nel tempo.

C7. Dimensione internazionale e strategica

Con riferimento a quest'ultima categoria un primo aspetto da considerare è la dimensione internazionale che IOPP, pur caratterizzandosi come un panel rappresentativo della popolazione italiana, aspira ad acquisire.

Anzitutto essendo inserito all'interno dell'infrastruttura FOSSR, beneficerà del percorso di adesione all'*European Open Science Cloud* (EOSC) che a breve sarà attivato. Inoltre, è importante evidenziare che fin dalla fase di progettazione e costruzione, la dimensione collaborativa (fra soggetti, competenze e esperienze) ha rappresentato un elemento strutturale, addirittura necessario alla buona riuscita dell'impresa. Proprio perché IOPP rappresenta un *unicum* nel contesto italiano, è stato quasi naturale confrontarsi con esperienze analoghe ma già sedimentate nel contesto europeo. Panel come il LISS, il GESIS, lo *Swedish citizens Panel* o il *Social Study* sono fin da subito apparse come esperienze da cui sarebbe stato possibile apprendere molto in termini di strategie e scelte epistemologiche effettuate nonché di pratiche attuate. Le relazioni sviluppate in queste prime fasi potranno consolidarsi nel tempo, favorendo la nascita di

network transazionali che favoriscano la collaborazione, la condivisione dei dati, e diano un contributo specifico alla cooperazione internazionale nella produzione dei dati.

A ciò si aggiunge l'integrazione di IOPP con altre tre infrastrutture transazionali (SHARE, GGP, GUIDE) avviata nel corso del finanziamento PNRR, strategica per la costruzione di un potenziale *Life Course Observatory*, che potrebbe rappresentare un insieme di servizi complementari di ricerca in ambito sociale.

Un secondo aspetto, riconducibile alla dimensione strategica che IOPP potrà assumere è il suo ruolo nella promozione del dialogo tra scienza e politica. Un panel di questo tipo fornisce dati affidabili e condivisi che possono offrire elementi di dibattito ai decisori politici su questioni globali e locali (es. cambiamenti demografici, migrazioni, disuguaglianze sociali, sostenibilità), favorendo una governance basata su evidenze empiriche.

8.4 Conclusioni

Il concetto di sostenibilità presenta una natura multiforme, articolandosi in diverse categorie e dimensioni che formano un ecosistema integrato. In tale sistema, la *performance* di ogni singola componente è interdipendente rispetto all'efficacia delle altre, e le vulnerabilità che emergono in un ambito specifico tendono a generare effetti a cascata sulle restanti dimensioni, minando la stabilità dell'intera infrastruttura.

Lo studio condotto evidenzia quanto sia importante, per promuovere la sostenibilità nel contesto di una infrastruttura di ricerca, agire in maniera sistemica *vis-à-vis* dei fattori endogeni ed esogeni e di sistema che possono influenzarla. In questo capitolo, riflettendo sulla sostenibilità di medio-lungo periodo del Panel IOPP all'interno dell'Infrastruttura di ricerca FOSSR, abbiamo cercato di puntare un faro sugli aspetti che si vanno via via consolidando e sulle criticità emergenti. Se da una parte l'eccellenza scientifica, la robustezza metodologica, una solida *governance*, il costante riferimento ai principi dell'*Open Science* e della *fairness*, le relazioni tessute a livello internazionale rappresentano dei sicuri punti di forza, la sostenibilità finanziaria si pone come un evidente elemento di criticità. Come ab-

biamo visto, l'assenza di una componente finanziaria strutturale, che consenta una programmazione quantomeno di medio periodo da una parte ostacola l'attrattività e il mantenimento delle risorse umane che rappresentano uno dei pilastri fondamentali a garanzia dell'eccellenza scientifica e conseguentemente anche della rilevanza dell'impatto socioeconomico generato, e dall'altra espone IOPP alle fluttuazioni e alle esigenze del mercato, minandone sia la stabilità che l'indipendenza scientifica.

Il contesto specifico in cui IOPP è stato progettato, caratterizzato da una disponibilità di fonti finanziarie senza precedenti grazie al PNRR, ha consentito un avvio favorevole, se confrontato con le esperienze già realizzate a livello europeo. Tuttavia, questa condizione eccezionalmente favorevole di avvio rischia di trasformarsi in un elemento di debolezza, se non si troveranno altre fonti strutturali che forniscano un sostegno di base all'attività.

La storia dei panel europei con cui IOPP si confronta e che hanno rappresentato il *benchmark* di riferimento in questa fase di avvio, ci ha mostrato quanto il passaggio alla fase di messa a regime sia legato alla integrazione effettiva dell'infrastruttura panel nel sistema-ricerca del paese in cui si inserisce e al corrispondente supporto da parte di fondi pubblici strutturati.

Un altro aspetto che copre un ruolo centrale per garantire la sostenibilità di IOPP è la sua capacità di inserirsi in *network* transnazionali. Tale processo si sviluppa attraverso due direttrici principali: l'adesione alle reti dei panel probabilistici europei, che consente di potenziare la collaborazione scientifica e la condivisione dei dati, configurandosi come contributo specifico alla cooperazione internazionale nella generazione di basi dati; e l'intensificazione delle relazioni con infrastrutture di ricerca quali SHARE, GUIDE e GGS, con cui sono già state stabilite forme di interazione nella fase di avvio, in prospettiva della realizzazione di un *Life Course Observatory* che configuri un ecosistema integrato di servizi di ricerca complementari in ambito sociale.

8.5 Riferimenti bibliografici

- Bezuidenhout L.M., Leonelli S., Kelly A. H., Rappert B., 2017, *Beyond the digital divide: Towards a situated approach to open data*, in *Science and Public Policy*, 44(4), 464-475, <https://doi.org/10.1093/scipol/scw036>.
- Charmaz K., 2014, *Constructing grounded theory. 2nd Edition*, London, Sage.
- CODATA / International Science Council, 2020, *Open Science for a Global Transformation*.
- Corbin J., Strauss A., 2015, *Basics of qualitative research. Techniques and procedures for developing grounded theory. Fourth Edition*, London, Sage.
- D'Ambrosio G., Marchesini N., Pennacchiotti C., Primerano I., 2025, *Creazione e sviluppo di un'infrastruttura di ricerca nelle scienze sociali in Italia: il progetto FOSSR*, in IRPPS Working papers 149, 38.
- Das M., 2025, *Building on the LISS Experience: Future Challenges and Opportunities for Online Probability-based Panels*, in Conference "European Online Probability Panels (EOPPs): opportunities for the social research", Milan 5-6 May 2025.
- de Lillo A., 2010, *Il mondo della ricerca qualitativa*, Torino, Utet.
- EIROForum, 2015, *Discussion paper: Long-term sustainability of Research Infrastructures*.
- ESFRI, 2017, *Long-Term Sustainability of Research Infrastructures*, ESFRI Scripta Volume II ISBN PDF: 978-88-901562-8-1.
- European Commission, 2025, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy on Research And Technology Infrastructures*, COM (2025) 497 final/2.
- European Commission, 2017, *Sustainable European research infrastructures : a call for action: Commission staff working document: long-term sustainability of research infrastructures*, Publications Office. <https://data.europa.eu/doi/10.2777/76269>.
- Fricker M., 2007, *Epistemic injustice: Power and the ethics of knowing*, Oxford University Press.
- Hallonsten O., 2020, *Research Infrastructures in Europe: The Hype and the Field*, in *European Review*, 28(4), 617-635. doi:10.1017/S106279872000-009.
- Hoffert E., et al., 2025, *ESFRI Report: Funding of Research Infrastructures*, DOI 10.5281/zenodo.14770890.
- Kocar S., Kaczmirek L., 2023, *A meta-analysis of worldwide recruitment rates in 23 probability-based online panels, between 2007 and 2019*, in *International Journal of Social Research Methodology*, 27(5), 589-604. doi: 10.1080/13645579.2023.2242202.

Lynn P., 2012, *Longitudinal survey methods for the household finances and consumption survey*, Report prepared for the European Central Bank.

Menard S., 2002, *Longitudinal Research. Quantitative Applications in the Social Sciences*, Thousand Oaks, CA, SAGE Publications, Inc.

OECD, 2017, *Strengthening the effectiveness and sustainability of international Research Infrastructures*, in OECD Science, Technology And Industry Policy Papers, n. 48, OECD Publishing.

Parlamento Europeo 2013, Regolamento (UE) n. 1291/2013 - Articolo 2(6).

Pennacchiotti C., D'Ambrosio G., Primerano I., *Bringing together different data sources in Italy: the FOSSR project*, Plenary Session, 34.

Santurro M., et al., 2025, *Designing the Italian Online Probability Panel: Innovations and Challenges to Foster Open Science*, in di Bella E., Gioia V., Lagazio C., Zaccarin S. (eds), *Statistics for Innovation I*, in SIS 2025. Italian Statistical Society Series on Advances in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-031-96736-8_43.

Scherpenzeel A., 2011, *Data Collection in a Probability-Based Internet Panel: How the LISS Panel Was Built and How It Can Be Used*, in Bulletin of Sociological Methodology, 109, 56-61.

Taddei L., Paolucci M., (eds.), 2026, *Longitudinal Data Infrastructures in Europe: Tools for Open Science in Social Science Research*, Cham, Springer Nature Switzerland.

Zhang Y., Wildemuth B.M., 2009, *Qualitative Analysis of Content*, in Wildemuth B.M., (a c. di), *Applications of Social research methods to questions in information and library science*, Santa Barbara, Libraries Unlimited. Testo disponibile al sito: https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf (consultato il 15/04/2025).

Conclusioni

di *Mario Paolucci*

Eccoci, dunque, alla conclusione di questo libro. Studenti e studentesse, ricercatrici e ricercatori, portatori e portatrici di interesse in ambito politico e sociale, lettrici e lettori – questo volume vi ha accompagnato, attraverso una pletora di acronimi e innumerevoli riferimenti bibliografici e riferimenti web, alla scoperta di alcune complesse operazioni condotte dagli esseri umani per ottenere informazioni sulla società. Se l’iper-connessione vi dà l’impressione che si sappia tutto di tutto, sappiate invece che questo non è vero; che la società, le relazioni sociali, le condizioni di salute, le famiglie, le abitudini e i comportamenti cambiano nel tempo. E se si vuole conoscere per agire, o anche solo conoscere per capire, è bene che queste complesse operazioni – non una novità in sé, censimenti sono stati effettuati sin dall’epoca degli antichi Romani o della dinastia Han – vadano organizzate, sostenute, gestite. Questo volume propone una ricetta per tale lavoro abbastanza complicata da giustificare, per queste operazioni, il nome di infrastrutture (Taddei e Paolucci 2026).

L’idea di infrastruttura di ricerca si impone dove lo sforzo collettivo diventa preponderante su quello del singolo, ed è un buon grimaldello per smontare il dominio del sistema di incentivi creatosi dopo la Seconda guerra mondiale. Nel tempo, questo sistema, centrato sul successo e sulla valutazione individuale, ha prestato il fianco allo sfruttamento da parte di alcuni grandi editori (Grossmann e Brems 2021). Per porre termine a questo sfruttamento, si propone un ruolo maggiormente attivo nella pubblicazione da parte delle biblioteche. Ecco, se l’idea di “una vera e propria cultura infrastrutturale all’interno della comunità scientifica nazionale” vi lascia ancora

freddi, se dopo la lettura di questo libro avete bisogno di una messa a terra, può essere utile considerare che il prototipo di un'infrastruttura per le scienze sociali non è altro che una biblioteca. E se questo delle pubblicazioni scientifiche vi sembra un settore marginale, considerate che il meccanismo di cattura del sistema delle pubblicazioni scientifiche è lo stesso meccanismo di cattura oligopolistica che si è manifestato più tardi nella colonizzazione del web da parte dei colossi delle reti sociali (Doctorow 2025).

Abbiamo visto come le infrastrutture di ricerca rappresentino una risposta collettiva a queste logiche estrattive. Ma come si traduce questa visione in azione concreta? E da qui come si prosegue? Con le informazioni in questo libro, la lettrice ed il lettore possono costruire il loro panel longitudinale (se accade, contattateci, lo collegheremo volentieri ad IOPP). Oppure possono approfondire le materie presentate. Come spinta iniziale in questa, forse più fattibile direzione, concludo con spunti di approfondimento e di ricerca legati ai capitoli del libro in forma di domande.

Per quanto riguarda il primo capitolo, perché FAIR? Da dove viene la cultura dell'accessibilità e dell'apertura? Quali potrebbero essere gli svantaggi di quest'approccio se divenisse dominante? Come suggerimento, cos'è cambiato dal web a Facebook? Come possiamo assicurarci il funzionamento nel tempo di strumenti aperti come *Laravel* e *Lime survey*?

I capitoli 2 e 3 vi rendono in grado di realizzare il vostro panel, o almeno di evitare alcuni degli errori da noi commessi nel farlo; che fare con il logoramento (*attrition*) e il *bias* che questo introduce?

Nel capitolo 4 viene presentata la strategia di campionamento probabilistico, che viene anche comparata alle alternative. La differenza tra un campione probabilistico e uno non probabilistico è una scelta fondante che IOPP propone con forza. Ma vale la pena di analizzare anche questa scelta con spirito critico. Dove, e per quali domande di ricerca, un panel non probabilistico, più semplice da realizzare, sarebbe abbastanza? Dove e per quali domande causerebbe risposte errate?

Nel capitolo 5 i due assi del sistema degli incentivi, condizione e natura, presentati nella loro generalità, sono sicuramente meritevoli di approfondimenti. Nessuna scelta del progettista è realmente neu-

tra; le condizioni agiscono sul sistema dei valori dei potenziali intervistati, potenzialmente sbilanciando il campione. Qui si potrebbe approfondire la questione con esperimenti sul campo ed in laboratorio; in quali casi un incentivo funziona? In quali altera il campione?

Nel capitolo 6 avrete letto quali dimensioni sono state incluse nel questionario base di IOPP: le caratteristiche sociodemografiche e socioeconomiche dell'individuo, i principali tratti della personalità, gli orientamenti valoriali e politici, nonché le percezioni, le opinioni e i comportamenti individuali. Quali potrebbero essere altre domande di primaria rilevanza? Quali avrebbero potuto essere trascurate, ad esempio perché ricavabili in altro modo?

Per il capitolo 7, è evidente l'importanza del *Data Management Plan*. Cosa dovrebbe comprendere un buon *management plan* per un'infrastruttura? Che esempi si trovano online?

Per il capitolo 8 vorrei evidenziare criticamente il punto dell'eccellenza scientifica. Si tratta di una parola chiave che permea il discorso sulla rappresentazione della scienza, a partire dalla valutazione per finire con l'immagine pubblica della ricerca e di ricercatrici e ricercatori. Ora, le parole non sono neutre (Guccini 1979; Moretti 1989). Se la scienza va divisa tra eccellente e meno eccellente, chi traccia il confine? A che scopo?

Con questo vi lasciamo alle vostre ricerche, ai vostri dati, e magari al vostro futuro panel. In caso di dubbio, la biblioteca è sempre un buon punto di partenza.

Riferimenti bibliografici

Doctorow C., 2025, *Enshittification: Why Everything Suddenly Got Worse and What to Do about It*, First edition, New York, MCD, Farrar, Straus and Giroux.

Grossmann A, Björn B., 2021, *Current Market Rates for Scholarly Publishing Services*, in *F1000 Research*, 10, 20. doi:10.12688/f1000research.27468.2.

Guccini F., 1979, *Statale 17*, Album Concerto.

Moretti N., 1989, Palombella Rossa, Titanus Distribuzione.

Taddei L., Paolucci M. (a c. di), 2026, *Longitudinal Data Infrastructures in Europe: Tools for Open Science in Social Science Research*. Cham: Springer Nature Switzerland.

Gli autori*

Caterina Ambrosio è dottoranda in Scienze Sociali e Statistiche presso l'Università degli Studi di Napoli Federico II. La sua borsa di dottorato è finanziata dal progetto FOSSR - *Fostering Open Science in Social Science Research*. I suoi principali interessi di ricerca vertono su metodologia della ricerca sociale, sociologia digitale, *digital methods* e *digital awareness*.

Ferruccio Biolcati Rinaldi è professore di Sociologia presso l'Università degli Studi di Milano. La sua attività di ricerca si concentra sulla religiosità, il cambiamento dei valori e la metodologia della ricerca survey. È co-fondatore di spsTREND, un laboratorio di ricerca sul cambiamento sociale e politico. Biolcati Rinaldi è responsabile scientifico per l'Italia sia di *European Values Study* (EVS) sia di *World Values Survey* (WVS) e attualmente presiede il Comitato Scientifico di EVS.

Gabriella D'Ambrosio è dottore di Ricerca in Comunicazione, Ricerca, Innovazione – Curriculum in Metodologia delle Scienze Sociali presso Sapienza Università di Roma. È attualmente Ricercatrice presso l'Istat - Istituto Nazionale di Statistica.

Giovanna De Santis è *Data Curator* presso DASSI - *Data Archive for Social Sciences in Italy*. È laureata in Sociologia e ha conseguito un dottorato nel programma di Studi Urbani URBEUR, dove ha studiato le trasformazioni urbane e la mobilità residenziale. Il suo lavoro si concentra sulla gestione dei dati nelle scienze sociali, sviluppando strumenti e workflow per l'archiviazione e la documentazione dei dati di ricerca. I suoi principali interessi riguardano l'*housing*, la mobilità residenziale e il data management.

Dario Germani è dottore di Ricerca in Metodologia delle Scienze Sociali. È stato assegnista di ricerca nel 2022 presso CNR-IRCrES e presso l'Università degli Studi di

* Per Gabriella D'Ambrosio, Nicolò Marchesini e Luciana Taddei si specifica che pareri e opinioni nel contributo hanno carattere personale e quanto espresso non rappresenta e non impegna l'Istituto di appartenenza.

Cagliari nel 2023. Attualmente è tecnologo III livello presso CNR-IRPPS nell'ambito del progetto FOSSR – *Fostering Open Science in Social Science Research*.

Frank Heins, geografo e demografo, dal 2001 al 2022 è stato Primo Ricercatore e Dirigente di ricerca del Consiglio Nazionale delle Ricerche presso l'Istituto di Ricerche sulla Popolazione e le Politiche Sociali (CNR-IRPPS). Le sue ricerche sono focalizzate sugli aspetti territoriali delle strutture e delle dinamiche dei processi demografici e socio-economici, in particolare le migrazioni, sia interne sia internazionali.

Nicolò Marchesini è demografo presso l'Istituto Nazionale di Statistica (Istat). Esperto di analisi statistiche quantitative e di pianificazioni di indagini, i suoi principali interessi di ricerca riguardano la costruzione e la collezione dei dati in ottica femminista e intersezionale, e le questioni di genere da una prospettiva maschile in particolare riguardo l'evoluzione delle maschilità e delle paternità.

Rocco Paolillo è ricercatore presso CNR-IRPPS all'interno del progetto FOSSR. I suoi interessi di ricerca includono la produzione di dati sintetici e l'applicazione di metodi computazionali come simulazioni e *text-as-data* allo studio di politiche e fenomeni sociali.

Mario Paolucci è direttore dell'Istituto di Ricerche sulla Popolazione e le Politiche Sociali (IRPPS) del CNR. Svolge ricerca interdisciplinare nel campo delle scienze sociali computazionali e dei modelli ad agenti.

Angela Papparuso, demografa, è Prima Ricercatrice del Consiglio Nazionale delle Ricerche presso l'Istituto di Ricerche sulla Popolazione e le Politiche Sociali (CNR-IRPPS). I suoi principali interessi di ricerca si focalizzano sui processi e sulle politiche di immigrazione e di integrazione e sul benessere soggettivo degli immigrati in Italia e in Europa.

Francesco Piacentini è tecnologo presso il Dipartimento di Scienze Sociali e Politiche dell'Università degli Studi di Milano. Si occupa di ricerca survey, comportamento elettorale e partiti politici

Michele Santurro, dottore di Ricerca in Scienze Sociali Applicate, è attualmente ricercatore a tempo determinato di terzo livello presso l'Istituto di Ricerche sulla Popolazione e le Politiche Sociali del Consiglio Nazionale delle Ricerche. I suoi principali interessi di studio riguardano le disuguaglianze sociali e le tematiche demografiche e migratorie, nonché gli aspetti metodologici e statistici connessi alla ricerca sociale.

Domingo Scisci è *Data Manager* presso DASSI – *Data Archive for Social Sciences in Italy*, *Service Provider* italiano di CESSDA ERIC. Si occupa di gestione, valorizzazione e distribuzione dei dati per le scienze sociali. Progetta soluzioni tecniche per ottimizzare i workflow di data management e sviluppa software e infrastrutture interoperabili, automatizzando processi e migliorando accesso e riuso dei dati per la comunità scientifica.

Valentina Tudisca, dottorata in fisica e con un master in comunicazione della scienza, è Ricercatrice dell'Istituto di Ricerche sulla Popolazione e le Politiche Sociali del Consiglio Nazionale delle Ricerche (CNR-IRPPS), dove coordina il gruppo di ricerca Studi Sociali sulla Scienza, Educazione, Comunicazione, la cui principale finalità è la comprensione e la valorizzazione delle relazioni tra scienza, politica e società.

Cristiano Vezzoni è professore di Sociologia Politica presso l'Università degli Studi di Milano e responsabile scientifico del laboratorio spsTREND. Dal 2021 presiede ITANES (*Italian National Election Studies*). A livello internazionale ha lavorato e insegnato nelle università di Groningen (NL), Essex (UK) e Leuven (BE). I suoi interessi di ricerca riguardano il comportamento elettorale, l'opinione pubblica e la metodologia, con particolare attenzione alla pratica della survey.

Francesco Visconti è ricercatore *Tenure Track* in Scienza Politica alla Luiss Guido Carli di Roma e membro del Centro Italiano Studi Elettorali. Ha lavorato come *research assistant* presso la University of Leicester, come assegnista di ricerca all'Università Statale di Milano e come ricercatore presso il Consiglio Nazionale delle Ricerche. Si occupa di analisi dell'opinione pubblica, del comportamento di voto, della solidarietà europea, dell'immigrazione e del cambiamento climatico.

Questo 
LIBRO

 ti è piaciuto?

Comunicaci il tuo giudizio su:
www.francoangeli.it/opinione



**VUOI RICEVERE GLI AGGIORNAMENTI
SULLE NOSTRE NOVITÀ
NELLE AREE CHE TI INTERESSANO?**



ISCRIVITI ALLE NOSTRE NEWSLETTER

SEGUICI SU:



FrancoAngeli

La passione per le conoscenze

Vi aspettiamo su:

www.francoangeli.it

per scaricare (gratuitamente) i cataloghi delle nostre pubblicazioni

DIVISI PER ARGOMENTI E CENTINAIA DI VOCI: PER FACILITARE
LE VOSTRE RICERCHE.



Management, finanza,
marketing, operations, HR

Psicologia e psicoterapia:
teorie e tecniche

Didattica, scienze
della formazione

Economia,
economia aziendale

Sociologia

Antropologia

Comunicazione e media

Medicina, sanità



Architettura, design,
arte, territorio

Informatica, ingegneria
Scienze

Filosofia, letteratura,
linguistica, storia

Politica, diritto

Psicologia, benessere,
autoaiuto

Efficacia personale

Politiche
e servizi sociali



FrancoAngeli

La passione per le conoscenze

Il volume descrive la nascita e lo sviluppo dell'*Italian Online Probability Panel* (IOPP), il primo panel probabilistico online in Italia, realizzato nell'ambito del progetto *Fostering Open Science in Social Science Research* (FOSSR) del Consiglio Nazionale delle Ricerche (CNR). IOPP rappresenta uno strumento, conforme ai principi FAIR, accessibile alla comunità scientifica per osservare i cambiamenti sociali nel tempo. Attraverso sfide, scelte e soluzioni, il libro ripercorre le fasi di costruzione dell'infrastruttura di ricerca: uno spazio collettivo per sostenere l'innovazione metodologica e il dialogo tra scienza e società.

Loredana Cerbara è ricercatrice senior, esperta in statistica per le scienze sociali, si occupa di disegni di indagine e metodi di analisi statistica. Referente CNR-IRPPS per l'accredito al Comstat e per il trattamento dati, è membro del sottogruppo CID Etica, Unità di supporto per i profili di privacy di progetto del Centro Interdipartimentale per l'Etica e l'Integrità nella Ricerca del CNR; è responsabile della realizzazione di IOPP ed autrice di oltre 100 pubblicazioni a carattere scientifico.

Claudia Pennacchiotti è tecnologa presso CNR-IRPPS, dove lavora nel progetto FOSSR e alla costruzione di IOPP. Laureata in Filosofia, si occupa di comunicazione della scienza, comprensione e valorizzazione delle relazioni tra scienza, politica e società, *Open Science*, sistemi educativi e pratiche di educazione aperta, inclusione e didattica partecipata mediante l'uso di metodologie volte al coinvolgimento dei vari attori sociali.

Ilaria Primerano è tecnologa presso CNR-IRPPS, dove collabora al *Work Package 4 Improving longitudinal data infrastructures in Italy* del progetto FOSSR e alla realizzazione di IOPP. Ha conseguito il dottorato in Ingegneria ed Economia dell'Innovazione e svolge la sua attività di ricerca nell'ambito della statistica sociale, occupandosi di analisi di strutture di dati complesse integrando Analisi Multidimensionale dei dati e *Social Network Analysis*.

Luciana Taddei è metodologa, specializzata in sociologia e ricerca sociale avanzata, con un dottorato conseguito in cotutela tra Genova e Buenos Aires. Insegna all'Università Magna Græcia di Catanzaro e all'Università della Calabria ed è autrice di oltre 50 pubblicazioni scientifiche. Presso il CNR-IRPPS, ha contribuito allo sviluppo del progetto FOSSR, e alla progettazione e all'avvio di IOPP. Dal 2026 è ricercatrice presso Istat.