

La condivisione dei dati *deve* diventare prassi comune per la ricerca psicologica
Data Sharing *Must* Become Common Practice for Psychological Research

Cristina Zogmaister*
Dipartimento di Psicologia
Università di Milano - Bicocca

*Cristina Zogmaister
Dipartimento di Psicologia
Università di Milano - Bicocca
Piazza dell'Ateneo Nuovo, 1
20126 Milano
E-mail: cristina.zogmaister@unimib.it

Riassunto

La condivisione dei dati di ricerca è un aspetto fondamentale della scienza, ma troppi psicologi manifestano ancora riluttanza a praticarla. La prima parte di questo lavoro discute alcuni importanti vantaggi della condivisione dei dati, sia per la ricerca psicologica nel suo complesso, sia per il ricercatore che la mette in atto. I risultati empirici, anche quelli più recenti, mostrano che la condivisione dei dati è frenata da diversi timori e spesso anche da un'insufficienza di conoscenze sui modi con cui attuarla. Anche alla luce della considerazione che timore e scarsa conoscenza spesso vanno a braccetto, la seconda parte del lavoro fornisce le indicazioni di base per orientarsi nelle diverse pratiche di condivisione dei dati.

Abstract

Data sharing is a key aspect of science, but too many psychologists are reluctant to practice it. The first part of the present work discusses some important advantages of data sharing, both for psychological research as a whole, and for the researcher. Even the most recent empirical results show that data sharing is slowed down by fears and by a lack of knowledge on how to implement it. Also in the light of the consideration that fear and lack of knowledge often go arm in hand, the second part provides some basic information on how to share one's own data.

Keywords

Open Data; Open Science; Data Sharing; Data Reuse

La condivisione dei dati *deve* diventare prassi comune

Storicamente, gli psicologi hanno sempre manifestato una scarsa propensione a condividere i dati delle loro ricerche. Wolins, nel 1962, descrisse l'esperienza scoraggiante di un suo studente di master, che chiese agli autori di 37 articoli empirici pubblicati in riviste APA tra il 1959 e il 1961 di poter avere i dati relativi a questi articoli per uno studio per sua tesi e ottenne risposta affermativa da solo nove di essi. Una decina d'anni dopo, Craig e Reese (1973) riscontrarono risultati simili, ottenendo i dati da due quinti dei 53 autori contattati. In epoca più recente, Wicherts, Borsboom, Kats e Molenaar (2006) contattarono gli autori di 141 articoli pubblicati su riviste APA, chiedendo di poterne analizzare i dati, per testare l'effetto della presenza di outlier sulla solidità dei risultati. La loro richiesta era corredata dalla descrizione dello studio, che era stato approvato dal comitato etico, e i richiedenti garantivano che non avrebbero condiviso i dati con nessun altro. Nonostante tutto questo, e sebbene gli autori avessero firmato un documento di impegno a fornire i dati se richiesti per analisi (questo infatti è un obbligo standard richiesto dalle riviste APA), Wicherts e colleghi ottennero i dati per soltanto un quarto degli studi e non furono perciò in grado di eseguire la ricerca che si erano prefissi. In un articolo scientifico che risale a trentacinque anni fa, Ceci e Walker (1983, cfr. anche Ceci, 1988) misero in luce alcune delle ragioni della scarsa propensione degli psicologi a condividere i dati delle loro ricerche: preoccupazioni di natura etica e relative alla sicurezza dei dati; paura che nuove analisi portassero alla luce errori nelle analisi effettuate; timore che, in seguito alla divulgazione dei dati, altri facessero delle ricerche simili togliendo valore ai dati stessi; costi legati alla condivisione.

Gli accadimenti che hanno caratterizzato il mondo scientifico - ed in particolare l'ambito della psicologia - nell'ultimo decennio, hanno portato a una profonda riflessione sul nostro modo di fare scienza (cfr. Grassi, 2018; Perugini, 2014, 2018), ad un ripensamento sulle condotte da adottare nei nostri laboratori (Crepaldi, 2018) e a una riflessione sull'importanza dell'uso di buone pratiche di analisi statistica (Agnoli, 2018); anche grazie ai concomitanti sviluppi tecnologici ed informatici, abbiamo assistito allo

sviluppo di modi totalmente nuovi di fare ricerca scientifica. La condivisione dei dati è un aspetto centrale del cambiamento a cui stiamo assistendo ed è un aspetto necessario ai fini della trasparenza (Perugini, 2014). Per chi è dentro il movimento Open Science, la condivisione dei dati è considerata qualcosa di ovvio, indispensabile, da dare quasi per scontato. Tant'è vero che negli ultimi anni abbiamo assistito al nascere dei progetti di ricerca multi-laboratorio, caratterizzati dalla collaborazione tra numerosi gruppi di scienziati sparsi in tutto il mondo (es. Klein et al., 2014; Ebersole et al., 2016). Questi progetti, che permettono di ottenere dataset grandi ed eterogenei, garantiscono maggiore potenza e generalizzabilità, e in ultima analisi maggiore fiducia nei risultati, non potrebbero nemmeno esistere se non ci fosse condivisione (dei dati, delle procedure, di tutto). Nuova enfasi è stata posta sul tema del riuso dei dati (Tenopir et al., 2015). In vari campi, si stanno facendo spazio ricercatori che vengono chiamati, in tono (non sempre) scherzoso, parassiti e simbionti (Longo e Drazen, 2016), i quali grazie alle analisi secondarie danno nuovo valore a dati che, in molti casi, avrebbero altrimenti cessato di essere utili. Si assiste al sorgere grandi progetti collaborativi di analisi, basati sul riuso di dati già disponibili (Martone, Garcia-Castro, & VandenBos, 2018; NIMH Collaborative Data Synthesis for Adolescent Depression_Trials Study Team, 2013): questi progetti sostituiscono la pratica della meta-analisi tradizionale con analisi secondarie più sofisticate e potenti, effettuate direttamente sui dati grezzi (cfr. Martone e coll., 2018).

Dall'aggregazione e dal riuso di dati diversi emergono opportunità nuove. Per esempio, con campioni molto grandi è possibile vedere le osservazioni atipiche in una luce diversa: se nei piccoli campioni gli outlier sono essenzialmente fonti di disturbo, che spesso obbligano a ricorrere a strategie per arginarne l'effetto distorcente sui risultati (Aguinis, Gottfredson, & Joo, 2013), in campioni grandi le code delle distribuzioni possono divenire, esse stesse, oggetto di studio e permettere di capire meglio i comportamenti o le caratteristiche più rare, così come avviene in altri ambiti di analisi di big data (Monteith, Glenn, Geddes e Bauer, 2015). Inoltre, l'integrazione di dati provenienti da fonti diverse e misurati con tecniche diverse può permettere una migliore comprensione dell'oggetto di studio (Ferguson, Nilson, Cragin, Brandowski e Martone, 2014). In breve, condividere i dati non aumenta solo la trasparenza e la fiducia nei risultati scientifici, ma facilita l'impresa scientifica nel suo complesso.

Mettere a disposizione i propri dati, in maniera ordinata e comprensibile anche a terzi, richiede un certo impegno. Non tutti sono così altruisti da farlo ‘per il bene della scienza’. È allora importante sottolineare che la condivisione ha anche diversi benefici diretti per chi condivide. È stato evidenziato, in diversi ambiti scientifici, che gli studi corredati da open data hanno maggiore visibilità e ricevono un maggior numero di citazioni rispetto a studi simili, ma non accompagnati dai dati (McKiernan et al., 2016), senza contare che, se altri ricercatori decidono di effettuare analisi secondarie, è possibile che chi ha messo a disposizione i dati venga coinvolto nelle analisi e nell’eventuale pubblicazione. Inoltre, un numero crescente di riviste incoraggia a rendere disponibili i dati, riconoscendo spesso tale disponibilità con appositi badge. Alcune riviste richiedono obbligatoriamente che i dati vengano resi pubblici, rendendo di fatto impossibile la pubblicazione a chi non si è preparato per tempo a questa condivisione. Anche molti enti che finanziano la ricerca sono sensibili alle tematiche dell’Open Science e richiedono che i dati siano resi pubblicamente disponibili. È plausibile che, con il passare del tempo, la maggioranza dei ricercatori renderà pubblici i propri dati e chi non lo farà diventerà una mosca bianca, che faticherà sempre più a trovare riviste su cui pubblicare le proprie ricerche. Vale anche la pena di ricordare che stanno nascendo riviste espressamente dedicate alla pubblicazione di *data papers*, ossia articoli che descrivono dataset: un esempio per il nostro settore è il *Journal of Open Psychology Data* (<https://openpsychologydata.metajnl.com/>). Ogni *data paper* è corredato da un DOI, come gli altri articoli scientifici, ed è perciò citabile da chi usa i dati.

In breve, i cambiamenti intercorsi negli ultimi decenni hanno portato al centro dell’attenzione l’importanza, i vantaggi e le difficoltà della condivisione dei dati di ricerca (cfr. anche Crepaldi, 2018; Grassi, 2018). Come ha inciso tutto questo sulla propensione degli psicologi a condividere i loro dati?

Gli psicologi condividono i dati di ricerca?

In epoca immediatamente successiva alla crisi di credibilità che ha scosso la ricerca scientifica (Pashler & Wagenmakers, 2012), nel 2013, Vanpaemel, Vermorgen, Deriemaecker e Storms (2015) hanno chiesto agli autori di quasi quattrocento articoli, pubblicati in quattro riviste APA nel 2012, l’accesso ai dati. Solo una minoranza – circa due quinti dei contattati – ha risposto inviando i dati. Molti non hanno dato

alcuna risposta e circa un autore su cinque ha esplicitamente rifiutato di condividerli o ha dichiarato di non poterlo fare. Tenopir e colleghi (2015), confrontando le risposte date a un questionario da un campione di psicologi autori di articoli scientifici nel 2009-2010 e nel 2013-14, hanno riscontrato un leggero miglioramento nella disponibilità a condividere i dati, ma anche in questa ricerca la maggior parte degli intervistati dichiarava di non aver mai reso disponibili i propri dati ad altri o di averne condiviso solo alcuni. In epoca ancora più recente, Hardwicke e Ioannidis (2018) arrivano a risultati altrettanto scoraggianti con il progetto Data Ark.

Il progetto Data Ark. Poiché i dati di gran parte delle ricerche psicologiche non sono disponibili in modalità Open Access, ossia in un archivio pubblico da cui siano facilmente recuperabili, Hardwicke e Ioannidis hanno avviato un progetto per l'archiviazione dei dati scientifici in un deposito online, il Data Ark (<https://osf.io/view/DataArk/>). Hanno deciso di iniziare con i 111 articoli più citati in ambito psicologico e psichiatrico tra il 2006 e il 2016. Per prima cosa, hanno verificato per quanti di questi articoli i dati fossero disponibili in un archivio pubblico e hanno trovato che lo erano in soli 22 casi: in 17 casi i dati erano disponibili pubblicamente senza restrizioni e in 5 casi con restrizioni all'accesso. Per i restanti 89 articoli, Hardwicke e Ioannidis hanno contattato gli autori, spiegato il progetto e suggerito di condividere i dati con una di tre modalità: (a) caricare i dati sull'archivio Data Ark, con accesso libero; (b) caricare su Data Ark, con accesso ristretto (gli autori che sceglievano quest'opzione dovevano specificare le restrizioni richieste); (c) condividere i dati solo con il team di ricerca che ne faceva richiesta. Se gli autori non erano disponibili a condividere i loro dati in nessuna delle tre modalità, veniva chiesto di spiegare le ragioni di questo rifiuto. La richiesta di Hardwicke e Ioannidis ha portato alla disponibilità senza limitazioni per 10 studi e un altro è stato reso disponibile con restrizioni su Data Ark, mentre i dati relativi a due studi sono stati resi disponibili solo al team di ricerca. Complessivamente, si è arrivati a una qualche forma di disponibilità per 35 dei 111 studi: meno di un terzo. La grande maggioranza dei dati continua a rimanere non disponibile pubblicamente. L'aspetto che induce maggior pessimismo è che, differentemente da quanto era stato riscontrato Tenopir e colleghi (2015), Hardwicke e Ioannidis non hanno trovato risultati sostanzialmente diversi per i dati relativi a ricerche pubblicate tra il 2006 e il 2011 e quelli di ricerche pubblicate tra il 2014 e

il 2016. Questi risultati trovano conferma nei risultati di una ricerca di self report di Houtkoop e coll. (2018), alla quale hanno partecipato 600 ricercatori che hanno pubblicato recentemente articoli in ambito psicologico: la maggioranza infatti ha risposto di aver condiviso i dati in un archivio online per meno del 10% dei propri progetti, solo una minoranza indicava di essere intenzionata a farlo nel futuro.

Nel complesso, questi risultati mostrano un quadro preoccupante, evidenziando che molto poco è cambiato, rispetto all'attitudine alla condivisione che si registrava in passato. Quello della condivisione dei dati deve rimanere un tema centrale nella discussione, perché non è ancora una pratica che si possa dare per scontata.

Perché i ricercatori sono così riluttanti a condividere i dati? Le ricerche empiriche evidenziano ragioni di varia natura e per molti versi simili a quelle elencate da Ceci e Walker nel 1983. In alcuni casi il rifiuto di rendere disponibili i dati delle proprie ricerche è legato al non avere accesso ai dati stessi, averli persi (per esempio, perché s'è rotto il computer), o non esserne i proprietari. C'è inoltre la questione della mancanza di tempo per preparare i dati e metadati da rendere disponibili, oltre al timore che questo sforzo sia inutile: alcuni ricercatori temono infatti che lo sforzo non valga il beneficio, poiché i dati non serviranno a nessun altro, dato che, secondo loro, tutte le analisi che potevano essere fatte sono già state fatte. Alcuni giustificano il rifiuto di fornire i dati con la presenza di regolamenti locali che vietano la condivisione, o la mancanza di un compenso economico per la condivisione, altri affermando di non averne l'autorizzazione (per esempio l'approvazione del comitato etico o il consenso dei partecipanti non include un permesso esplicito a condividere i dati).

Talvolta i ricercatori esprimono preoccupazioni legate alla tutela della privacy dei partecipanti. In alcune situazioni la condivisione dei dati grezzi completi può essere obiettivamente inappropriata. Questo accade quando i dati contengono informazioni sensibili. Anche eliminando dai file le variabili che identificano i partecipanti, c'è il rischio che qualcuno riesca a ri-identificarli ricorrendo ad archivi esterni (Lundberg, Narayanan, Levy, & Salganik, 2018). In alcuni casi, per esempio nei casi in cui il dato grezzo è costituito da riprese video di persone che raccontano informazioni molto intime, anonimizzare i filmati è molto difficile.

Ci sono poi una serie di timori: che qualcuno trovi errori nei dati o nelle statistiche e ciò abbia un effetto negativo per l'autore¹, che altri ricercatori facciano le analisi pianificate dall'autore che li ha raccolti e lo precedano nella pubblicazione, o infine che i dati siano troppo complicati per essere compresi adeguatamente da terzi e perciò da eventuali analisi secondarie possa scaturire solo cattiva ricerca (cfr. Martone et al., 2018; Hardwicke e Ioannidis, 2018; Houtkoop et al., 2018; Vanpaemel et al., 2015).

Come avviene per altri ambiti del comportamento umano, è plausibile che alcuni dei timori e delle reticenze che ostacolano la condivisione dei dati siano legati a un'insufficiente conoscenza delle modalità attraverso le quali tale condivisione può avvenire e degli strumenti per proteggersi da conseguenze indesiderate. La mancanza di conoscenze adeguate, infatti, viene talvolta annoverata tra le cause per cui non si condivide (Houtkoop e coll., 2018). Molti degli ostacoli evidenziati qui sopra possono essere superati – almeno per le ricerche future – se la possibilità di rendere pubblici i dati viene considerata fin dall'inizio della ricerca. Ecco allora, qui di seguito, alcune informazioni di base che possono essere utili in questo processo.

Condividere i dati: dove, come e quando

È possibile condividere i propri dati scientifici in modi diversi, che possiamo immaginare collocati lungo un continuum. Ad un estremo, i dati di una ricerca scientifica vengono posti in un archivio pubblico online o nello spazio web messo a disposizione da una rivista e chiunque voglia può scaricarli ed utilizzarli, senza limitazioni di scopo o di ri-distribuzione. All'estremo opposto, lo scienziato non salva i dati in uno spazio pubblico ma, qualora altri ricercatori lo richiedano, li mette a loro disposizione esclusivamente per ripetere le analisi effettuate per gli specifici scopi della ricerca per cui sono stati raccolti. La possibilità di non condividere i dati in nessun modo non è presente in questo continuum per la ragione che tutti i dati dovrebbero essere disponibili almeno per la verifica delle analisi, prima di tutto per questioni etiche e di trasparenza della scienza e poi perché, in molti casi, il ricercatore si impegna esplicitamente a rendere disponibili i dati nel pubblicare i propri articoli (in particolare sulle riviste APA o che ne seguono il modello).

Dove condividere?

¹ È tuttavia plausibile che, con la diffusione delle pratiche di condivisione, la scelta di non condividere pubblicamente i dati sarà guardata con sospetto e potrà portare, paradossalmente, a maggiori richieste di controllo dei dati stessi.

Tenere i dati esclusivamente su supporti personali, per esempio sul proprio computer o altre memorie a cui soltanto il ricercatore ha accesso, in generale non è una buona idea. Per prima cosa, ci sono ragioni pratiche: i supporti si danneggiano, vengono persi, diventano obsoleti; il ricercatore stesso, un giorno, si godrà il meritato riposo in lidi soleggiati e diventerà irraggiungibile per i colleghi scienziati. Un dato condiviso esclusivamente su richiesta, inoltre, ha minor probabilità di entrare nel circolo virtuoso del riuso. La scelta, pertanto, dovrebbe essere tra (1) mettere i dati in un archivio/repository, con un indirizzo web o un Digital Object Identifier (DOI) che permetta di collegarli all'articolo in cui vengono descritti i risultati delle analisi; (2) caricarli sul sito internet della rivista, come materiali supplementari di una pubblicazione di testo; (3) caricarli su siti internet personali o istituzionali.

Tra le tre, l'opzione di pubblicare i dati sul proprio sito personale è sconsigliabile perché è quella che fornisce minor sicurezza, garanzia di persistenza nel tempo, visibilità. Pubblicare i dati in un archivio è, in generale, una scelta migliore rispetto a caricarli come materiali supplementari, per due ragioni. La prima ragione è che i dati pubblicati come materiali supplementari non sono identificati come oggetti digitali separati dal testo dell'articolo e perciò è possibile trovarli solo attraverso la corrispondente pubblicazione di testo. I dati pubblicati negli archivi, invece, sono inclusi in cataloghi online e possono perciò essere ritrovati sia a partire dall'articolo scientifico sia attraverso ricerche di dati: questo ne accresce la visibilità e il potenziale contributo alla ricerca scientifica. La seconda ragione è, a differenza di ciò che accade con i materiali supplementari caricati sui siti delle riviste, molti archivi – ma non tutti – permettono la scelta del tipo di licenza da accordare ai dati, consentendo un maggiore controllo sulle modalità di condivisione (su questo argomento tornerò in seguito).

Se si decide di caricare i propri dati in un archivio, si potrà optare per uno di tipo generalista oppure un archivio disciplinare. Nello scegliere tra gli archivi disponibili è importante tener conto di varie caratteristiche. Una di queste è la localizzazione geografica², poiché la legislazione vigente, i regolamenti della propria organizzazione o le linee guida dei finanziatori possono imporre dei vincoli a questo proposito.

² Più in particolare, è importante verificare in quale Paese vengono immagazzinati i dati, poiché esistono archivi, come quello dell'Open Science Framework, che permettono di scegliere di archiviare i propri dati in Paesi diversi.

Un'altra caratteristica che può essere utile tenere in considerazione è se l'archivio fornisca un DOI. Il sistema DOI infatti identifica i dati in maniera permanente, fornisce un facile accesso online ed inoltre è un sistema ampiamente riconosciuto, che permette l'inserimento dei dati nelle banche dati e nei sistemi di misurazione delle citazioni. Infine, è importante verificare con quale tipo di licenza i dati vengono condivisi. Alcuni archivi permettono di scegliere la licenza, mentre altri ne impongono una, non sempre adeguata ai dati che si vogliono condividere. Gli archivi si differenziano inoltre per i tipi di servizi offerti (per esempio, alcuni si limitano ad immagazzinare i dati mentre altri forniscono un servizio di controllo sulla qualità tecnica ed accessibilità del materiale), per l'essere gratuiti o a pagamento, nonché per una serie di caratteristiche tecniche. Per orientarsi tra le diverse possibilità di archiviazione, consiglio di consultare il registro degli archivi <https://www.re3data.org/>. Questo registro permette di effettuare ricerche mirate sulla base delle caratteristiche che ho elencato e anche di altre.

Come condividere?

Ci sono alcune semplici prassi, che potremmo chiamare di buon senso, che è opportuno seguire nell'archiviare i propri dati per la condivisione. Tra queste, mi limito a citare l'opportunità di condividere i dati grezzi (ossia non trasformati), in un formato aperto – come possono essere per esempio i formati .txt o .csv nel caso di dati alfanumerici – che non richieda software specifici (e a volte costosi); accompagnarli con adeguati meta-dati che descrivano chiaramente la struttura dei dati e il significato delle variabili, e contengano il riferimento all'articolo scientifico a cui fanno riferimento; corredarli se possibile con le sintassi utilizzate per le analisi riportate nell'articolo. È importante che la documentazione sia abbastanza completa da permettere anche chi non ha partecipato alla raccolta di comprendere bene i dati ed effettuare future analisi.

Una questione importante è la licenza. Assegnare una licenza ai dati permette, da un lato, di comunicare agli altri ricercatori *se e a quali condizioni* possono utilizzarli e, dall'altro, di assicurarsi che chi li usa riconosca appropriatamente l'autorship a chi li ha condivisi. Accordare una licenza che limiti le possibilità d'utilizzo può inoltre aiutare a rispondere alle preoccupazioni etiche che i ricercatori possono porsi sui modi in cui i loro dati di ricerca possono essere usati da altre persone.

Una volta che ci si è assicurati di avere il diritto di assegnare una licenza ai dati (o in altre parole che non siano terzi ad avere il diritto su quei dati), è importante verificare che il tipo di licenza imposto dalla rivista o dall'archivio sia adatto alle caratteristiche dei dati e conforme alle condizioni stabilite nel documento di consenso informato, alle linee guida eventualmente fornite dall'ente finanziatore e alla legislazione vigente o, nel caso si possa scegliere tra licenze diverse, optare per quella più adatta. Le licenze più usate nell'ambito scientifico sono le Creative Commons (2016), perciò vale la pena di spendere qualche parola su come funzionano. La licenza di base (indicata come CC BY) richiede che chi utilizza o condivide i dati ne attribuisca l'autorship al creatore. Ci sono altre cinque licenze che combinano l'obbligo di riconoscimento dell'autorship con uno o più elementi aggiuntivi: il divieto di usare commercialmente i dati (indicato dalla sigla NC, Non-Commercial), il divieto di modificarli (ND, che significa No Derivative Works), l'obbligo di ri-condividere eventuali adattamenti del materiale con la stessa licenza (SA, ossia Share-Alike). Esiste, infine, una licenza CC0, che consiste nella rinuncia a tutti i diritti sui dati: se viene concessa questa licenza, chiunque potrà usare e ridistribuire i dati senza restrizioni, anche per usi commerciali. Anche nel caso della licenza CC0, tuttavia, vanno rispettate le regole imposte dalle leggi sulla protezione dei dati e andrebbero rispettate le regole scientifiche circa l'attribuzione dell'autorship.

Un discorso a parte va fatto per le situazioni in cui i dati non possano essere resi pubblicamente disponibili. Ci sono legittime eccezioni alla condivisione pubblica dei dati, legate alla legislazione vigente, agli accordi presi tra ricercatore e partecipanti nel momento del consenso informato, alla tutela della privacy e alla sicurezza di chi – persona o organizzazione – ha fornito il dato. Anche in questi casi, tuttavia, i dati andrebbero resi disponibili alla comunità scientifica, almeno per permettere il controllo delle analisi statistiche effettuate. Una soluzione è quella di anonimizzare i dati per quanto possibile e immagazzinarli in un archivio senza renderli pubblici, condividendoli solo su richiesta e con adeguata descrizioni degli scopi per i quali possono essere utilizzati (nella condizione più restrittiva: solo per verificare le analisi effettuate nella ricerca), chiedendo agli utilizzatori secondari di firmare un accordo che ne vieti la divulgazione e imponga la cancellazione dei dati grezzi dopo un periodo di tempo prestabilito (cfr. Schönbrodt, Gollwitzer e Abele-Brehm 2016, per una proposta di licenza di questo tipo).

Quando condividere?

In generale, i dati vanno resi disponibili alla comunità scientifica insieme all'articolo scientifico che descrive la ricerca e vanno condivisi con i revisori durante il processo di valutazione dell'articolo per la pubblicazione. Come si è visto in precedenza, però, una delle cause per cui i ricercatori non condividono i loro dati è l'intenzione di fare nuove analisi in futuro. Questo può succedere per esempio nel caso di un grande insieme di variabili, quando un ricercatore scrive una prima pubblicazione che riguarda solo una parte di esse. Quella di tutelarsi dal fatto che altri facciano prima di noi le analisi per cui avevamo raccolto i dati, oppure dopo aver visto i nostri dati effettuino e pubblichino ricerche simili, diminuendo così il valore dei dati stessi per chi li ha raccolti, è una preoccupazione legittima. Una possibile soluzione è rendere pubblici da subito solo i dati relativi alle variabili analizzate nell'articolo pubblicato e prevedere un periodo d'embargo per il dataset completo. Un'altra soluzione è rendere inizialmente disponibili i dati soltanto su richiesta e con condizioni d'uso restrittive (cfr. quanto scritto nella sezione precedente a proposito delle licenze d'uso) e renderli pubblici dopo un embargo temporale.

La questione della condivisione dei dati, però, va considerata ben prima della pubblicazione dell'articolo scientifico (Crepaldi, 2018). È cruciale considerare le problematiche relative alla condivisione già nella stesura del documento di consenso informato e nella richiesta di parere al Comitato Etico (cfr. Meyer, 2018; Schönbrodt et al., 2016). In particolare, talvolta nel documento di consenso informato e/o nella richiesta al Comitato Etico si indica che i dati verranno utilizzati soltanto per determinati e precisi scopi scientifici, quelli della ricerca iniziale: questo pone vincoli al riuso che sono spesso inutili, mentre è plausibile che i partecipanti sarebbero favorevoli al fatto che i loro dati venissero in futuro riutilizzati per altri studi scientifici. Sarebbe meglio specificare in questi documenti che i dati verranno usati per gli scopi indicati dal ricercatore e potranno in futuro essere usati dalla comunità scientifica per altre ricerche. Analogamente, il vincolo secondo cui 'solo il ricercatore e i suoi collaboratori avranno accesso ai dati' è problematico e andrebbe sostituito con un'informazione sui modi con cui altri avranno accesso ai dati - per esempio, specificando se saranno resi disponibili in maniera pubblica / su richiesta di altri ricercatori, che verranno eliminate le informazioni identificative (come nome, cognome e data di nascita) e descrivendo le

condizioni a cui i dati potranno essere riutilizzati. In alcuni casi il documento di consenso informato vincola a cancellare i dati raccolti dopo un determinato periodo di tempo: anche questo pone un obbligo spesso inutile. In certi casi, i dati personali possono essere indispensabili, per esempio per studi longitudinali, ma in molti altri casi la distruzione delle informazioni sull'identità dei partecipanti dopo un certo lasso di tempo può essere ragionevole ai fini della tutela della privacy. Perciò se le informazioni identificative dei partecipanti non hanno un'utilità scientifica è ragionevole impegnarsi a eliminare queste, anziché promettere di cancellare tutti i dati.

Poiché siamo in un periodo di transizione, può accadere che i dati siano già stati raccolti e il documento di consenso informato non dica nulla a proposito di condivisione o, peggio ancora, prometta di non condividere i dati con altri ricercatori. Cosa fare in questi casi? Ritengo che, anche in questi casi, non ci si possa o debba esimere dalla condivisione ai fini del controllo delle analisi effettuate da parte della comunità scientifica, con adeguate misure per tutelare l'interesse alla privacy dei partecipanti. È importante ricordare che la disponibilità alla condivisione è esplicitamente prevista nel Codice Etico per la ricerca in Psicologia (articolo 9.9): "Chi svolge attività di ricerca in psicologia deve essere disponibile, per un periodo di almeno tre anni dalla pubblicazione dei risultati, a mostrare i dati grezzi, ancorché anonimi, a chi ne faccia richiesta per ulteriori verifiche" (AIP, 2015)³. E' inoltre plausibile che i partecipanti valuterebbero positivamente le misure prese per assicurarsi che i risultati della ricerca a cui hanno partecipato siano corretti e perciò sarebbero favorevoli a questo tipo di condivisione dei dati.

Per quanto concerne la condivisione dei dati per analisi secondarie, se in cui il foglio di consenso informato è silente sulla possibilità o meno di condivisione e riutilizzo dei dati, è importante trovare un giusto bilanciamento tra gli interessi dei partecipanti, che hanno il diritto alla tutela della privacy e hanno acconsentito a partecipare a uno studio che aveva determinate finalità di ricerca, con l'interesse scientifico che i dati vengano usati nel miglior modo possibile. Perciò credo che la situazione vada valutata di volta in volta, considerando quali analisi verrebbero effettuate: per esempio, la condivisione dei dati con altri

³ Analoghi articoli sono presenti nei codici etici di altre associazioni nazionali di psicologia, come quello dell'American Psychological Association e della Deutsche Gesellschaft für Psychologie

ricercatori per un'analisi integrata di più studi con scopi simili a quelli dello studio in oggetto non contravverrebbe all'accordo preso con il partecipante circa l'oggetto della ricerca.

Se il ricercatore, all'atto del consenso informato, si è esplicitamente impegnato a non condividere i dati con altri, la situazione è ancora più problematica. Da un lato, si potrebbe sostenere che probabilmente i partecipanti hanno accettato queste condizioni per il semplice fatto che non ne sono state proposte di diverse, ma avrebbero probabilmente accettato anche altre opzioni meno restrittive. D'altro lato, l'impegno preso è un vincolo esplicito e un'eventuale decisione di contravvenire a questo vincolo deve essere il frutto di un'attenta valutazione, che dovrebbe probabilmente essere accompagnata dal parere del Comitato Etico.

Conclusione

Gli studi qui descritti evidenziano che gli psicologi hanno una lunga strada da percorrere, sul cammino della condivisione dei dati. Molti sono riluttanti non solo a caricare i loro dati in archivi pubblici, ma persino a condividerli con altri ricercatori che ne facciano richiesta. Questi comportamenti pongono problemi etici e di trasparenza della ricerca e sono un serio limite alle possibilità di progresso scientifico. I cambiamenti tecnologici a cui abbiamo assistito negli ultimi anni e la disponibilità di archivi, in molti casi gratuiti e facilmente fruibili, hanno eliminato alcune delle difficoltà legate alla condivisione. Rimangono barriere culturali, paure che spero possano mitigarsi con una migliore conoscenza degli strumenti a disposizione, difficoltà oggettive dovute alle risorse – soprattutto di tempo – che devono essere inizialmente investite per avviare le prassi di condivisione all'interno del proprio laboratorio. Si tratta tuttavia di prassi necessarie, che vanno a beneficio della scienza nel suo insieme ma anche del singolo ricercatore.

Bibliografia

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, 16(2), 270-301. DOI:

10.1177/1094428112470848

Agnoli, F., & Carollo, A.G. (2018). Uso e (abuso) di prassi di ricerca problematiche in psicologia.

- Associazione Italiana di Psicologia (2015). Codice etico (Approvato nell'Assemblea Generale dei Soci AIP del 27 marzo 2015). Recuperato il 7 ottobre 2018 <https://www.aipass.org/node/11560>
- Ceci, S. J. (1988). Scientists' attitudes toward data sharing. *Science, Technology, & Human Values*, 13(1-2), 45-52. DOI: 10.1177/0162243988013001-206
- Ceci, S. J., & Walker, E. (1983). Private archives and public needs. *American Psychologist*, 38, 414. DOI: [10.1037/0003-066X.38.4.414](https://doi.org/10.1037/0003-066X.38.4.414)
- Craig, J. R. , & Reese, S. C. (1973). Retention of raw data: A problem revisited. *American Psychologist*, 28, 723. Wolins, L. (1962). Responsibility for raw data. *American Psychologist*, 17, 657–658. DOI: [10.1037/h0035667](https://doi.org/10.1037/h0035667)
- Creative Commons. (2016). *Frequently asked questions*. Recuperato il 7 ottobre 2018 da <https://creativecommons.org/faq/>
- Crepaldi, D. (2018). Open Science, Fair Science: Garantire la trasparenza della scienza attraverso l'organizzazione della pratica quotidiana in laboratorio
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., ... & Brown, E. R. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68-82. DOI: [10.1016/j.jesp.2015.10.012](https://doi.org/10.1016/j.jesp.2015.10.012)
- Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E., & Martone, M. E. (2014). Big data from small data: data-sharing in the 'long tail' of neuroscience. *Nature neuroscience*, 17(11), 1442-1447; DOI: [10.1038/nn.3838](https://doi.org/10.1038/nn.3838)
- Grassi, M. (2018). Crisi della riproducibilità e Open Science: alle porte di un cambio di paradigma
- Hardwicke, T. E., & Ioannidis, J. P. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLoS one*, 13(8), e0201856. DOI: [10.1371/journal.pone.0201856](https://doi.org/10.1371/journal.pone.0201856)
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data Sharing in Psychology: A Survey on Barriers and Preconditions. *Advances in Methods and Practices in Psychological Science*, 1(1), 70-85. DOI: [10.1177/2515245917751886](https://doi.org/10.1177/2515245917751886)

- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., ... Nosek, B. A. (2014). Investigating variation in replicability: A “many labs” replication project. *Social Psychology*, 45, 142–152. DOI: [10.1027/1864-9335/a000178](https://doi.org/10.1027/1864-9335/a000178).
- Longo, D. L., & Drazen, J. M. (2016). Data Sharing. *The New England Journal of Medicine*, 374, 276 –277. DOI: [10.1056/NEJMe1516564](https://doi.org/10.1056/NEJMe1516564)
- Lundberg, I., Narayanan, A., Levy, K., & Salganik, M. J. (2018). Privacy, ethics, and data access: A case study of the Fragile Families Challenge Recuperato il 1 settembre 2018. *arXiv preprint arXiv:1809.00103*.
- Martone, M. E., Garcia-Castro, A., & VandenBos, G. R. (2018). Data sharing in psychology. *American Psychologist*, 73(2), 111. DOI: [10.1037/amp0000242](https://doi.org/10.1037/amp0000242)
- McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Spies, J. R. (2016). Point of view: How open science helps researchers succeed. *Elife*, 5, e16800. DOI: [10.7554/eLife.16800](https://doi.org/10.7554/eLife.16800)
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1, 131–144. DOI: [10.1177/2515245917747656](https://doi.org/10.1177/2515245917747656)
- Monteith, S., Glenn, T., Geddes, J., & Bauer, M. (2015). Big data are coming to psychiatry: a general introduction. *International journal of bipolar disorders*, 3, 21. doi: [10.1186/s40345-015-0038-9](https://doi.org/10.1186/s40345-015-0038-9)
- NIMH Collaborative Data Synthesis for Adolescent Depression_Trials Study Team. (2013). Advancing science through collaboratedata sharing and synthesis. *Perspectives on Psychological Science*, 8, 433–444. <http://dx.doi.org/10.1177/1745691613491579>
- Pashler, H., & Wagenmakers, E. J. (2012). Editors’ introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, 7(6), 528-530. DOI: [10.1177/1745691612465253](https://doi.org/10.1177/1745691612465253)
- Perugini, M. (2014). La crisi internazionale di credibilità della psicologia come un'opportunità di crescita: Problemi e possibili soluzioni. *Giornale italiano di psicologia*, 41(1), 23-46. DOI: [10.1421/77194](https://doi.org/10.1421/77194)
- Perugini, M. (2018). Separare il segnale dal rumore: Alcuni suggerimenti.
- Schönbrodt, F., Gollwitzer, M., & Abele-Brehm, A. (2016). *Data management in psychological science: Specification of the DFG guidelines*. Advance online publication. Recuperato il 7 ottobre 2018

da http://www.dgps.de/fileadmin/documents/Empfehlungen/Data_Management_in_Psychological_Science_20160928.pdf

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PloS one*, *10*(8), e0134826. DOI: 10.1371/journal.pone.0134826

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra: Psychology*, *1*, Article 3. doi:10.1525/collabra.13

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*(7), 726-728. <http://dx.doi.org/10.1037/0003-066X.61.7.726>

Wolins, L. (1962). Responsibility for raw data. *American Psychologist*, *17*, 657–658. <http://dx.doi.org/10.1037/h0038819>