

L'importanza della cooperazione e la condivisione dei dati e nella ricerca scientifica

Giovanni Destro Bisol^{1,2*} Paolo Anagnostou^{1,2} e Marco Capocasa²

¹ Università di Roma "La Sapienza", Dipartimento di Biologia Ambientale, Roma Italy

² Istituto Italiano di Antropologia, Roma, Italy

email: destrobisol@uniroma1.it

L'avvento di Internet e delle tecnologie associate hanno fatto letteralmente esplodere la quantità di informazione liberamente disponibile, una vera e propria rivoluzione che ha avuto, e sta avendo, conseguenze profonde sui più diversi aspetti della nostra vita quotidiana (Hendriks, 1999; Morrison et al, 2001). Questa nuova era della comunicazione ha inevitabilmente attivato dei processi di cambiamento anche per le attività scientifiche. Grazie agli sforzi congiunti di ricercatori, comunicatori della scienza, responsabili di istituzioni e di riviste è in atto in molti ambiti di ricerca un ampio movimento per rendere più accessibili non solo i contenuti scientifici, ma anche dati e metodi che ne sono alla base (Neylon e Wu, 2008; Boulton et al. 2012). Al tempo stesso, il principio dell'*open access* è stato recepito anche da grandi iniziative scientifiche, prima fra tutte Horizon 2020, il nuovo programma di finanziamento per le attività di ricerca della Commissione europea per il prossimo settennio. Infine, si moltiplicano le iniziative di divulgazione degli "open data", mentre sembrano assumere un'importanza crescente modelli di *governance* basati sulla condivisione delle conoscenze finalizzate ad un maggiore coinvolgimento dei cittadini nelle scelte politiche e sociali (vedi il contributo di Andrea Cerroni in questo volume).

Dato questo insieme di azioni convergenti, potrebbe sembrare che la piena e responsabile condivisione dei saperi scientifici sia un processo ormai ben avviato, il quale necessita semplicemente di tempo e perseveranza per il suo completamento. Sfortunatamente le cose non stanno esattamente in questo modo. Sono molti gli ostacoli ancora da superare e le resistenze da vincere a vari livelli per il pieno sviluppo di una scienza aperta. L'argomento specifico della condivisione dei dati, o "data sharing", dimostra, forse meglio di ogni altro, quanto lunga sia ancora la strada da percorrere, e quanto urgente e vitale sia per la ricerca scientifica intensificare gli sforzi nella direzione dell'apertura, della condivisione e della trasparenza.

Partendo da questa premessa, vogliamo, in primo luogo, chiarire i principali punti a favore e contro la condivisione dei dati della ricerca. A questo seguirà la trattazione di tre aspetti che, nonostante la loro importanza, sembrano non avere ancora uno spazio sufficiente nel dibattito attuale: (1) Il ruolo della cooperazione nei processi di produzione della scienza; (ii) l'importanza dell'inclusività, intesa come coinvolgimento delle diverse comunità scientifiche nel movimento degli "Open Data"; (iii) passando a un piano più operativo, l'utilità degli studi empirici delle modalità di condivisione (o trattenimento) dei dati nelle diverse pratiche scientifiche al fine sia di integrare le classiche strategie "top down" per il *data sharing*, che di sviluppare modelli più inclusivi di Open Science. Infine, descriviamo brevemente "Opening Science to Society", un'iniziativa recentemente proposta dal nostro gruppo di ricerca, con l'ambizione di contribuire alla discussione sulla condivisione dei dati nei suoi aspetti scientifici, didattici e sociali.

Data sharing: quali sono i pro e i contro

Come documentato da editoriali pubblicati recentemente in alcune delle più prestigiose riviste scientifiche (e da quanto discusso in alcuni dei contributi a questo volume), sembra essere davvero ampio il consenso

circa l'importanza della condivisione dei dati nella comunità scientifica (Fischer e Zigmond, 2010; Lang 2011; King 2011). Effettivamente, è possibile identificare almeno 3 aspetti che motivano in maniera convincente il ruolo dell'apertura dei dati per il progresso delle conoscenze (vedi anche gli esempi descritti nel BOX 1). In primo luogo, la condivisione amplia gli scopi e le prospettive delle ricerche scientifiche. Lavorare insieme, mettendo in comune risultati e capacità, è un fattore fondamentale per acquisire nuovi saperi. Anche disporre semplicemente di insiemi di dati, o *dataset*, che riuniscono i risultati di diversi laboratori, o poter rianalizzare, seguendo prospettive ancora inesplorate, i dati prodotti da altri può sicuramente aumentare le ricadute delle iniziative di ricerca. Secondo, la condivisione permette di utilizzare più efficacemente le risorse a disposizione dei ricercatori. Esse sono non solo finite, ma spesso inadeguate, a maggior ragione in periodi di crisi e in paesi dove la ricerca non sembra essere tra le principali priorità. Ecco allora che diventa fondamentale evitare duplicazioni nella produzione di dati e dei relativi esperimenti, il che permette anche di ridurre i disagi a carico dei donatori e degli animali da laboratorio. Infine, la condivisione migliora i rapporti e aumenta la fiducia tra i ricercatori. Sono ormai numerosi gli esempi di come la verifica e la ripetizione di analisi di dati scientifici possa permettere di scoprire frodi, distorsioni o, semplicemente, errori sperimentali. Più in generale, la consapevolezza che i propri risultati saranno resi pubblici scoraggia i ricercatori dal portare avanti pratiche scientifiche non rigorose. Tutto questo aumenta la fiducia e la disponibilità tra ricercatori, creando un clima di trasparenza e facilitando l'apertura.

Tuttavia, nonostante questi innegabili vantaggi, è evidente ad uno sguardo attento come il tema della condivisione dei dati non possa essere affrontato in maniera aprioristica, insistendo semplicemente su assiomi di ordine etico o epistemologico. Per comprendere questo punto di vista è necessario prendere consapevolezza almeno di alcuni degli aspetti che sono alla base della scarsa o insufficiente disponibilità a mettere in comune i risultati scientifici da parte di molti ricercatori. Tra questi vanno considerati: (i) il tempo e i costi economici necessari per le attività di condivisione, un problema che può avere un peso maggiore per piccoli gruppi e/o campi di ricerca in cui i finanziamenti sono maggiormente limitati; (ii) ragioni legate alla competizione per i fondi e la carriera, così come le opportunità legate alla brevettazione nei settori che hanno maggiori possibilità di commercializzazione delle scoperte; (iii) timori legati ad un uso dei dati non adeguato, o addirittura foriero di rischi per la collettività (per esempio bioterrorismo); (iv) in particolare negli studi biomedici, i rischi per i donatori di essere identificati, in modo tale che i risultati dello studio possano essere utilizzati per rendere loro difficile o più oneroso il ricorso a forme di previdenza sanitaria e di cura.

Per quanto sbrigativa, questa trattazione dei pro e contro speriamo sia sufficiente a illustrare la complessità dell'argomento e la necessità da parte della comunità scientifica di definire strategie efficaci. Un importante passaggio intermedio è, a nostro parere, rappresentato dalla definizione di un modello di produzione dei saperi scientifici che permetta di inserire la condivisione dei dati in una visione coerente e identificarne le basi epistemologiche.

La cooperazione nelle pratiche scientifiche, da Cartesio a oggi

Si attribuisce comunemente a Cartesio il primo esplicito riconoscimento dell'importanza della cooperazione nelle attività scientifiche:

Le migliori menti sarebbero indotte a contribuire ad ulteriori progressi, ciascuno secondo le sue inclinazioni e capacità, allestendo gli esperimenti necessari e comunicando al pubblico tutto ciò che hanno imparato, in modo che un uomo possa iniziare dove un altro si è fermato; così, combinando il lavoro di molti nel corso del tempo, molti più progressi sarebbero fatti da tutti insieme rispetto a quanto chiunque potrebbe fare da solo.

tradotto da: *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences, Sixième Partie* (René Descartes, 1637)

Partire da questa definizione, oltre a essere storicamente doveroso, è utile non solo a illustrare quanto i processi di cooperazione siano stati ben presto percepiti come elemento fondamentale di ogni impresa scientifica, ma anche quanto essi si siano modificati, nel senso di una maggiore articolazione, nel corso degli ultimi quattro secoli. Speriamo che i lettori vogliano perdonare la riduzione implicita nella nostra scelta di confrontare gli antipodi di un processo così lungo e complesso come quello dell'evoluzione del pensiero e del metodo scientifico.

Come primo punto, è interessante notare quanto il concetto di attori o "player" della ricerca si sia ampliato rispetto alla definizione di "migliori menti" di Cartesio, fino ad integrare la classica figura dei ricercatori con i cosiddetti *stakeholder*. Responsabili di istituzioni accademiche (università e centri di ricerca) ed enti finanziatori, comunicatori della scienza o anche semplici cittadini (si pensi alla *citizen science*) fanno parte di questa, per certi versi eterogenea, categoria, in cui i ruoli non sono sempre così rigidamente separati come potrebbe sembrare. Ma è ancor più importante osservare come il contributo e addirittura le modalità assunte dalle forme di cooperazione si siano evolute rispetto all'idea di un passaggio del testimone da un ricercatore ad un altro, così come traspare dalle parole di Cartesio. In definitiva, il processo di produzione scientifica allo stato attuale può essere definito come un sistema cooperativo, in cui entrano in gioco tre fattori principali: attori, obiettivi e strumenti. Mentre i primi sono stati appena definiti, nella seconda categoria, anche se in modo un po' schematico, possiamo identificare, almeno da un punto di vista nominalistico, la conoscenza di base e le applicazioni. Sempre in modo semplificato, teorie e pratiche scientifiche contribuiscono a definire la categoria degli strumenti. Non sfuggerà ai lettori la provvisorietà dei confini appena delineati e, in definitiva, l'osmosi necessaria sia tra che all'interno di ciascuno delle tre componenti del sistema. Per esempio, molte ricerche nate per alimentare conoscenze di base sono approdate a risultati che si sono tradotti, immediatamente o in un secondo tempo, in applicazioni scientifiche. Così come, l'evoluzione su base tecnologica delle pratiche scientifiche ha talvolta finito con il mettere in evidenza l'inadeguatezza delle teorie scientifiche, contribuendo a una loro rivisitazione o addirittura ad una loro profonda revisione. Pur riconoscendo la semplificazione implicita nella schematizzazione proposta, è comunque evidente che le varie componenti e sub-componenti che partecipano al processo di produzione scientifica necessitano di una stretta sinergia per un funzionamento ottimale del sistema nella sua totalità.

Crediamo si possano identificare almeno due elementi essenziali nei processi cooperativi necessari per realizzare tale sinergia. In primo luogo, un flusso illimitato di informazioni, che siano il prodotto di pratiche scientifiche non solo rigorose ma anche trasparenti. Non meno importante, una capacità inclusiva che si manifesti con: (i) un'ampia partecipazione e interazione tra le varie comunità scientifiche; (ii) il trasferimento dei concetti tra ricercatori giovani e maturi; (iii) la comunicazione delle evidenze e dello stesso metodo scientifico al pubblico, fino a un coinvolgimento attivo di quest'ultimo in una prospettiva di *governance* della gestione delle iniziative scientifiche e delle loro ricadute. Una volta definito un modello di questo tipo, quegli stessi elementi dei processi cooperativi che abbiamo appena citato assumono il ruolo di fattori indispensabili (anche se non sufficienti, si consideri ad esempio l'importanza degli investimenti) per favorire il progresso scientifico. Ovviamente, i dati, ed i relativi metadati, sono parte essenziale dell'informazione, così come, la capacità di condividerli è fondamentale per la capacità inclusiva cui si faceva riferimento.

Si noti che nell'ottica proposta la condivisione e la cooperazione caratterizzano il processo scientifico, il che finisce con il coincidere con quanto teorizzato su base etica da Robert King Merton nel suo saggio del 1942 "*The Normative Structure of Science*" tramite i concetti di "communalism" e "organized skepticism".

Integrare strategie *top down* e *bottom up* per una maggiore e più efficace condivisione dei dati

Cerchiamo ora di comprendere cosa accade nella realtà scientifica per quanto riguarda la condivisione dei dati, partendo dalla domanda: "Le attuali strategie per aumentare la condivisione dei dati sono coerenti con il modello proposto, ovvero sono realmente ispirate da principi di cooperazione?". Gli approcci correnti sembrano insistere principalmente su tre linee d'azione: creare e migliorare gli strumenti (e.g. infrastrutture e standard), definire norme (*policy* e linee guida) e motivare i ricercatori (*moral suasion*, incentivi) (cfr Piwowar, 2008). Tutte queste sono però più spesso il prodotto di valutazioni e decisioni prese da *elite*, secondo un approccio *top-down*, piuttosto che nascere da un coinvolgimento della comunità scientifica. Questo ne limita la capacità inclusiva, sia perché il disegno stesso delle strategie può partire da un orizzonte forzatamente limitato, sia perché tali approcci nascono generalmente per rispondere alle esigenze di specifici progetti, programmi di funzionamento e tematiche scientifiche.

Esiste tuttavia un'alternativa, rappresentata dagli studi empirici delle pratiche di condivisione dei dati, sia attraverso questionari o analisi della letteratura scientifica. Queste due vie possono essere a loro volta complementari: i questionari possono permettere di esplorare le ragioni alla base dei comportamenti, mentre lo scrutinio dei lavori scientifici può aiutare a definire un quadro fattuale di quanto e come si condividano i dati. Anche se non possono sostituire le attuali strategie (si pensi solamente all'importanza delle infrastrutture), tali approcci possono fornire risposte quantitative a domande concernenti l'efficacia delle norme (sono le politiche realmente efficaci?), le motivazioni alla base dei comportamenti di condivisione (perché i dati sono condivisi o trattenuti?) e l'adeguatezza degli strumenti (quanto vengono utilizzati database e *standard* per condividere i dati scientifici). Ma nella nostra ottica, è di particolare interesse anche la possibilità che essi offrono di includere un maggior numero di ricercatori e comunità scientifiche nel movimento a favore degli *open data*. Infatti, gli studi empirici possono essere intrapresi da qualsiasi comunità scientifica o parte di essa, secondo un approccio *bottom-up*, ed essere disegnati in modo da adattarsi alle pratiche di condivisione dei più diversi campi di ricerca ed applicabili ai contesti istituzionali tra loro indipendenti. In una prospettiva ampia, una raccolta sistematica di informazioni relative alle principali aree scientifiche può essere molto utile per identificare le diverse necessità, ma anche i rispettivi punti di forza e di debolezza per lo sviluppo di modelli avanzati di condivisione, definire strategie e linee guida comuni e sviluppare modelli che possano facilitare le interazioni disciplinari. Un'esigenza, quest'ultima, molto attuale, come dimostrato ad esempio dagli sviluppi della ricerca in neurobiologia o nelle necessità di integrare le varie "omiche" (e.g. genomica, trascrittomica e proteomica).

Un caso studio: il *data sharing* nella ricerca genetica e antropologica

Il campo di ricerca della variabilità genetica umana può essere considerato come un settore ideale per il raggiungimento di una piena condivisione dei dati. Ciò è dovuto alla natura codificata dell'informazione genetica, alla disponibilità di strumenti per la diffusione dei dati e, ovviamente, all'importanza che assume la ricerca in questo campo dal punto di vista della prevenzione, della diagnosi e dello sviluppo di terapie per innumerevoli patologie (Nelson et al. 2008; Peltonen 2011).

Ma la condivisione dei dati è realmente una prassi comune nelle pubblicazioni prodotte in questo settore di ricerca? Per rispondere a questa domanda abbiamo deciso di analizzare il grado di condivisione dei dati genetici nell'ambito degli studi di variabilità genetica umana. Il nostro studio, pubblicato nel giugno del 2012 sulla rivista PLoS One (Milia et al. 2012), è stato basato sull'analisi della condivisione di dati in 543 studi riguardanti la variabilità genetica nelle popolazioni umane pubblicati dal 2008 al 2011. Abbiamo concentrato la nostra attenzione sui polimorfismi del DNA mitocondriale e del cromosoma Y, marcatori diffusamente utilizzati in diversi campi di ricerca della genetica umana, come l'antropologia molecolare, la genetica forense e la genetica medica (King e Jobling 2009; Destro Bisol et al. 2010; Mishmar e Zhidkov 2010).

L'analisi delle pubblicazioni ha innanzitutto messo in luce come una parte sostanziale dei *dataset* (23,2%) non sia immediatamente condivisa. Inoltre, né l'applicazione di politiche editoriali esplicite né un elevato *impact factor* si sono dimostrati efficaci nell'aumentare il tasso di condivisione oltre l'80,5%. Disaggregando i dati in base al campo di ricerca, abbiamo potuto osservare una minore propensione alla condivisione in genetica medica rispetto alla genetica evolutiva e alla genetica forense (vedi figura 1 per ulteriori dettagli).

Il nostro studio suggerisce quindi che anche i ricercatori operanti in diversi sottosectori della genetica umana incontrano spesso delle difficoltà nell'attuare una piena condivisione dei loro dati. Questo può dipendere dal fatto che si tratta di una tipologia di dati sensibili, per il cui trattamento è necessario tenere in considerazione una serie di aspetti etici e legali. Il trattamento dei dati del DNA può, innanzitutto, mettere a rischio la *privacy*. Tale problema può verificarsi nel caso in cui i dati ottenuti dall'analisi genetica siano associati a metadati individuali (ad esempio il luogo di nascita, l'età, il sesso o le eventuali relazioni di parentela con altri donatori). Informazioni di questo tipo possono, tuttavia, risultare estremamente utili per approfondire alcuni aspetti particolari. Ad esempio, negli studi microevolutivi sulla variabilità genetica umana, possono aiutare a valutare gli effetti di particolari comportamenti sociali sulla struttura genetica delle popolazioni umane (Destro Bisol et al. 2012). Pertanto, il loro uso deve però essere attentamente calibrato in modo da minimizzare il rischio di violare la riservatezza dei donatori. Tale raccomandazione si applica a maggior ragione agli studi basati di genomica umana (Homer et al. 2008).

Come logico sviluppo dello studio, dopo l'analisi dei dati, abbiamo cercato di mettere a fuoco quali fossero le implicazioni dei nostri risultati in grado di contribuire al perfezionamento delle strategie di condivisione dei dati. Tre aspetti sembrano, a nostro avviso, i più importanti. In primo luogo, le minori percentuali di condivisione osservate per i lavori nel campo della genetica medica suggeriscono che i fattori che limitano la condivisione dei dati possono variare anche tra settori di ricerca strettamente affini. Questo mostra come l'identificazione di accorgimenti specifici per ogni settore di ricerca sia di fondamentale importanza per aumentare realmente la condivisione dei dati. La differenza osservata potrebbe scaturire anche dai potenziali conflitti tra condivisione dei dati e *privacy* e/o dalla mancanza di consapevolezza dei ricercatori in ambito bio-medico sull'utilità dei dati per altri campi di ricerca (in particolare dati relativi ai gruppi di controllo). Altre possibili spiegazioni sono state discusse in un recente studio effettuato sui dati di espressione genica su *microarray*, in cui è stato dimostrato che gli autori di lavori incentrati sullo studio dei tumori in soggetti umani fossero meno propensi a rendere i loro dati disponibili (Piwowar, 2011). L'autore di questo articolo suggerisce che la percezione da parte dei ricercatori che l'ambiente accademico legato a questo tipo di studi sia molto competitivo, l'influenza esercitata dall'industria farmaceutica e i problemi legati alla *privacy* rendono i ricercatori meno propensi a condividere i loro dati. Le prime due condizioni sono probabilmente più presenti nel campo medico rispetto a quello forense o della genetica evolutiva.

La seconda considerazione parte dalla constatazione che, nonostante la percentuale di risposte positive da parte dei *corresponding authors* alle nostre richieste e-mail per ottenere i *dataset* "trattenuti" sia stata superiore a quella ottenuta in precedenti studi (Savage e, Vickers, 2009; Wicherts et al. 2006), anche nel nostro caso un gran parte dei dati richiesti (90 su 126, 71,4%) resta non disponibile. Le difficoltà nel recupero dei dati dopo la loro pubblicazione implica che una completa ed efficace condivisione dovrebbe essere vista, attraverso politiche editoriali più restrittive, come requisito fondamentale per l'accettazione e la pubblicazione del lavoro, piuttosto che una semplice raccomandazione. Tra le numerose politiche editoriali esaminate, quelle del *International Journal of Legal Medicine* (IJLM) e *Forensic Science International Genetics* (FSIG) possono essere prese come un utile modello (Carracedo et al. 2010, Parsons e Rower 2010, Carracedo et al., 2013). Gli autori per sottomettere i propri lavori in queste due riviste devono prima depositare i propri dati a Y Chromosome Haplotype Reference Database (www.yhrd.org) ed al European mtDNA Population Database (<http://empop.org/>) (Parson et al. 2004; Willuweit et al. 2007).

Dopo che i dati hanno superato il controllo di qualità, i lavori sono sottoposti a *peer review*. In caso di accettazione finale, i dati devono essere riportati come aplotipi individuali, di solito come materiale supplementare online. Abbiamo osservato che la percentuale di condivisione dei dati pubblicati in IJLM e FSIG contribuiscono all'elevato livello di condivisione riscontrato in ambito forense. Pertanto, le politiche editoriali di queste due riviste possono avere un impatto sostanziale sulla disponibilità di dati forensi di alta qualità. È altresì importante notare che le riviste possono beneficiare dall'adozione di regole sulla condivisione dei dati in quanto i lavori i cui *dataset* sono disponibili senza restrizioni hanno una maggiore probabilità di essere citati rispetto a quelli i cui dati non sono condivisi (Piwowar et al. 2007). Naturalmente, questo può contribuire ad aumentare il loro *impact factor*; IJLM e FSIG sono infatti le riviste con *impact factor* più alto nella loro categoria "*Medicine, Legal*" del Science Citation Index (versione 2010).

Infine, attraverso il nostro studio, abbiamo messo in luce un aspetto importante e concreto che potrebbe contribuire a diffondere una cultura della condivisione dei dati più efficace tra i giovani ricercatori e in tutta la comunità scientifica. Infatti, abbiamo stimato che tra il 29,8% e il 32,9% delle risorse impiegate in tutte le ricerche considerate nello studio sono state spese per produrre dati non condivisi. Un dato davvero notevole, che indica quanto sia importante rendere la comunità scientifica e gli *stakeholder* consapevoli della necessità di migliorare il rapporto tra benefici (le informazioni disponibili alla comunità scientifica) e costi (risorse impiegate) delle attività di ricerca.

In conclusione, il nostro studio dimostra che la maggioranza dei dati pubblicati nell'ambito della variabilità genetica umana sono a disposizione della comunità scientifica. Tuttavia, abbiamo anche mostrato che ulteriori sforzi sono ancora necessari per rendere la condivisione dei dati una pratica comune in questo settore di ricerca. Noi sosteniamo che la ricerca in questo campo possa realmente diventare un *forerunner* per la creazione di una condivisione dei dati diffusa, rendendo le politiche editoriali più rigorose, adattando le strategie alle caratteristiche di ogni specifico ambito di ricerca e divulgando i vantaggi della condivisione dei dati in termini di utilizzo ottimale delle risorse. Più in generale, ci auguriamo che questo studio possa aprire la strada ad ulteriori indagini in altre aree di ricerca.

Il progetto "Opening Science to Society"

La priorità assunta dal tema della condivisione per la ricerca biologica si riflette in un moltiplicarsi di pubblicazioni, progetti e risorse online che hanno come motivazione principale la relazione tra quantità di dati disponibili e progresso scientifico (e.g. www.dataone.org, www.biosharing.org). Cominciano, tuttavia, ad affacciarsi anche altri punti di interesse. La comunità scientifica si sta rendendo finalmente consapevole della necessità di diffondere la cultura della condivisione tra i giovani ricercatori e di attrarre su questo tema anche l'attenzione del grande pubblico (Destro Bisol et al. 2014).

Allo scopo di andare incontro a questa visione più ampia del *data sharing* stiamo lanciando il progetto di ricerca "Opening Science to Society" (Destro Bisol et al. 2013). Con questa iniziativa vogliamo affrontare non solamente gli aspetti scientifici, ma anche quelli formativi e divulgativi del *data sharing* con un approccio integrato. Come primo passo vogliamo chiarire quanto e come i dati scientifici vengano condivisi in diversi settori di ricerca, estendendo lo studio già condotto nell'ambito della genetica umana (Milia et al. 2012). Il potenziale del *data sharing* in ambito formativo verrà sviluppato attraverso la messa a punto di strumenti educativi che avranno come punto di partenza la valutazione delle conoscenze e della percezione degli studenti universitari rispetto al tema della condivisione dei dati e delle sue implicazioni. Infine, il nostro progetto vuole mettere a punto modelli di partecipazione attiva e consapevole adatti ad avvicinare la cittadinanza a progetti di ricerca scientifica. Coerentemente con quanto è stato discusso prima, tutti e tre questi obiettivi mettono al centro la cooperazione e, in particolar modo, la capacità di includere tutti i protagonisti dei processi di produzione e comunicazione scientifica.

Siamo consapevoli del fatto che per raggiungere questi obiettivi dobbiamo essere capaci di aprire uno

spazio di ulteriore “condivisione” di competenze ed esperienze, di idee e di opinioni. Questa è la prima sfida che il nostro progetto deve affrontare: sviluppare una collaborazione fra esperti e studiosi di discipline diverse, accomunati dall’esigenza di sviluppare una scienza aperta alla società.

Vi invitiamo a visitare il sito <http://isita-org.com/openingsciencetosociety> per saperne di più e dare il vostro *feedback*.

RINGRAZIAMENTI

Questo lavoro è stato possibile grazie al supporto economico dell’Istituto Italiano di Antropologia (progetto “La condivisione dei dati scientifici nell’era digitale: un’indagine sulla variabilità genetica nelle popolazioni umane”).

BIBLIOGRAFIA

- Boulton G., Campbell P., Collins B., Elias P., Hall W., Laurie G., O'Neill O., Rawlins M., Thornton J., Vallance P. & Walport W. (2012) *Science as an open enterprise*. The Royal Society, London.
- Carracedo A, Butler JM, Gusmao L, Parson W, Roewer L, Schneider PM. (2010) Publication of population data for forensic purposes. *Forensic Sci Int Genet* 4: 145–147.
- Carracedo A, Butler JM, Gusmao L, Parson W, Roewer L, Schneider PM. (2013) New guidelines for the publication of genetic population data. *Forensic Sci Int Genet* 7: 217–220.
- Destro-Bisol G, Jobling MA, Rocha J, Novembre J, Richards MB, et al. (2010) Molecular anthropology in the genomic era. *J Anthropol Sci* 88: 93–112.
- Destro Bisol G, Capocasa M, Anagnostou P. (2012) When gender matters: new insights into the relationships between social systems and the genetic structure of human populations. *Mol Ecol* 21: 4917-4920.
- Destro Bisol G, Capocasa M, Anagnostou P, Greco P. (2013). Opening Science to Society, a new initiative of the Istituto Italiano di Antropologia . *J Anthropol Sci*, 91: 233-235.
- Destro Bisol G, Anagnostou P, Capocasa M, Bencivelli M, Cerroni A, Contreras J, Enke N, Fantini B, Greco P, Heeney C, Luzi D, Manghi P, Mascalzoni M, Molloy J, Parenti F, Wicherts JM and Boulton G. Bridging perspectives on Open Science: a report from the meeting “Scientific data sharing: an interdisciplinary workshop”. in preparazione.
- Fischer BA, Zigmund MJ. (2010) The essential nature of sharing in science. *SciEng Ethics* 16: 783–799. doi:10.1007/s11948-010-9239-x.
- Hendriks P. (1999) Why Share Knowledge? The Influence of ICT on the Motivation for Knowledge Sharing. *Knowledge and Process Management*, 6: 91–100.
- Homer N, Szelinger S, Redman M, Duggan D, Tembe W, Muehling J, Pearson JV, Stephan DA, Nelson SF, Craig DW (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*, 4(8): e1000167.
- King G. (2011) Ensuring the data-rich future of the social sciences. *Science*, 331: 719–721.
- King TE, Jobling MA. (2009) What’s in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet* 25: 351–360.
- Lang T. (2011) Advancing global health research through digital technology and sharing data. *Science*, 331: 714–717.
- Merton RK (1942) reprinted in *The sociology of science: theoretical and empirical investigations*. University of Chicago Press. 1972 pag 267–278.

- Milia N, Congiu A, Anagnostou P, Montinaro F, Capocasa M, Sanna E, Destro Bisol G. (2012) Mines, yours, ours? Sharing data on human genetic variation. *PLoS One*, 7: e37552.
- Mishmar D, Zhidkov I. (2010) Evolution and disease converge in the mitochondrion. *Biochim Biophys Acta* 1797: 1099–1104.
- Morrison JB, Pirolli P & Card SK (2001). A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. *Proceedings CHI 2001*, 163-164.
- Murdoch C, Caulfield T. (2009) Commercialization, patenting and genomics: researcher perspectives. *Genome Med* 1: 22.
- Nelson B. (2009) Data sharing: Empty archives. *Nature* 461: 160–163.
- Neylon C. & Wu S. (2008) Open science: tools, approaches, and implications. *Nature Precedings*, doi:10.1038/npre.2008.1633.1.
- Parson W, Brandstatter A, Alonso A, Brandt N, Brinkmann B, et al. (2004) The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Sci Int* 139(2–3): 215–226.
- Parson W, Roewer L. (2010) Publication of population data of linearly inherited DNA markers in the *International Journal of Legal Medicine*. *Int J Legal Med* 124: 505–509
- Peltonen L. (2001) The molecular dissection of human diseases after the human genome project. *Pharmacogenomics J* 1: 5–6.
- Piwovar HA, Becich MJ, Bilofsky H, Crowley RS. (2008) Towards a data sharing culture: recommendations for leadership from academic health centers. *PLoS Med* 5: e183.
- Piwovar HA, Day RS, Fridsma DB. (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2: e308.
- Piwovar HA, Day RS, Fridsma DB. (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2: e308.
- Savage CJ, Vickers AJ. (2009) Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE* 4: e7078.
- Wicherts JM, Borsboom D, Kats J, Molenaar D. (2006) The poor availability of psychological research data for reanalysis. *Am Psychol* 61: 726–728.
- Willuweit S, Roewer L; International Forensic Y Chromosome User Group. (2007) Y chromosome haplotype reference database (YHRD): update. *Forensic Sci Int Genet.* 1: 83–87.

BOX 1. PERCHE' E' IMPORTANTE CONDIVIDERE I DATI

Il neutrino: più veloce della luce... o invece no!

Nel settembre del 2011 OPERA, un progetto di collaborazione internazionale con sede al CERN, ha diffuso la notizia che i neutrini superano la velocità della luce, una scoperta capace di rovesciare i fondamenti stessi della Fisica così come la conosciamo. I ricercatori, "sparando" un fascio di neutrini muonici dall'acceleratore di particelle del CERN al laboratorio del Gran Sasso, posto ad una distanza di 731Km, hanno registrato che esso arriva con un anticipo di 60 nanosecondi rispetto a quanto ci si aspetterebbe se viaggiasse alla velocità della luce. Questa scoperta ha suscitato scalpore e scetticismo nella comunità scientifica, tanto da spingere il CERN a rendere accessibili tutti i dati relativi all'esperimento attraverso l'archivio online arXiv.org. Da allora, nello stesso archivio, sono apparsi circa 200 lavori che cercavano di spiegare questa "anomalia", focalizzandosi, in particolare, sulla tecnica utilizzata per rilevare il tempo di percorrenza del fascio di neutrini. Il 23 febbraio del 2012, prendendo atto di questi lavori, i ricercatori di OPERA hanno annunciato che i risultati dell'esperimento erano stati falsati da due problemi tecnici, uno nella fibra ottica e l'altro nel *master clock* del laboratorio di Gran Sasso. La replicazione dell'esperimento utilizzando quattro strumenti separati ha portato i ricercatori ad annunciare che il neutrino non supera la velocità della luce, confermando di fatto il supposto errore sperimentale.

Dalla teoria alla pratica: l'epidemia di E. coli del 2011

I benefici della condivisione dei dati sono emersi in maniera preponderante a seguito della diffusione di una severa infezione gastrointestinale, potenzialmente fatale, dovuta ad un ceppo particolarmente virulento di *Escherichia Coli*. Il primo caso clinico attribuibile a questo ceppo è stato registrato nel maggio del 2011 ad Amburgo in Germania, da dove l'epidemia si è propagata anche in altri stati infettando un totale di 3950 persone, delle quali 53 sono decedute. La mortalità particolarmente elevata di questo ceppo di E. coli è dovuto alla presenza di gene che produce una tossina (verocitotossina di sierogruppo O104) capace di scatenare la sindrome emolitico uremica (SEU) e la gastroenterite emorragica. Il ceppo in questione stato analizzato inizialmente da laboratori in Cina (BGI-Shenzhen) e Amburgo. Tre giorni dopo una prima versione del genoma del batterio è stata messa a disposizione della comunità scientifica internazionale con una licenza aperta. Ventiquattro ore dopo, grazie al lavoro di bioinformatici di tutto il mondo, il genoma era stato già analizzato ed assemblato e meno di una settimana più tardi, nel sito *open-source* dedicato all'analisi del ceppo di *E. Coli*, venivano pubblicati due dozzine di rapporti. Queste analisi tempestive hanno fornito cruciali informazioni sulla virulenza e la resistenza agli antibiotici del ceppo batterico e sulla sua diffusione, permettendo così di mettere a punto cure adeguate e fermare l'epidemia.

Il virus dell'influenza aviaria, un caso Italiano di buona pratica

L'influenza aviaria è una malattia infettiva altamente diffusiva dovuta al virus orthomyxovirus. Colpisce diverse specie di uccelli sia domestici che selvatici. Tuttavia ci sono alcune evidenze che mostrano una trasmissione del virus anche all'uomo con un altissimo tasso di mortalità (circa il 60% degli infettati). Nel 2003 un focolaio di influenza aviaria del sottotipo H5N1 viene registrato nel sud-est asiatico e si diffonde rapidamente divenendo una minaccia globale per la salute degli uomini e degli animali. La ricerca di una cura definitiva per questo virus è resa difficile dalla sua capacità di mutare e combinarsi con altri ceppi di virus influenzali. In questo contesto, per capire la sua evoluzione, risulta fondamentale poter comparare i genomi dei ceppi nuovi con quelli già disponibili. Tuttavia, i dati su questi ultimi non sono sempre disponibili perché i ricercatori sono riluttanti al loro rilascio prima che essi siano pubblicati in un articolo scientifico. Questo atteggiamento rallenta il progresso della conoscenza su questo virus con gravi conseguenze per la salute pubblica. Nel 2006 la dottoressa Ilaria Capua nel suo laboratorio dell'Istituto Zooprofilattico Sperimentale delle Venezie ha isolato il ceppo Africano del virus H5N1 e, piuttosto che depositare i dati genetici in un archivio protetto accessibile solamente ad altri 15 lavoratori, ha deciso di rilasciarli in un archivio pubblico aperto a tutti (GenBank) rinunciando così al primato della pubblicazione dei dati e privilegiando l'interesse della collettività.

Figura 1 - Percentuali di condivisione di dati in pubblicazioni riguardanti la variabilità nelle popolazioni umane di polimorfismi del DNA mitocondriale e del cromosoma Y. Le barre verticali indicano intervalli di confidenza al 95%. (A) Nella categoria "immediate sharing" (sharing immediato), abbiamo riportato la percentuale di *dataset* condivisi nel testo principale, nel materiale supplementare o banche dati on-line che sono stati esplicitamente indicati nel lavoro, le email "will provide on request" (fornirò i dati su richiesta) sono state inviate agli autori corrispondenti che esplicitamente dichiaravano nel lavoro la loro disponibilità all'invio dei dati; Le email "all authors" sono state inviate a tutti gli autori dei *dataset* "trattenuti". I risultati riportati nelle caselle B, C e D sono stati ottenuti utilizzando le percentuali di condivisione ottenute dopo le risposte positive alle e-mail "will provide on request". Abbiamo considerato come negative le risposte in cui gli autori hanno richiesto informazioni dettagliate circa l'uso dei dati e / o richieste di coauthorship prima dell'invio dei dati. (B) I *dataset* sono stati assegnati a ciascun settore di ricerca in relazione alle finalità di ricerca, come indicato nel lavoro. Quando l'assegnazione di un lavoro risultava essere non possibile per via obiettivi di ricerca ambigui o non esplicitati, è stato utilizzato come criterio aggiuntivo la categoria ISI della rivista scientifica. (C) Il tipo di politica editoriale è stata valutata utilizzando le informazioni riportate nelle linee guida per gli autori di ogni giornale: "politica editoriale debole" sono quelle in cui gli autori sono invitati a condividere i dati, mentre "politiche editoriali forti", quelle in cui la condivisione dei dati è indicata come obbligatoria. (D), La classificazione è stata fatta in base ai valori del fattore di impatto ISI rilasciate da Reuters nel giugno 2009. Si consulti Milia et al. (2012) per ulteriori dettagli.

