



Research data e scienze umane: un quadro europeo

Maura Funari

1 Definire i research data

Il tema dei research data¹ è trattato in quei contesti che si rivolgono all'ambito della ricerca scientifica, dalle teorizzazioni sui diversi aspetti della ricerca, alle politiche gestionali dei risultati della stessa e ai documenti di indirizzo delle organizzazioni sovranazionali. Una comune tendenza è quella di scavalcare la necessità di definire cosa intendere in senso stretto per research data. Nella pratica, però, non è difficile imbattersi in ambiguità concettuali: a seconda dell'ambito di applicazione vengono individuati aspetti propri che, quindi, variano, seppure in taluni casi quasi impercettibilmente. Una definizione, a carattere univoco ed universale, non è ad oggi esistente.

¹La traduzione italiana non ha trovato precisa definizione: nei documenti in versione italiana della Commissione Europea si parla sia di "dati della ricerca" (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*) che di "dati di ricerca" (*Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*). Si è scelto di mantenere per la presente trattazione il termine anglosassone perché già comunemente accettato.

L'UNESCO (Swan) qualifica i research data come un genere di "research output" insieme a riviste, atti di convegni peer-reviewed e libri. Tale categorizzazione sembra individuare in questa tipologia di dati uno strumento per l'esternalizzazione dei prodotti di una ricerca, riconoscendo, altresì, un incremento generale dell'attenzione attribuitagli nelle politiche sull'accesso aperto.

Un insieme di definizioni è dato dall'Australian National Data Service (ANDS) che, specificando in premessa quanto ogni definizione dipenda dal contesto che la richiede, raccoglie quelle enunciate nelle politiche di gestione dei dati di alcune università australiane: University of Melbourne, Monash University, Griffith University (Australian National Data Service). Le prime due definizioni richiamano elementi comuni nella caratterizzazione dei research data, riconoscendo varietà di forme e contenuti ed escludendo così una predeterminazione sulla base di tali criteri. Il carattere fondante dei research data, cioè il dato usato come fonte primaria, o su cui la teoria della ricerca stessa si basa, è un aspetto che si aggiunge soltanto nella prima delle tre definizioni citate: "Research Data means data [...] on which an argument, theory, test or hypothesis, or another research output is based" (University Of Melbourne). La definizione della Griffith University si rifà testualmente, nella prima parte, a quella della Organisation for Economic Cooperation and Development (OECD) (13) dove i research data sono definiti come "factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings". Tale definizione stabilisce, dunque, come caratterizzante per i research data, la condizione di materiale comunemente accettato come necessario, ai fini della validazione dei risultati di ricerca, da parte della comunità scientifica. Di simile impostazione è la definizione di scientific data riportata nel Memorandum for the Heads

of Executive Departments and Agencies² dell'Executive Office of the President, Office of Science and Technology policy, della White House. Tali dati sono definiti come: "the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications". Anche qui la condizione di materiale necessario alla validazione dei risultati della ricerca, il quale non è definito "factual record" ma "factual material (digital recorded)", è caratterizzante.

L'ANDS è, altresì, fondatore della RDA (Research Data Alliance),³ col supporto della Commissione Europea (tramite il progetto iCordi di cui si parlerà nel paragrafo dedicato alle infrastrutture) e degli Stati Uniti (tramite la National Science Foundation). L'organizzazione internazionale ha lo scopo di accelerare e migliorare l'innovazione e la ricerca data-driven, favorendo le azioni legate ai research data (quali scambio, condivisione, uso e ri-uso, standard e visibilità) e realizzando lo sviluppo e l'adozione di infrastrutture, policies, pratiche, standard e servizi.

Tornando alle definizioni di research data, la Comunicazione della Commissione Europea (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca* 3), nell'evidenziare la crescente attenzione rivolta al miglioramento dell'accesso ai dati della ricerca, qualifica questi ultimi come "risultati sperimentali, osservazioni e informazioni generate dal computer che costituiscono la base dell'analisi quantitativa sulla quale poggiano molte pubblicazioni scientifiche". Tenendo in considerazione l'eterogeneità che emerge dalle suddette definizioni è possibile, comunque,

²http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

³<http://rd-alliance.org>.

stabilire che per research data si possano intendere i dati, in diverse forme e contenuti, che sono alla base di una ricerca scientifica, come risorsa primaria e fondamento dei risultati della ricerca stessa. Poiché il valore intrinseco e le necessità di raccolta, conservazione, condivisione, sono qualità che variano sulla base di diversi fattori, dalla natura della ricerca all'ambito disciplinare, fondamentale per stabilire quali dati prodotti entro un certo settore rientrino nella categoria dei research data, è l'individuazione di criteri prestabiliti. Tale compito dovrebbe spettare prevalentemente alle politiche di trattamento dei risultati della ricerca, elaborate da centri, enti ed istituzioni coinvolti.

Da un punto di vista sia qualitativo che quantitativo, molti dati possono essere prodotti durante le fasi di una ricerca ma un aspetto che, certamente, costituisce la premessa indispensabile all'interesse per questa tipologia nel campo della scienza dell'informazione e delle politiche di trattamento dei risultati, è il loro valore potenziale. Questo può variare in relazione a forma, natura e origine del dato (National Science Foundation. National Science Board12-13). Tale relatività è acuita dalle differenze che emergono tanto tra i macrosettori scienze naturali/scienze umane, quanto tra le discipline individualmente considerate, all'interno dei due insiemi.

È doveroso specificare, inoltre, che il dato della ricerca può svolgere il duplice ruolo di prodotto (come risultato o risultante da una ricerca specifica) e di fonte (in qualità di dato già prodotto da qualcun altro e riutilizzato come base di una nuova ricerca): circostanza che è stata designata in riferimento agli antitetici concetti di output e input.⁴ Tale duplicità rende evidente la configurazione di un sistema di circolazione e scambio di conoscenza che, laddove si prospetti un livello aperto di condivisione, trova le sue fondamenta nell'azione

⁴"Data are outputs of research, inputs to scholarly publications, and inputs to subsequent research and learning" (Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*115).

del riutilizzo (Murray-Rust). Lo stesso Murray-Rust, richiamato negli studi italiani (De Robbio e Giacomazzi), rileva una differenza nelle pratiche di pubblicazione e utilizzo dei dati tra le tipologie di "Large Science" e "Small Science".⁵ Da un punto di vista terminologico,⁶ il livello aperto di condivisione è, inoltre, la condizione discriminante tra research data e open research data: quest'ultimo si riferisce ai soli dati aperti, mentre il primo non li esclude.

2 I research data nelle scienze umane

Le scienze naturali si differenziano dalle umane, oltre che per campo d'indagine e metodologie, per maggiore quantità di dati prodotti (conseguentemente al livello tecnico e in funzione dell'oggettività che è alla base delle scienze naturali), per tipologia degli stessi (che ne influenza anche il livello di elaborazione) e per grado di necessità ed abitudine nella pratica della loro condivisione e del loro riutilizzo. Tali circostanze hanno reso i dati della ricerca scientifica-naturale protagonisti, non solo degli studi teorici, ma anche delle realizzazioni di sistemi per la loro raccolta, gestione e condivisione, e delle politiche di indirizzo in tema di ricerca. Lo svantaggio che affligge le scienze umane in questo campo, come in quello propinquo dell'open access (Suber), deriva sia dalle caratteristiche della ricerca scientifica-umanistica e, quindi, dei suoi risultati e delle sue fonti, che da correlate motivazioni di natura economica e culturale, relative soprattutto alla diffusione e ai tempi (nella ricerca scientifica-naturale la necessità di condivisione, sia nell'azione del deposito

⁵La distinzione attuata da Murray-Rust tra "Large science" e "Small science" è basata sulla dimensione dell'unità di ricerca che è estesa nel primo caso e ristretta (singolo individuo o laboratorio) nel secondo.

⁶I requisiti che soddisfano l'attributo Open relativamente ai dati sono molteplici. Si veda la definizione di "apertura" proposta dalla Open Knowledge Foundation: <http://opendefinition.org/okd>.

dei risultati che in quella dell'accedere agli stessi, è caratterizzata da una rapidità che si presenta molto più modesta nell'ambito delle scienze umane).⁷

Se per le scienze naturali l'individuazione di ciò che può corrispondere ad un research data è piuttosto intuitiva, per quanto concerne le scienze umane, la questione è più complessa.⁸ La National Science Foundation (12-13) individua diverse categorie di dati ("observational, computational, or experimental") in conseguenza del loro originare da un'osservazione, un calcolo o un esperimento. Nelle le scienze naturali, queste azioni corrispondono alla normale metodologia di indagine, nell'interpretazione e nello studio dei fenomeni, intesi come eventi osservabili, mentre in quelle umane che hanno ad oggetto dei propri studi entità non fisiche, la corrispondenza non è così evidente.

La sovrapposizione tra il concetto di dato e quello di "primary source" (Burrows) costituisce il punto chiave della problematica.⁹ Il

⁷Peter Suber («Promoting open access in the humanities») nella sua diagnosi sul lento avanzare dell'open access nelle scienze umane rispetto alle naturali, individua nove differenze relative alla ricerca nei due diversi ambiti. Sebbene l'autore ponga ad oggetto dell'analisi i risultati della ricerca nella forma di articoli scientifici, molte circostanze osservate sono valide anche nel caso dei research data, denotando un contesto in cui l'accesso aperto nelle scienze umane si presenta come "less urgent and harder to subsidize than in the sciences".

⁸In Borgman («The digital future is now: A call to action for the humanities») la domanda "What constitute data in the humanities?" è indagata ma non trova risposta definitiva. In conclusione all'opera l'autrice afferma, in riferimento alla suddetta domanda e ad altre quattro problematiche relative al digital humanities: "Answering these questions will enable the digital humanities community to be more articulate about its scope and its goals, and better positioned to identify their requirements for infrastructure" (il tema è affrontato anche in Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* 215-217).

⁹Burrows («Sharing humanities data for e-research: conceptual and technical issues») sostiene che non attuare una distinzione tra "primary source" e "data" nelle scienze umane "would be analogous to describing the stars and galaxies as an astronomer's 'data'".

legame tra i concetti è evidente nella sopracitata definizione della OECD (*Principles and Guidelines for Access to Research Data from Public Funding*) dove i research data sono i "factual records" usati come "primary sources". Tale ruolo è relativo al processo di ricerca: il dato come registrazione effettiva del reale diviene fonte primaria della ricerca. Eppure, difficilmente si può pensare ad una configurazione dei dati prodotti ed utilizzati dalla ricerca scientifica-umanistica come factual records, anche se tale circostanza non è da escludere appieno se si pensa, ad esempio, ai risultati di un text mining in analisi testuale, o ai dati archeografici nel campo dell'archeologia. È indubbio, comunque, che da un punto di vista quantitativo e qualitativo, tale tipologia non sia quella preminente, tenendo anche conto della natura dell'oggetto di indagine costituito spesso da entità astratte, concretizzate in forma di rappresentazione. Piuttosto che di dati veri e propri, nella maggior parte dei casi si tratta di data sources (fonti dei dati e dell'informazione), nelle forme più disparate, che vanno da testi ad oggetti.

In sintesi, per individuare cosa si può intendere per research data nell'ambito delle scienze umane, tenendo imprescindibilmente conto del margine di discrezionalità conferito dalla qualifica fondante "primaria" che caratterizza le definizioni e il ruolo dei research data stessi, è necessario distinguere:

- dati intesi come elementi immediatamente presenti alla conoscenza, risultanti da osservazioni di fenomeni, realtà, esperimenti, calcoli (esempi: risultati di text mining in analisi testuale, dati archeografici, questionari, rilevazioni audio su campo, etc.);
- fonti di dati intesi come elementi indagati che forniscono il dato, o su cui il dato si basa (esempi: dipinto, opera letteraria, manoscritto musicale, reperto archeologico, etc.).

Quest'ultima tipologia, ponendosi alla base del dato, laddove la ricerca porti a risultati della prima categoria (che costituiscono i research data veri e propri), ne è appunto la fonte. Nella ricerca scientifica-umanistica la prima tipologia, in quanto a produzione ed utilizzo, non è, come detto, quantitativamente e qualitativamente consistente come in quella scientifica-naturale, dove lo studio di fenomeni fisici dà vita a produzione computazionale o meno di dati grezzi, direttamente registrati. È invece rilevante l'utilizzo di risorse, le quali sono costituite sia da dati che da fonti (es. documenti, immagini, testi etc.), tanto che, come si vedrà nel paragrafo dedicato all'argomento, le infrastrutture europee raccolgono dichiaratamente digital resources. Non si può affermare che, nel caso delle scienze umane, i factual records corrispondano alle primary sources e, quindi, che queste ultime vadano a costituire i research data, ma è possibile rilevare che svolgano lo stesso ruolo, a fondamento di una ricerca. Nelle discipline umanistiche, inoltre, la tecnologia offre in molti casi la possibilità diretta di registrazione, riproduzione, rappresentazione grafica, accesso e collegamento¹⁰ (circostanza meno probabile nel caso delle scienze naturali). La digitalizzazione, qualora integrata ad infrastrutture accessibili, rende difatti possibile la raccolta, la condivisione e l'utilizzo di collezioni di materiale, seppur escludendone la fisicità.

3 I research data nelle politiche europee

L'attenzione rivolta ai dati della ricerca in seno alla Comunità Europea si è esplicitata tramite una serie di Comunicazioni da parte della Commissione Europea (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato economico e sociale europeo*).

¹⁰Un esempio di collezione di fonti di dati digitalizzate, collegate a dati, è Europea: www.europeana.eu.

Sull'informazione scientifica nell'era digitale: accesso, diffusione e conservazione comunicazione; Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza; Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni. Un'agenda digitale europea; Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca), fino a raggiungere la forma di una Raccomandazione (Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione) agli Stati membri. Tale interesse è diretto ai research data e, più in generale, all'informazione scientifica, prodotti in tutti i campi di studio, tra i quali sono citate le scienze umane.¹¹ I research data, come forma parallela ma anche associata¹² alle pubblicazioni vere e proprie, costituiscono l'informazione scientifica, la cui diffusione ampia e rapida ha un ruolo fondamentale in termini di innovazione, progresso, efficienza ed eccellenza della ricerca. L'auspicabilità di tale diffusione, valida comunque per principio, si rende necessaria soprattutto in caso di risultati, siano essi articoli scientifici o dati,

¹¹"La comparsa di una 'scienza dei megadati' è un fenomeno globale e riflette l'importanza crescente che assumono i dati grezzi, sperimentali e desunti dall'osservazione, praticamente in tutti i campi scientifici (scienze umane, biodiversità, fisica delle alte energie, astronomia ecc.)" (Commissione Europea, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza*9).

¹²La Comunicazione parla di un "Continuum dello spazio dell'informazione scientifica, dai dati grezzi alle pubblicazioni in Paesi o comunità diversi". Internet e i nuovi strumenti di informazione e comunicazione consentono, infatti, di utilizzare i dati della ricerca che provengono da esperimenti ed osservazioni, associandoli ad altre fonti di informazione, per estrarne significati (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato economico e sociale europeo. Sull'informazione scientifica nell'era digitale: accesso, diffusione e conservazione comunicazione*3).

nati nell'ambito di ricerche finanziate con fondi pubblici.¹³

Già nel 2009, in tema di infrastrutture elettroniche TIC (Tecnologie dell'informazione e delle comunicazioni) per la e-scienza, il presupposto della nascita di nuovi metodi di ricerca che sfruttano risorse computazionali e collezioni di dati, nonché la consapevolezza dell'"importanza crescente che assumono i dati grezzi, sperimentali e desunti dall'osservazione" (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza*⁹), avevano portato ad individuare come obiettivo l'adozione, da parte dell'Europa, di un "ecosistema coerente e gestibile di archivi di dati scientifici" (11). Gli Stati membri e le comunità scientifiche venivano invitati a rafforzare gli investimenti nelle infrastrutture per questi dati, previsti anche da parte della Commissione stessa nell'ambito del Settimo Programma Quadro, al fine di "sostenere le politiche di conservazione e di accesso ai dati". È su questo terreno di fertile interesse nei confronti dei research data che si colloca la Comunicazione (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*) seguita da Raccomandazione (*Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*), in cui l'attenzione verso gli stessi si fa più dettagliata. Viene riconosciuta, infatti, la tradizionale concentrazione del dibattito sulle sole pubblicazioni e, contestualmente, l'accresciuta importanza del miglioramento dell'accesso ai dati della ricerca, di cui viene fornita la già citata definizione. L'inefficienza degli investimenti pubblici nella ricerca emerge laddove i risultati esistenti in forma di dati non

¹³Vengono comunque contemplati come giustificabili un certo ritardo per la prima utilizzazione dei ricercatori e lo sfruttamento commerciale (3).

sono messi a disposizione di un ampio pubblico per la verifica e l'eventuale utilizzo. Sono individuati i diversi ostacoli allo sviluppo di questa nuova branca di condivisione della conoscenza ed enunciate le iniziative già attuate dalla Commissione (OpenAIRE) e quelle in programma (finanziamenti alle infrastrutture di dati e alla ricerca sulla conservazione digitale). Gli ostacoli individuati relativi allo sviluppo di un accesso ai dati della ricerca e al loro uso e riuso sono costituiti da:

- carenze organizzative e scarsa chiarezza sulle responsabilità;
- assenza di modelli di finanziamento atti ad assicurare l'accesso a lungo termine;
- persistenza del problema dell'interoperabilità tra Paesi e discipline;
- ricercatori e imprese innovative riluttanti per diverse ragioni (dati percepiti di loro proprietà, tempo da dedicare alle operazioni di deposito, assenza di meccanismi di ricompensa e riconoscimento quali meccanismi di citazione e misurazioni del fattore d'impatto)(*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*⁷).

Sebbene tali problematiche siano reali, altri e più specifici (qui omessi) fattori incidono negativamente sul processo. Si può parlare di barriere¹⁴ strettamente relative all'accesso e all'uso dei research data nell'ambito delle infrastrutture ad essi dedicati. Tali barriere sono di tipo legale (diritto d'autore, licenze restrittive, politiche editoriali

¹⁴Sulla definizione dei diversi tipi di "barriers" si vedano Murray-Rust («Open Data in Science») e Suber («Promoting open access in the humanities»).

limitative), finanziario (abbonamenti per accesso ai datasets, pagamento per utilizzo dei materiali) e tecnico (limitata visibilità, limitata durata, impossibilità d'accesso/uso/riuso). La loro esistenza e il peso che esse rivestono nella difficoltà alla realizzazione e all'efficienza di sistemi per la libera ed aperta condivisione dei risultati della ricerca in forma di dati, sono dovuti al valore commerciale che alcuni dati molto spesso possiedono, all'attribuzione della qualifica di opera creativa pertanto suscettibile al diritto d'autore, all'assenza di programmazioni di sostenibilità volte a massimizzare l'investimento e a garantirne gli effetti nel lungo termine.

L'importanza attribuita ai dati della ricerca emerge anche nell'ambito del nuovo programma quadro di ricerca e innovazione (2012-2020) denominato Horizon 2020.¹⁵ La proposta di decisione (*Proposta di decisione del Consiglio che stabilisce il programma specifico recante attuazione del programma quadro di ricerca e innovazione (2014-2020) - Orizzonte 2020*) del Consiglio della Commissione Europea che reca attuazione del detto programma, riconosce nella diffusione e comunicazione dei risultati della ricerca su scala continentale, un "valore aggiunto" al fine di aumentarne l'incidenza (22). In linea con tale consapevolezza si collocano le azioni mirate a sostenere la creazione, lo sviluppo e il funzionamento delle infrastrutture TIC con l'obiettivo di "conseguire entro il 2020 uno spazio europeo unico e aperto per la ricerca on line in cui i ricercatori beneficino di servizi di rete e informatici all'avanguardia, ubiquitari e affidabili, e un accesso aperto e senza soluzione di continuità agli ambienti scientifici on line e alle risorse di dati mondiali" (36). La ricerca nelle scienze sociali ed umane è dichiaratamente integrata in questo specifico obiettivo relativo alle infrastrutture di ricerca e in quelli generali del programma (21). Sono pertanto previste, nell'ambito dello strumento Horizon 2020, delle azioni mirate a realizzare l'accesso aperto ai dati della

¹⁵http://ec.europa.eu/research/horizon2020/index_en.cfm.

ricerca. La Comunicazione prevede, infatti, il varo di un "progetto pilota sull'accesso aperto ai dati della ricerca scaturiti da progetti realizzati in determinati settori di Orizzonte 2020, e sul relativo riuso" (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*⁹). La Raccomandazione denuncia l'urgenza dell'adozione di interventi politici sull'accesso ai dati e, pertanto, raccomanda agli Stati membri di definire politiche chiare che prevedano obiettivi e indicatori di progresso, piani di attuazione e pianificazione finanziaria, al fine di assicurare che "i dati di ricerca prodotti nell'ambito di attività di ricerca finanziate con fondi pubblici siano pubblicamente accessibili, utilizzabili e riutilizzabili per mezzo di infrastrutture elettroniche digitali" (*Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*³).

L'approccio della Commissione Europea nel definire politiche relative ai research data è caratterizzato, si è detto, da un'attenzione rivolta all'informazione scientifica nel suo complesso. I dati della ricerca, qualora aperti ed accessibili, sono uno degli strumenti (insieme agli altri risultati) per potenziare il sistema della ricerca, in un contesto europeo interoperabile e collaborativo tanto al suo interno quanto in ambito extra europeo. Tale impostazione, la quale nella pratica esclude approfondimenti ed analisi rivolte ai research data come un elemento autonomo, scaturisce sia dalla natura dei documenti prodotti dalla Commissione (non si tratta di guide o testi base a carattere tecnico) che dal ruolo propulsivo che essa riveste, come centro di coordinazione tra i vari Stati membri, in merito allo sviluppo e al miglioramento del sistema di ricerca scientifica nel suo insieme.

Di diversa impostazione è il contributo, in forma di guida, realizzato

dalla OECD (*Principles and Guidelines for Access to Research Data from Public Funding*) e rivolto ai singoli Stati per promuovere uno scenario internazionale ed efficiente di scambio e uso dei research data, superando la varietà di leggi, politiche e pratiche nazionali. Il testo offre un'analisi dei diversi aspetti inerenti la definizione di politiche da parte delle istituzioni di ricerca e degli enti finanziatori. Fornisce, inoltre, una definizione chiara ed esaustiva, come visto, di tale categoria di dati, ponendola in una dimensione a sé stante, comunque non secondaria o necessariamente parallela agli altri risultati della ricerca. Dopo una premessa in cui è specificato che il principio di apertura e libero scambio di idee, informazioni e conoscenza è alla base dei sistemi pubblici scientifici degli Stati membri dell'Organizzazione, si riconosce che le nuove tecnologie hanno creato "new fields of application for not only the results of research, but the sources of research: the base material of research data" (9) e che un accesso "effective" a questi dati, apportando benefici e vantaggi enunciati, fornirebbe rendimenti maggiori dai pubblici investimenti. L'oggetto di questo accesso sono i dati derivanti da ricerche che utilizzano i fondi pubblici. I riconosciuti vantaggi apportati dalla condivisione aperta di questi dati sono di carattere sia universale che individuale per i singoli Stati membri. I principi cui l'accesso ai research data deve rispondere (openness, flexibility, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality, security, efficiency, accountability, sustainability) sono, si è detto, enunciati ed analizzati nei loro vari aspetti.

Anche l'UNESCO (Swan) riconosce il valore dei research data definiti, si è visto, come una categoria di "research output" che vanno sempre più acquisendo un ruolo centrale nelle politiche in tema di accesso aperto. Il testo, che ha lo scopo di promuovere l'open access negli Stati membri facilitando la comprensione degli aspetti

connessi, è incentrato sulle problematiche ad esso relative. Sebbene vengano affermate l'appartenenza dei research data all'insieme dei research output (quindi alla research information in generale) e l'inclusione di tale tipologia di prodotti della ricerca nel concetto di accesso aperto, il "target" primario ed originale dell'accesso aperto, è individuato nella "journal literature" (10). È proprio questa categoria l'oggetto principale della trattazione. Nel documento, tuttavia, si sottolineano il protagonismo delle scienze data-intensive nel processo di condivisione aperta dei risultati della ricerca e le differenze in fatto di norme e data management tra le discipline. In merito alle strategie per promuovere l'accesso aperto si ammette la crescente difficoltà nel separare l'accesso aperto alla "literature", dall'open data (riferito ai research data), e la conseguente necessità di includere nelle future strategie a favore dell'accesso aperto, quelle per gli open data. Nonostante questo riconosciuto legame¹⁶ è evidenziata l'esigenza di sviluppare politiche diversificate per questi ultimi, che tengano conto di problematiche relative alla privacy e di circostanze che impediscono la diffusione per altre ragioni.

Una politica sull'accesso ai prodotti della ricerca è modellata sull'oggetto che disciplina: nel definire contenuti, metodologie per il deposito, standard e quanto altro, non si può prescindere dalle caratteristiche dell'oggetto del processo di raccolta, conservazione, condivisione e riutilizzo. Il prodotto nella forma di dato si configura diversamente rispetto ad un articolo. Alcuni aspetti relativi alla suddetta categoria, richiedono, nella definizione di politiche, particolare attenzione: ad esempio i dati possono contenere informazioni riservate (basti pensare ai dati raccolti nel campo di studi delle social sciences) o possedere un valore commerciale. In questi casi si rende necessario, dunque, individuare apposite misure volte a stabilire obblighi e limiti specifici. Le caratteristiche dei dati della

¹⁶Swan (27) parla di un "ecosystem of 'open' issues".

ricerca si differenziano non solo rispetto a quelle degli articoli scientifici (motivo per cui una politica separata per le due categorie di research output eviterebbe omissioni e lacune) ma anche tra i singoli ambiti di ricerca. È stato infatti rilevato che "a generic approach to data curation will not be sufficient to cope with the different data-related needs and expectations of researchers working in different disciplines other than at a superficial level" (Key Perspectives2).

Nonostante la varietà di contesti nell'ambito in cui il tema dei research data è affrontato, a livello di politiche europee, la quale implica diversità di intenti e di approccio alla questione, può dirsi unanimemente riconosciuto il valore dei research data (siano essi legati alla ricerca scientifica-umanistica o scientifica-naturale) ai fini del miglioramento del processo di ricerca e di quanto ad esso correlato. Tale valore si realizza nella condizione di accesso aperto ai dati stessi tramite infrastrutture TIC al fine di renderli pubblicamente accessibili, utilizzabili e riutilizzabili.

I presupposti alla necessità di rendere i dati disponibili al pubblico sono validi per principio poiché i risultati prodotti in seno a progetti e ricerche finanziati da fondi pubblici, dovrebbero tornare alla comunità che li ha finanziati. In questo modo, non solo si rispetta il diritto dei cittadini finanziatori di usufruire potenzialmente dei prodotti finali, ma si ottimizza l'investimento stesso: la ripetizione delle ricerche è limitata (se i dati di ricerche precedenti non sono disponibili, è necessario ripetere la ricerca), i tempi di ricerca sono accelerati grazie alla velocità nel reperimento delle informazioni (anche a livello interdisciplinare), si annullano le spese per l'accesso (ad esempio a certi datasets a pagamento), il sistema della ricerca nel suo complesso si fa più trasparente (la qualità dei dati può essere verificata, così come l'utilizzo dei datasets è misurabile). Tutto ciò influisce positivamente, in modo più o meno diretto, in termini di crescita, progresso, sviluppo a livello economico e sociale (per un

elenco dettagliato dei vantaggi in termini di ottimizzazione dell'investimento pubblico nella ricerca scientifica si veda Organisation for Economic Cooperation and Development¹⁰).

4 La posizione italiana

La concretizzazione in forma reale del valore dei research data necessita di sistemi finalizzati alla raccolta, gestione, conservazione, condivisione, valorizzazione dei risultati della ricerca (e tutto ciò connesso a tali operazioni) e dell'azione degli Stati a sostegno e favore di tali iniziative. Poiché si tratta di veri e propri "sistemi" che, per definizione, corrispondono ad insiemi di strumenti, meccanismi, elementi, l'azione coordinativa risulta essenziale, tanto a livello interno, quanto esterno. Come è facile intuire, anche l'aspetto finanziario rientra nelle azioni a favore di tali sistemi, insieme alla diffusione di una educazione alla cultura del dato della ricerca. L'Italia si trova, attualmente, dinanzi a due circostanze che costituiscono il punto di partenza per l'attuazione di politiche mirate alle suddette imprese: la prima è un vincolo politico, che scaturisce dalle Raccomandazioni e Comunicazioni della Commissione Europea, per il quale, come parte di un organismo aggregativo, dovrebbe conformarsi alle linee di indirizzo comuni; la seconda circostanza è un aspetto pratico costituito dai vantaggi in termini di progresso culturale, economico e sociale, derivanti da sviluppo, miglioramento, crescita, valorizzazione, efficienza, del processo di ricerca scientifica. In questa congiuntura si collocano le partecipazioni italiane alle iniziative intra ed extra europee a favore dello sviluppo dei sistemi di raccolta, gestione, condivisione, dei research data (es. CLARIN, DARIAH, ARIADNE, MAPPA, RDA e iCordi, che saranno indagate nel paragrafo seguente). A confronto con le realtà di altri paesi europei, all'avanguardia in tema di iniziative a sostegno di progetti inerenti

i dati della ricerca, come quella del Regno Unito, è evidente che l'Italia si trovi in una posizione di arretratezza. Sono assenti, innanzitutto, politiche nazionali volte a disciplinare la materia, in termini di definizione dei diversi aspetti legati alla diffusione dei prodotti della ricerca (dall'individuazione dei contenuti, alle responsabilità); manca un punto di riferimento nazionale che agisca come centro di coordinamento; non sono stati delineati modelli finanziari ed organizzativi, tali da garantire una sostenibilità nel lungo periodo delle infrastrutture.¹⁷ Relativamente alle infrastrutture esistenti, oltre al suddetto limite costituito dall'assenza di una pianificazione pluriennale atta a garantire la sostenibilità nel lungo termine, sono state rilevate una limitata capacità di sfruttare i benefici socio-economici connessi alla realizzazione e al funzionamento di una infrastruttura per la ricerca e un'insufficiente presenza di infrastrutture di livello europeo sul territorio nazionale. La suddetta partecipazione ai progetti di infrastrutture europei è effettiva ma opportunamente definita come "assicurata da gruppi di ricerca di massa subcritica" (Ministero dell'Istruzione, dell'Università e della Ricerca66): non è sufficiente dunque ad innescare un fenomeno di larga portata. Il documento Horizon 2020 Italia redatto dal MIUR (8), in tema di comunicazione della ricerca e dei suoi risultati, parla di "limitata capacità" italiana di "trasferimento, diffusione e valorizzazione" che dipenderebbe sia dalle caratteristiche dei ricercatori italiani, sia da servizi di supporto ancora insufficienti da parte delle università e degli EPR nazionali. Lo stesso documento sottolinea, inoltre, la

¹⁷Le strutture per la conservazione sono spesso create per progetti specifici, e i finanziamenti limitati ad un certo periodo (in merito agli ostacoli all'accesso ai dati della ricerca e al loro uso e riuso e alla conservazione a lungo termine si veda Commissione Europea, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*7).

manifestazione di "criticità importanti per il posizionamento del sistema europeo della ricerca e dell'innovazione e, al suo interno, in maniera più accentuata, di quello italiano" (5). Coerentemente col programma comunitario, l'Italia si propone di realizzare "un sistema della ricerca sano ed efficiente, non frammentato e privo di duplicazioni, forte, coeso e strategicamente orientato" (26). Nell'ambito di questo obiettivo, la circolazione e la condivisione dei risultati della ricerca scientifica, assumono una posizione di rilievo. È riconosciuta, infatti, l'importanza che l'accesso (aperto, gratuito e in formato interoperabile) ai dati e alle informazioni che scaturiscono da attività finanziate con fondi pubblici riveste, in relazione al rapporto tra scienza e società e all'ottimizzazione dell'investimento finanziario stesso. Il portale ResearchItaly¹⁸ è individuato come gateway aggregativo delle iniziative sull'accesso aperto del sistema di ricerca nazionale e come piattaforma per elencare repositories locali di università e centri di ricerca e per ospitare un repository nazionale. Per le infrastrutture di ricerca, nel cui sistema italiano sono evidenziate una serie di già citate criticità, è previsto un piano nazionale (PNIR) finalizzato al miglioramento delle stesse. È atteso, altresì, il potenziamento delle infrastrutture esistenti e la realizzazione di nuove, in accordo con l'European Strategy Forum on Research Infrastructures (ESFRI), ed è identificato come valido lo strumento legale European Research Infrastructure Consortium (ERIC)¹⁹ per partecipare a progetti di infrastrutture pan europee.

Le prospettive enunciate nel programma Horizon 2020 Italia, aprono lo sguardo su uno scenario nuovo per il Paese: se questo scenario diverrà realtà, saremo un paese europeo che potrà finalmente partecipare attivamente alla circolazione della conoscenza.²⁰

¹⁸<https://www.researchitaly.it/>.

¹⁹http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric1.

²⁰In riferimento alla circolazione della conoscenza, Janez Potocnik, commissario europeo per la Scienza e la Ricerca, nel 2007, in occasione della presentazione del

5 Infrastrutture europee

La crescente attenzione rivolta ai research data, anche nelle scienze umane, al di là delle teorie, si è concretizzata nella realizzazione di progetti ed infrastrutture europee, a partecipazione internazionale o ad iniziativa di singoli Paesi o centri, che raccolgono rendendo disponibili all'accesso queste risorse, o forniscono gli strumenti per supportare ed agevolare tali iniziative. Il processo si inserisce in molti casi in un contesto di interesse diretto alle risorse digitali relative alla ricerca, in senso allargato e generale. Nella pratica, la distinzione tra research data e dati, infatti, non è netta; i primi sono una sottocategoria dei secondi ed essendo le infrastrutture, come detto, rivolte alla raccolta di digital data o resources derivanti dalla ricerca, principalmente a quest'ultima finalizzata, anche i data sources fanno spesso parte delle collezioni di dati, insieme ai risultati della ricerca intesi come raw. Questa circostanza rende difficoltosa l'individuazione della natura delle risorse effettivamente conservate ed accessibili nelle diverse infrastrutture e quasi impossibile una loro certa e chiara classificazione.

La lista di repositories di research data pubblicata da DataCite,²¹ realizzata tramite l'iniziativa DataBib²² (strumento per l'individuazione e la localizzazione di tali infrastrutture), evidenzia come, tra i Paesi europei, la maggior parte degli archivi che hanno ad oggetto discipline umanistiche, sia localizzata nel Regno Unito. Per altri Paesi come Francia, Svezia, Paesi Bassi, il risultato è solo uno per ciascuno.

Libro verde «The European Research Area: New Perspectives», ha parlato di "quinta libertà". Le altre quattro libertà sono quelle del mercato comune (libera circolazione delle persone, dei servizi, delle merci e dei capitali) enunciate nel Trattato CEE (1957). Per ulteriori informazioni si veda: http://cordis.europa.eu/fetch?CALLER=NEWSLINK_IT_C&RCN=27454&ACTION=D.

²¹<http://datacite.org/repolist>, la lista è in continuo aggiornamento.

²²<http://databib.org>.

Il Registry of research data repositories re3data.org²³ fondato dalla DFG (German Research Foundation) consente di effettuare una ricerca per soggetto dell'archivio e alla categoria *Humanities and social sciences*, corrispondono nove risultati. Per la maggior parte di questi (ad eccezione di due casi) si tratta di repositories europee. Tra i singoli Stati una posizione all'avanguardia nella realizzazione di infrastrutture dedicate alle risorse digitali, comprendenti i research data, spetta, come detto, al Regno Unito.

Il servizio Arts and Humanities Data Service (AHDS)²⁴, fondato da Joint Information Systems Committee (JISC) e Arts and Humanities Research Council (AHRC), è nato nel 1996 come servizio nazionale con lo scopo di raccogliere, conservare e promuovere le risorse elettroniche derivanti dalla ricerca e dall'insegnamento nel campo delle arti e delle discipline umanistiche. Finanziato fino alla fine di marzo 2008, è attualmente decentrato tra istituzioni ospitanti. Suddiviso in cinque aree disciplinari (archeologia, storia, letteratura lingue e linguistica, arti dello spettacolo, arti visive), ingloba delle infrastrutture che condividono i medesimi obiettivi e che raccolgono, rendendo accessibili, quelle che sono definite digital resources tramite dei data archives. L'area disciplinare Archaeology è ospitata da Archaeology Data Service (ADS),²⁵ fondato da un consorzio costituito dal Council for British Archaeology e le Università di Birmingham, Bradford, Glasgow, Kent at Canterbury, Leicester, Newcastle, Oxford e York. Lo scopo del servizio è raccogliere, descrivere, catalogare, conservare e fornire supporto agli utenti relativamente alle "digital resources that are created as a product of archaeological research". L'ADS promuove, inoltre, standard e linee guida per le pratiche di creazione, descrizione, conservazione e uso delle informazioni archeologiche. In collaborazione con enti nazionali e locali che si

²³<http://www.re3data.org>.

²⁴<http://www.ahds.ac.uk>.

²⁵<http://archaeologydataservice.ac.uk>.

occupano di archeologia o che sono coinvolti nel finanziamento alla ricerca archeologica (Arts and Humanities Research Council, Natural Environment Research Council, British Academy, Council for British Archaeology, English Heritage, Society of Antiquaries of London), raccoglie datasets di diversa provenienza e tipologia e risorse che vanno da mappe a text report. Il meccanismo di ricerca si configura all'utente in due differenti sistemi, entrambi riccamente strutturati. Archsearch per i records consente interrogazioni per parola chiave, categorie (What, Where, When) predefinite e risorse (intese come collezioni). Nel sistema denominato Archive è predisposta la ricerca per archivi classificati in soggetto, programma e regione. Sono inoltre presenti funzionalità aggiuntive che non solo consentono precisi restringimenti di ricerca ma offrono strumenti avanzati, come la funzione "mappa" e la ricerca esterna (per ora in fase sperimentale).

Una piattaforma a carattere nazionale francese, lanciata nel Dicembre 2010, la cui realizzazione è stata finanziata dal governo stesso, che raccoglie, arricchisce e fornisce una struttura unificata di accesso ai dati digitali delle scienze umane e sociali, è Isidore.²⁶ Creata dalla TGE ADONIS (oggi confluita nella TGIR Huma-Num) e realizzata dal Centre pour la Communication Scientifique Directe (CCSD) con la partecipazione delle società Antidot, Sword e Mondeca, è definita research platform e raccoglie attualmente 80 collezioni, 2026 fonti e 2.271.736 risorse.²⁷ L'agile interfaccia consente la ricerca immediata per parola chiave, per categorie (tipo di risorsa, categoria, periodo storico, disciplina, collezione etc.), per fonti e per repositories.

La TGIR Huma-Num, in una dedicata sezione al progetto nel proprio sito web,²⁸ affronta la questione della peculiarità che il mondo

²⁶<http://www.rechercheisidore.fr/index>.

²⁷Cifre rilevate in maggio 2013.

²⁸<http://www.huma-num.fr>.

delle scienze umane e sociali presentano in merito alle necessità di condivisione e raccolta dei dati prodotti. Riconoscendo un ruolo fondamentale alle fonti di informazione ("sources d'informations"), in particolare, viene evidenziato come, tra queste, il testo sia fondamentale nel processo di produzione della conoscenza in dette discipline. Il fattore tempo, che spesso rende velocemente superato un testo scientifico, non realizza il medesimo effetto su un manoscritto medievale, il quale può conservare la stessa importanza ed attualità dell'ultimo articolo pubblicato su una rivista internazionale.

Alcune infrastrutture a collaborazione internazionale offrono supporto e servizi allo scopo di facilitare l'accesso e l'uso ai risultati della ricerca nel campo delle discipline umanistiche, realizzando dei veri e propri network: un esempio è Digital Research Infrastructure for the Arts and Humanities (DARIAH)²⁹ che attualmente conta quattordici Paesi membri, tra cui l'Italia. Le origini del progetto risalgono al 2005 ma la fase preparatoria, destinata ad individuare ed istituire gli elementi (fisici, strategici ed umani) nonché gli aspetti giuridici e finanziari necessari all'impresa, è terminata nel Febbraio 2011. La dichiarata missione di DARIAH è sostenere e supportare la ricerca digitale nel campo delle arti e delle scienze umane, realizzando un'infrastruttura a carattere cooperativo, che riunisce gli sforzi nazionali, regionali e locali, in una rete interconnessa di strumenti, persone, informazioni, metodologie etc. L'istituzione di DARIAH è avvenuta nella forma di ERIC e tale quadro normativo favorisce la sostenibilità del progetto a lungo termine: le necessità finanziarie e tecnologiche sono affrontate in collaborazione tra i membri, in un quadro di coordinamento ed omogeneità complessivi. Quella che viene definita la grand vision del progetto, è volta a facilitare l'accesso e l'uso, a lungo termine, a tutti gli "European Arts and Humanities (A+H) digital research data". DARIAH opera

²⁹<http://www.dariah.eu>.

mediante quattro centri di competenza virtuale, i Virtual Competency Centres (VCCs), ciascuno dei quali si occupa di una specifica area (e-Infrastructure, Research and Education, Scholarly Content Management, Advocacy).

Common Language Resources and Technology Infrastructure (CLARIN),³⁰ dedicata esclusivamente alle risorse linguistiche, è un'altra infrastruttura pan europea cui partecipano membri ed istituzioni da trentatré Paesi, che si propone di offrire servizi per fornire un facile accesso alle risorse, tramite un sistema integrato ed interoperabile. Nello specifico, la finalità del progetto è favorire il progresso della ricerca nell'ambito delle scienze umane e sociali realizzando una piattaforma ad accesso unificato che integri, a livello europeo, le language-based resources ad avanzati strumenti, dando vita ad un'infrastruttura distribuita e condivisa al tempo stesso. La fase di preparazione prevista per il progetto è iniziata nel 2008 ed è terminata nel 2011; ad oggi CLARIN è ancora "under construction", eppure sono disponibili ed accessibili un insieme di servizi. Questi si dividono in due tipologie, la prima delle quali è costituita dai servizi destinati agli utenti relativi a ricerca, trasformazione, archiviazione delle risorse: il servizio di deposito, il Virtual Language Observatory (VLO), servizi web e di consulenza. La seconda tipologia è quella dei servizi dell'infrastruttura stessa, di natura tecnica, destinati ai centri CLARIN: CLARIN IdP (Identity Provider), CLARIN Discovery Service, Component registry for the Metadata Infrastructure (CMDI), ISOcat concept registry, Relation Registry.

In tema di infrastrutture dedicate ai dati scientifici, il progetto iCordi³¹ (tramite il quale, si è detto, la Commissione Europea supporta la Research Data Alliance) rappresenta un importante traguardo, soprattutto in termini di attenzione pratica rivolta alla interopera-

³⁰<http://www.clarin.eu>.

³¹<https://www.icordi.eu>.

bilità. Avviato nel Settembre 2012 e finanziato dalla Commissione Europea nell'ambito del Settimo Programma Quadro, il progetto persegue l'obiettivo di realizzare una piattaforma di coordinazione tra Europa e Stati Uniti, al fine di discutere e sviluppare l'interoperabilità delle infrastrutture ed estenderla a livello globale. Nello specifico, la dichiarata strategic vision è dare un importante contributo allo sviluppo della policy per la gestione e la curation degli scientific data, conducendo ad una policy comune che guidi lo sviluppo di un'infrastruttura globale più ampia. ICordi che, ad oggi, conta quattordici partners sarà guidato da un High Level Scientific Forum composto da esperti in materia di policy per gestione e curation, provenienti da entrambi i continenti coinvolti, i quali realizzeranno raccomandazioni strategiche volte a favorire la convergenza di integrazione, interoperabilità e infrastrutture dei dati. L'azione si baserà su tre programmi (analisi, prototipo e workshop): il primo sarà volto a valutare le tecniche di organizzazione dei dati e le soluzioni che emergeranno da varie comunità scientifiche; quello di prototipo coordinerà le attività tra importanti progetti europei e statunitensi, sostenendo esperimenti cross-infrastructure sull'interoperabilità EU-USA; il programma di workshop indagherà la convergenza delle infrastrutture con particolare attenzione ad un ampio insieme di discipline scientifiche.

Per quanto riguarda singole discipline, importanti sviluppi (e quasi esclusive iniziative italiane)³² si sono realizzati nel campo dell'archeologia, conseguentemente alle caratteristiche della materia stessa che, per necessità e metodologie, rendono la condivisione dei dati della ricerca un'esigenza, più che un valore aggiunto. Il

³²L'Italia, tramite istituzioni e centri, partecipa ai progetti di infrastrutture europee, come CLARIN e DARIAH. In campo archeologico il progetto Metodologie Applicate alla Predittività del Potenziale Archeologico (MAPPA) ha realizzato MOD (MAPPA Open Data), un archivio digitale archeologico, disponibile all'indirizzo: <http://mappaproject.arch.unipi.it/mod/Index.php>.

progetto Advanced Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE)³³ lanciato nel Febbraio del 2013, per la durata prevista di quattro anni e finanziato nell'ambito del Settimo Programma Quadro della Commissione Europea, mira alla creazione di un'infrastruttura di dati archeologici. L'obiettivo è quello di realizzare un'infrastruttura che consenta un accesso transnazionale ai dati, agli strumenti e alle linee guida, attuando una raccolta di molteplici banche dati, offrendo un punto d'accesso unificato e strumenti per mettere le nuove tecnologie a servizio della ricerca. Coordinato dal Polo Universitario Città di Prato dell'Università degli studi di Firenze, con la partecipazione del Ministero per i Beni e le Attività Culturali ed altri istituti italiani, raccoglie partner da sedici Paesi europei.

Questo rapido sguardo ad alcuni dei principali esempi di infrastrutture dedicate ai research data nelle discipline umanistiche, mostra come l'Europa si muova su due percorsi differenti ma paralleli ed intrecciati. Da un lato, infatti, si assiste alla creazione di infrastrutture di portata internazionale che raccolgono, supportano, indirizzano, rispondendo ad una necessità di coordinazione e collaborazione; dall'altro, l'oggetto di tale azione coordinativa sembra essere il processo di creazione e necessità di condivisione dei dati della ricerca, organizzati in altre infrastrutture tramite autonome iniziative, o meno. Il ruolo dei singoli Stati nel definire politiche volte a promuovere, sostenere, potenziare, valorizzare questo processo in atto, è determinante, non solo per rendere efficienti le singole imprese, ma anche al fine di limitare le disomogeneità, favorendo la realizzazione di sistemi facilmente individuabili ed utilizzabili di per sé ed integrati tra loro.

³³<http://www.ariadne-infrastructure.eu>.

6 Conclusioni

I research data costituiscono una tipologia di materiale che possiede un valore tale, nell'ambito della ricerca e nella più generale condivisione di conoscenza, da trovare posto nelle politiche europee e documenti di organizzazioni internazionali. La produzione e l'esistenza di questo tipo di dati, integrate alle possibilità offerte dalle tecnologie informatiche (sistemi di raccolta, gestione, conservazione, linguaggi e formati interoperabili, rete internet) hanno portato alla consapevolezza dei vantaggi, a diversi livelli, derivanti da una loro libera condivisione (e dal libero uso ed eventuale ri-uso). In particolare, nel caso di research data prodotti nell'ambito di ricerche finanziate con fondi pubblici, alla ragione dei vantaggi apportati alla scienza, al progresso, all'economia, alla società, si aggiunge quella relativa alla necessità (se non dovere) di rendere disponibili i risultati della ricerca allo stesso pubblico che l'ha finanziata. Se nella teoria (com'è possibile dedurre anche dai documenti della Commissione Europea, della OECD e dell'UNESCO) tutto ciò è riconosciuto e, nel caso della Commissione Europea, la consapevolezza indirizzata verso reali iniziative, nella pratica esistono e persistono delle difficoltà. Gli ostacoli alla realizzazione, allo sviluppo e all'efficienza di sistemi per la libera ed aperta condivisione dei risultati della ricerca in forma di dati sono di diversa natura: finanziari, organizzativi, tecnici, legali, culturali. Individuarli e trovare risposte e soluzioni è fondamentale al fine di realizzare validi e solidi modelli ed infrastrutture atti a garantire la corretta raccolta, diffusione, conservazione e condivisione dei research data. Fino ad allora i dati continueranno comunque ad essere prodotti ma molti non saranno reperibili, altri lo saranno per limitati periodi di tempo, altri ancora non lo saranno per tutti (per circostanze legate ad esempio a visibilità, formati o condizioni di accesso). Ciò renderà inefficienti (o efficienti in misura ridotta) gli investimenti: molte ricerche saranno ripetute, i tempi

per il reperimento di materiale potrebbero essere lunghi, le risorse investite in progetti senza adeguate pianificazioni finanziarie e di rischio potrebbero essere perse. La situazione europea, nel complesso, sembra comunque muovere verso una via favorevole alla realizzazione di politiche, sistemi ed infrastrutture dedicati ai research data, attenti a ciò che ne garantisce l'efficienza (sostenibilità finanziaria e tecnica, interoperabilità, etc.) basti l'esempio dei progetti DARIAH e CLARIN.

In Italia l'intenzione di affrontare la problematica relativa ai research data emerge per la prima volta nel documento Horizon 2020 Italia. Affinché questa dichiarata intenzione non rimanga tale ma si concretizzi, fondamentali saranno delle azioni da parte dello Stato, tra cui:

- nell'ambito dell'individuazione delle responsabilità, la realizzazione di un centro per il coordinamento nazionale in tema di accesso aperto alle risorse derivanti dalla ricerca italiana;
- investimenti finanziari in iniziative di portata nazionale e partecipazione a quelle pan europee, come ad esempio la realizzazione di un'infrastruttura nazionale, integrata a sistemi europei;
- promozione di una cultura all'accesso aperto, che allontani i timori (spesso scaturiti dalla non conoscenza di nuovi modelli e pratiche di diffusione, come le licenze Creative Commons), proponendo nuove soluzioni;
- definizione di politiche complete e chiare che tengano separate l'accesso aperto agli articoli scientifici, da quello ai research data. È inoltre importante mantenere concettualmente separato l'ambito dell'open data in senso generico (il quale comprende ad esempio i dati delle pubbliche amministrazioni) dai dati

della ricerca che possiedono caratteristiche e problematiche proprie e specifiche e necessitano di azioni mirate. Dette politiche dovrebbero stilare pianificazioni finanziarie pluriennali, ma anche strategiche (come piani di gestione di rischio), al fine di garantire la sostenibilità nel lungo termine.

In ultimo sarebbe opportuno ed auspicabile, che il tema dei research data nelle discipline scientifiche-umanistiche, il quale, si è visto, soffre di secondarietà rispetto all'attenzione rivoltagli in quelle scientifiche-naturali, sia considerato nelle peculiarità che contraddistinguono la ricerca e i risultati della stessa in tale ambito. Guardare alla realtà delle discipline umanistiche ponendo attenzione alla specificità che la caratterizza incentiverebbe maggiore interesse e sviluppo, senza nulla togliere al confronto e alla condivisione di soluzioni, pratiche, problematiche ed attributi tra i due settori.

Riferimenti bibliografici

- Australian National Data Service. *What is research data?* ANDS, 2011. <http://ands.org.au/guides/what-is-research-data.pdf>. (Cit. a p. 2).
- Borgman, Christine L. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge: MIT, 2007. (Cit. alle pp. 4, 6).
- . «The digital future is now: A call to action for the humanities». *Digital humanities quarterly* 3.4. (2009). <<http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html>>. (Cit. a p. 6).
- Burrows, Toby. «Sharing humanities data for e-research: conceptual and technical issues». (2011). <<http://ses.library.usyd.edu.au/bitstream/2123/7938/1/Burrows.pdf>>. (Cit. a p. 6).
- Commissione Europea. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza*. Bruxelles, 2009. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:IT:PDF>. (Cit. alle pp. 9, 10).
- . *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni. Un'agenda digitale europea*. Bruxelles, 2010. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:IT:HTML>. (Cit. a p. 9).
- . *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*. Bruxelles, 2012. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0401:FIN:IT:PDF>. (Cit. alle pp. 1, 3, 9–11, 13, 18).
- . *Comunicazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato economico e sociale europeo. Sull'informazione scientifica nell'era digitale: accesso, diffusione e conservazione comunicazione*. Bruxelles, 2007. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0056:FIN:IT:PDF>. (Cit. alle pp. 8–10).
- . *Proposta di decisione del Consiglio che stabilisce il programma specifico recante attuazione del programma quadro di ricerca e innovazione (2014-2020) - Orizzonte 2020*. Bruxelles, 2011. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0811:FIN:it:PDF>. (Cit. a p. 12).
- . *Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*. Bruxelles, 2012. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:194:0039:0043:IT:PDF>. (Cit. alle pp. 1, 9, 10, 13).

- De Robbio, Antonella e Silvia Giacomazzi. «Dati aperti con LODe». *Bibliotime* 14.2. (2011). <<http://www.aib.it/aib/sezioni/emr/bibtime/num-xiv-2/derobbio.htm>>. (Cit. a p. 5).
- Key Perspectives. *Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study*. Digital Curation Centre, 2007. <http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP%20SYNTHESIS.pdf>. (Cit. a p. 16).
- Ministero dell'Istruzione, dell'Università e della Ricerca. *Horizon 2020 Italia*, 2013. <https://www.researchitaly.it/uploads/50/HIT2020.pdf>. (Cit. alle pp. 18, 19).
- Murray-Rust, Peter. «Open Data in Science». *Serials Review* 34.1. (2008): 52–64. <<http://www.sciencedirect.com/science/article/pii/S009879130800004X>>. (Cit. alle pp. 5, 11).
- National Science Foundation. National Science Board. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. Washington: National Science Board, 2005. http://www.nsf.gov/nsb/meetings/2005/LLDDC_draftreport.pdf. (Cit. alle pp. 4, 6).
- Organisation for Economic Cooperation and Development. *Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publications, 2007. <http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf>. (Cit. alle pp. 2, 7, 14, 17).
- Suber, Peter. «Promoting open access in the humanities». *Syllecta Classica* 16. (2005): 231–246. <<http://www.earlham.edu/~peters/writing/apa.htm>>. (Cit. alle pp. 5, 6, 11).
- Swan, Alma. *Policy Guidelines For The Development And Promotion Of Open Access*. Paris: UNESCO, 2012. <http://unesdoc.unesco.org/images/0021/002158/215863e.pdf>. (Cit. alle pp. 2, 14, 15).

Ai fini di una corretta indicizzazione, si invitano i lettori a citare esclusivamente il testo in lingua inglese; l'unico, infatti, che presenta l'indicazione del numero di pagina, l'abstract, le keywords e le date del processo redazionale.

Funari, M. "Research data e scienze umane: un quadro europeo". *JLIS.it*. Vol. 5, n. 1 (Gennaio/January 2014): Art: #8927. DOI: [10.4403/jlis.it-8927](https://doi.org/10.4403/jlis.it-8927). Web.

