



**UNIVERSITÀ DI PISA**

Dipartimento di Filologia, Letteratura e Linguistica  
Corso di Laurea Triennale in Informatica Umanistica

**TESI DI LAUREA**

**LE PROMESSE (MANTENUTE E NON) DELL'OPEN SCIENCE**

**RELATORI**

Prof. Vittore CASAROSA

Prof.ssa Enrica SALVATORI

**CANDIDATO**

Alice BERGONZINI

**ANNO ACCADEMICO 2020-2021**

# INDICE

<b>1</b>	<b>INTRODUZIONE</b>	<b>4</b>
<b>2</b>	<b>OBBIETTIVI DELL'OPEN SCIENCE</b>	<b>7</b>
2.1	Un sistema oligopolistico	7
2.2	I metodi di valutazione	8
2.2.1	Peer Review	9
2.2.2	L'indice H	9
2.2.3	L'impact Factor	10
2.3	La frode scientifica	11
2.3.1	I risultati negativi	11
2.3.2	La questione delle ritrattazioni	12
2.3.3	La crisi della riproducibilità	13
2.4	Soluzioni <i>open</i>	13
2.4.1	Obbiettivi dell'Open Science	13
2.4.2	I dati FAIR	14
2.4.3	Pratiche e strumentazioni	15
2.4.4	Benefici	17
<b>3</b>	<b>POLITICHE E INIZIATIVE</b>	<b>18</b>
3.1	Le politiche europee	18
3.1.1	Politiche nazionali	19
3.1.2	In Italia	20
3.2	Iniziative	22
3.2.1	Plan-S	22
3.2.2	EOSC	24
3.3	Open Science e Covid-19	24
3.3.1	Open Science applicata	25
3.3.2	Aspetti da migliorare	26
<b>4</b>	<b>I CAMBIAMENTI CONCRETI</b>	<b>27</b>
4.1	Open Access nel tempo	27
4.1.1	Le vie dell'Open Access	28

4.1.2	Sviluppi nel mondo delle pubblicazioni	28
4.2	L'Open data nel tempo	34
4.2.1	Sviluppi nei repository di dati	35
4.2.2	Il punto di vista dei ricercatori	37
4.3	Il sito open-science.it	38
4.3.1	Ricerche preliminari	38
4.3.2	Lo schema di metadati	39
4.3.3	Inserimenti di prova	42
<b>5</b>	<b>CONCLUSIONI</b>	<b>44</b>
<b>6</b>	<b>APPENDICE</b>	<b>45</b>
6.1	XML Schema per lo schema di metadati "Documenti"	45
6.2	XML Schema per lo schema di metadati "Eventi"	46
	<b>BIBLIOGRAFIA</b>	<b>49</b>

# 1 INTRODUZIONE

L'Open Science è un movimento che nasce in contrapposizione a varie problematiche all'interno del sistema di comunicazione scientifica attuale e si fonda sull'apertura dei risultati della ricerca per consentire una loro trasmissione trasparente e democratica. In questa tesi, suddivisa in tre capitoli principali, si tenta di fare luce sul tema e osservarne le manifestazioni e gli sviluppi, ponendo particolare attenzione alle attività registrate nelle riviste e negli archivi. Infine, si illustra il lavoro svolto in collaborazione col CNR per il progetto open-science.it.

La comunicazione della letteratura accademica si basa da secoli sulla divulgazione tramite riviste scientifiche e simposi. La stampa ha rappresentato (e continua in parte a rappresentare ancora) lo strumento fondamentale per la diffusione delle conoscenze ottenute dalla ricerca scientifica. Questo sistema è iniziato ad andare in crisi con l'avvento prima di Internet e poi del Web. La possibilità di condividere informazioni globalmente in maniera istantanea ha infatti messo in discussione, in primo luogo, l'utilità dell'editoria cartacea e in secondo delle riviste (tradizionali ed elettroniche). Tuttavia, il sistema tradizionale di comunicazione scientifica resiste, anzi in qualche misura è riuscito a trovare nel web nuova linfa vitale e nel tempo ha provocato una serie di problematiche non indifferenti che limitano il potenziale della ricerca. Innanzitutto, ci troviamo di fronte a un sistema che trae profitto dalla mercificazione della conoscenza. I costi di produzione a cui devono andare incontro gli editori sono ormai estremamente ridotti perché con Internet le pubblicazioni vengono presentate in forma digitale online, ma nonostante questo i contenuti sono bloccati dal cosiddetto *paywall* che obbliga l'utente interessato a pagare dei prezzi eccessivi. Sovente i prezzi di acquisto di abbonamenti come di singoli saggi continuano ad aumentare in maniera apparentemente immotivata, visto che i *paper* vengono consegnati dai ricercatori agli editori in maniera gratuita. Gli aumenti – uniti al contestuale taglio delle risorse per la ricerca - hanno portato all'applicazione di prezzi talmente alti da costringere varie biblioteche e università a rinunciare ad abbonarsi a determinate riviste. Il sistema si mantiene in piedi grazie al prestigio che rappresentano alcuni editori e alla capacità di alcuni aggregatori di riviste di offrire alle istituzioni pacchetti di abbonamento. Riuscire a pubblicare nelle più famose e acclamate riviste scientifiche è un fattore spesso determinante per la carriera dei ricercatori. Inoltre, gli oligopoli che dominano questo sistema, controllano spesso anche i metodi di valutazione della ricerca che spesso consistono in metriche che valutano accessi e citazioni incrociate. Non a caso si è consolidato tra i ricercatori il motto *publish or perish*, in italiano “pubblica o muori”, emblema della pressione istituzionale che i ricercatori provano quotidianamente a pubblicare in maniera intensiva, spesso sacrificando la qualità della ricerca in sé. Alcune volte i ricercatori arrivano persino a compiere atti di frode scientifica, manipolando i risultati a favore di un articolo più appetibile agli occhi dell'editore. Ciò ha causato un aumento nel numero di articoli ritrattati ogni anno e una crisi della riproducibilità.

Questi evidenti limiti del sistema attuale, che di fatto ostacolano il progresso della ricerca, hanno portato alla nascita del movimento Open Science. Infatti, l'obiettivo della Scienza Aperta è quello di rendere la produzione della ricerca accessibile a tutti creando un ambiente democratico all'interno del quale ognuno ha la possibilità di dare il proprio contributo alla ricerca.

Open Science (OS) è un concetto ombrello che comprende sotto di sé tanti altri concetti, tra cui Open Access, Open Data, Open Source ecc. L'accesso aperto e gratuito alle pubblicazioni è l'obiettivo principale dell'Open Access, ma fare Open Science significa rendere trasparente e pubblico ogni passo e ogni tipo di risultati della ricerca. Significa, inoltre, seguire determinate pratiche che rendono i dati adatti all'apertura. Infatti, condividere non significa per forza fare Open Science. I principi che i ricercatori devono seguire per rendere i dati congrui alla filosofia *open* sono riassunti nell'acronimo *FAIR: Findable, Accessible, Interoperable e Reusable*. L'intento è quello di creare una rete di condivisione in cui i risultati della ricerca circolano liberamente per permettere una maggiore collaborazione anche tra le varie comunità scientifiche. Avere accesso ad una ricerca così trasparente significa anche avere a disposizione una ricerca più affidabile e più riproducibile. Lo scopo è quello di far diventare tutto questo una norma, ma per farlo c'è bisogno che le istituzioni intervengano nel modificare il sistema di incentivi e nel creare nuovi metodi di valutazione che valorizzino la qualità della ricerca. Col passare del tempo l'Open Science viene sempre più accolta anche come obiettivo di enti governativi, istituzioni ed enti di ricerca. Questi attraverso la stesura di regolamenti permettono ai ricercatori di potersi convertire a una ricerca basata sulle pratiche dell'Open Science ponendo le basi per un cambiamento concreto. La Commissione Europea (EC) ha posto l'Open Science come uno dei suoi obiettivi nella strategia per la ricerca e l'innovazione e ha avviato progetti come *Horizon 2020* e *Horizon Europe*, che prevedono l'apertura dei risultati di tutta quella ricerca finanziata dai fondi europei. Sono nate anche molte iniziative come l'EOSC, l'*European Open Science Cloud*, l'ambiente virtuale dedicato ai ricercatori europei per la condivisione e la gestione dei dati, oppure Plan-S, un'iniziativa creata da Coalition-S, un gruppo di enti finanziatori, in collaborazione con la Commissione Europea che prevede che tutta la ricerca finanziata dai membri di questo gruppo sia pubblicata su riviste Open Access. Tanti paesi europei, sulla scia della Commissione Europea, hanno provveduto a stendere delle politiche per l'Open Science. L'Italia, un po' in ritardo rispetto ad altri paesi europei, a fine 2020 ha inserito nel programma nazionale per la ricerca un piano per la Scienza Aperta. In generale, dunque, il beneficio universale dell'Open Science è universalmente riconosciuto. La domanda è, in che modo e con quale velocità si sta sviluppando l'Open Science? Per dare una risposta a questa domanda ho analizzato alcuni parametri sia nell'ambito dell'Open Access che nell'ambito dell'Open Data. Per quanto riguarda l'accesso aperto si registra un incremento marcato soprattutto negli ultimi anni. Il numero di riviste Open Access è circa 10 volte il numero di 15 anni fa e anche i vari archivi per i *preprint* sono aumentati di circa 24 volte rispetto al 2006.

L'OA sta diventando una modalità di pubblicazione sempre più utilizzata. L'insieme di articoli pubblicati ad accesso aperto rappresentano più di un terzo dell'intera attività di pubblicazione nel 2020. Nel mondo dell'Open Data, invece, si registra sempre un incremento sia nel numero di dataset ad accesso aperto, che nel numero di repository di dataset, ma leggermente più lento rispetto al mondo delle pubblicazioni. Dunque, si può affermare che l'Open Science sia in costante crescita e che sia un modo di fare ricerca a cui sempre più ricercatori si avvicinano. In questo è indispensabile il ruolo delle istituzioni che attraverso la stesura di *policy* incentivino i propri ricercatori a pubblicare secondo i principi dell'Open Science.

Come contributo alla promozione di queste pratiche ho potuto partecipare al progetto del CNR per lo sviluppo del portale open-science.it, un sito che entro la fine del 2021 sarà aperto al pubblico. Questo fungerà da punto di appoggio per ricercatori, enti finanziatori e chiunque voglia avvicinarsi alle pratiche dell'Open Science. Il sito ospiterà anche un catalogo di risorse digitali, nella cui strutturazione e nel cui sviluppo ho avuto il piacere di partecipare attivamente. In particolare, mi sono occupata di ragionare sulla struttura, sviluppare uno schema di metadati per le risorse e infine di inserire nel catalogo un primo insieme di oggetti. Per definire i metadati è stato necessario effettuare una ricerca per trovare delle soluzioni ottimali. Una volta ideato lo schema di metadati, questo è stato definito in maniera formale attraverso il linguaggio XML Schema che mi ha permesso di esplicitare alcuni aspetti, come numero di occorrenze e valori. Una volta completato il lavoro, lo schema è stato consegnato al personale tecnico che si è occupato di creare un'istanza del catalogo in cui ho potuto fare i primi inserimenti di prova. I metadati definiti sono stati trasformati in campi da compilare durante la creazione di una risorsa nel catalogo e questi, una volta creata la risorsa, vengono mostrati nella scheda descrittiva. Lo schema di metadati si è rivelato essere ottimale per la descrizione dei documenti che saranno ospitati dal catalogo.

## 2 OBIETTIVI DELL'OPEN SCIENCE

L'Open Science nasce con la volontà di contrapporsi ad alcuni meccanismi della comunicazione scientifica che con l'avanzare del tempo risultano controproducenti e limitanti. Il fondamentale scopo del movimento Open Science si può estrapolare già dal nome stesso: *open*. L'apertura e l'accesso libero ai risultati della ricerca sono alla base degli ideali che la Scienza Aperta si propone di portare avanti. Il sistema attuale, tuttavia, è colmo di dinamiche che ostacolano questo approccio e che non contribuiscono al regolare avanzamento del progresso scientifico, ma che, anzi, lo rallentano. Esistono varie problematiche all'interno di questo sistema che l'Open Science, come movimento e come comunità, cerca di affrontare e risolvere attraverso nuove metodologie e procedimenti per la divulgazione del sapere. Questo capitolo intende mettere in luce i singoli fenomeni che rappresentano una difficoltà nel sistema attuale e i modi in cui l'Open Science vuole far loro fronte.

### 2.1 Un sistema oligopolistico

Per fare ricerca e, soprattutto, per fare in modo che questa si trasformi in opportunità lavorative e avanzamenti di carriera, bisogna conformarsi a un sistema ben preciso, ovvero quello delle grandi riviste scientifiche. Per arrivare alla pubblicazione di un *research paper* il procedimento è decisamente lungo ed escludendo la parte di lavoro vero e proprio, cioè le attività di ricerca che porteranno alla scrittura dell'articolo, i restanti passi sono affidati completamente a questi grandi produttori commerciali. Una volta consegnato (in maniera totalmente gratuita) il lavoro del ricercatore alla rivista, questa si occupa di valutarlo, pubblicarlo e divulgarlo. Il problema, però, risiede principalmente nel cosiddetto *paywall*, il blocco all'accesso dell'articolo, sbloccabile esclusivamente tramite il pagamento di un abbonamento. La quantità di denaro richiesta è spesso sproporzionata a quello che ci viene offerto in termini di servizi soprattutto se pensiamo al lavoro concreto che sta dietro all'articolo, svolto nella sua interezza dai ricercatori, i quali consegnano gratuitamente il loro operato cedendo, così, i diritti di copyright interamente alle riviste. Queste, poi, chiudono le pubblicazioni dietro a un abbonamento che gli enti di ricerca, le biblioteche e le università devono pagare per accedervi. La ricerca, così, si trasforma in un prodotto commerciale, dando luogo al fenomeno di mercificazione della ricerca accademica [1].

Nonostante la fruizione di informazioni al giorno d'oggi sia sempre più immediata grazie al Web, questo sistema oligopolistico persiste e indubbiamente governa il mondo della ricerca. Si parla di oligopoli in quanto si osserva un grande sbilanciamento nel controllo delle pubblicazioni scientifiche. È noto, infatti, che il 50% di tutti gli articoli scientifici globalmente sono pubblicati e divulgati da soli cinque editori. La sproporzionata influenza che una rivista come Elsevier ha, non solo sulla divulgazione delle pubblicazioni scientifiche, ma su tutta l'infrastruttura della

ricerca, obbliga i ricercatori e le istituzioni ad appoggiarsi ad essa. Con l'avvento del Web e delle possibilità di aprire la conoscenza anche a chi non se la può permettere, questi grandi editori attuano strategie per mantenere il loro potere. In uno studio condotto nel 2017 è dimostrato come la rivista *Elsevier* abbia investito nel tempo in strumenti e servizi che coinvolgono tutto il ciclo di produzione con l'obiettivo di estendere la propria influenza sull'infrastruttura della ricerca nella sua interezza [2]. Parte di questa infrastruttura è anche la valutazione della ricerca che molte volte condiziona la carriera di un ricercatore. Pubblicare in queste riviste, infatti, rappresenta un prestigio nel mondo della ricerca, che poi si trasforma in opportunità lavorative ed avanzamenti di carriera.

A rafforzare la concentrazione del potere economico nel mondo della divulgazione scientifica ci sono anche le strategie di *bundling*, che attraverso la vendita di abbonamenti fatti a pacchetti beneficiano quelle riviste che da sole non verrebbero prese in considerazione formando così dei gruppi d'élite che si promuovono a vicenda. Le riviste accademiche che rimangono escluse non possono usufruire delle stesse strategie e finiscono per restare ai margini del mercato accademico. Il risultato è quello di un mercato esterno dominato da quelle riviste che sono parte dello stesso pacchetto in cui la competitività è annullata dal comune obiettivo di incrementare i profitti [3].

Questo accentramento di potere porta a svariate conseguenze negative. Una è la cosiddetta *crisi del prezzo dei periodici accademici* che consiste nell'aumento ingiustificato del prezzo degli abbonamenti del 300% tra il 1975 e il 1995 [4]. Ingiustificato perché è importante tenere presente che le riviste scientifiche non pagano i ricercatori per i loro lavori e nemmeno i revisori. La ricompensa in questi casi è rappresentata esclusivamente da benefici in termini di reputazione. In più grazie alla tecnologia moderna produrre una rivista non è più costoso come una volta. I costi smisurati e l'affermarsi di convenzioni e standard portano alcune realtà a rimanere escluse da questo sistema e quindi a rimanere ai margini della realtà scientifica internazionale. Tante biblioteche e università si sono trovate obbligate a disdire certi abbonamenti. Poi, le istituzioni di ricerca si ritrovano a essere totalmente dipendenti da questi editori rendendo così difficile la possibilità di fare ricerca al di fuori degli schemi convenzionali. Insieme, queste conseguenze confluiscono in un sistema difettoso che costruisce barriere intorno alla cultura e privatizza i contenuti della ricerca.

L'apertura dei risultati della ricerca rappresenta una maniera democratica di fare scienza che permette a chiunque di accedervi, eliminando privilegi e gerarchie che non danneggiano soltanto i più svantaggiati, ma l'intera comunità scientifica.

## **2.2 I metodi di valutazione**

Un aspetto molto importante da tenere presente quando si parla di comunicazione scientifica è la valutazione qualitativa delle riviste e delle pubblicazioni in sé. È fondamentale nel percorso

di ricerca poter stabilire l'attendibilità di un determinato risultato. Purtroppo, però, convenzionalmente vengono utilizzati degli indici quantitativi, come il numero di citazioni o altri tipi di metriche che, però possono portare a delle conclusioni fuorvianti.

### 2.2.1 Peer Review

La revisione paritaria, meglio conosciuta come *peer review*, è la valutazione a cui vengono sottoposti i lavori dei ricercatori prima di essere confermati per la pubblicazione. Il ruolo di *reviewer* è affidato a gruppi di specialisti, la maggior parte delle volte sono anch'essi ricercatori, e il loro compito ha due fondamentali funzioni:

- 1 Valutare e giudicare la validità di un lavoro in base alla metodologia usata, l'analisi e le argomentazioni
- 2 Fare una selezione editoriale sulla base dell'atteso impatto che un articolo possa avere

È chiaro che la *peer review* sia un fondamentale ed essenziale tassello nel ciclo di produzione di un articolo. Tuttavia, l'aumento del numero di ritrattazioni<sup>1</sup> registrato a partire dai primi anni del 2000, ha messo in dubbio l'efficacia di questo meccanismo di valutazione e dimostrato come la *peer review* non sia in grado di garantire al 100% la veridicità dei contenuti [5]. L'elemento che contribuisce maggiormente alla parziale inefficacia della *peer review* è l'anonimità e la mancanza di trasparenza. Questo dà ai *reviewer* il potere di sovvertire i risultati della valutazione e di fare preferenze sulla base di parametri soggettivi. In questo modo potrebbero avvenire discriminazioni e favoritismi e i *reviewer*, protetti dall'anonimità, non subirebbero alcuna conseguenza. L'Open Science risponde a queste difficoltà con l'*Open peer review*, un concetto per cui esistono tante definizioni diverse, ma tutte si accomunano per l'approccio più trasparente ai processi di valutazione [6].

### 2.2.2 L'indice H

Come risulta necessario la possibilità di valutare un lavoro di ricerca è altrettanto fondamentale poter determinare la qualità di un ricercatore di per sé. L'indice H, o anche indice di Hirsch, serve a quantificare l'impatto di un autore sul mondo della ricerca tramite un parametro specifico: il numero di citazioni. Questo fattore è determinante per le carriere dei ricercatori in quanto le istituzioni si basano su di esso per assegnare ruoli di rilievo. Quella che in letteratura si chiama *Legge di Campbell*<sup>2</sup> ci spiega che quando si ha una metrica quantitativa come questa è facile sfruttarla a proprio vantaggio dando origine a varie conseguenze negative,

---

<sup>1</sup> Col termine *ritrattazione* ci si riferisce a un articolo che è stato rimosso dall'archivio di una rivista scientifica nonostante fosse stato accettato tramite *peer review* in precedenza. Un articolo può essere ritrattato per vari motivi, ma di solito è perché ci si trova di fronte a un caso di frode scientifica

<sup>2</sup> Legge di Campbell: "Più un indicatore sociale quantitativo è utilizzato per il social-decision making, più sarà soggetto a pressioni di corruzione e più sarà adatto a distorcere e corrompere processi sociali che doveva misurare"

come la frode scientifica o l'uso inappropriato dell'autocitazione<sup>3</sup> per inflazionare l'indice [7]. Così viene premiato il lavoro di un ricercatore non sulla base del contenuto dei suoi studi, ma sulla base di quanti studi è riuscito a pubblicare. Questo modo di approcciarsi alle pubblicazioni con l'obiettivo di produrne il più possibile incentiva i ricercatori a cambiare i propri metodi in modo da intensificare la produttività a discapito della qualità.

### 2.2.3 L'impact Factor

Nel sistema di pubblicazioni attuale il prestigio e la qualità di una rivista scientifica vengono determinati attraverso i cosiddetti indicatori di impatto. Il più noto e più utilizzato sistema di classificazione delle riviste scientifiche è l'*Impact Factor* (IF), utilizzato per lo *Science Citation Index* e inventato da Eugene Garfield, e consiste nella misurazione del numero medio di citazioni ottenute da una rivista in un determinato lasso di tempo. Come per l'indice H, anche qui si tratta di una misura quantificabile e quindi anche in questo caso ci troviamo di fronte a un'unità di misura facilmente manipolabile. Ci si aspetterebbe di trovare ai primi posti dell'indice delle citazioni delle riviste più affidabili. Invece, è stato dimostrato che la frequenza con cui avvengono ritrattazioni è fortemente correlata all'Impact Factor di una rivista. Riviste ad alto IF tendono ad avere un più alto numero di ritrattazioni. Il problema è che i ricercatori e le istituzioni di ricerca si affidano a questi indici e quindi aspirano ad apparire nelle migliori riviste per ottenere una questione di prestigio. Si viene a creare, così, una sorta di ossessione per la valutazione [8, 9]. Il ruolo centrale che ha l'Impact Factor nella designazione dell'importanza delle riviste e degli autori che scrivono in essa, porta gli autori a servirsi di strumenti stilistici per far sembrare il proprio articolo più appetibile. Questo ha provocato la produzione di sempre più articoli cosiddetti *literature-driven* che hanno inflazionato la letteratura scientifica e causato una perdita di valore delle singole pubblicazioni [3].

Insieme, questi metodi di valutazione, creano un sistema di pubblicazioni gerarchico a cui le istituzioni si appoggiano e si affidano per la distribuzione dei posti di lavoro. Le carriere dei ricercatori dipendono da questo sistema portandoli ad acquisire un'attitudine improntata alla pubblicazione intensiva invece che al contributo scientifico. La qualità dei prodotti scientifici, determinata tramite questi sistemi bibliometrici<sup>4</sup>, è determinata da aspetti che prescindono dalla reale importanza che hanno per la ricerca. Si perde di vista lo scopo vero della ricerca, che è quello di contribuire al progresso scientifico. L'Open Science sostiene una valutazione diversa che porti il ricercatore a focalizzarsi sul contributo che può dare alla ricerca.

---

<sup>3</sup> L'autocitazione è quando in un saggio o in articolo si fa riferimento ai propri lavori precedentemente pubblicati

<sup>4</sup> Per indicatori bibliometrici si intendono delle metriche usate per valutare l'impatto delle riviste o dei singoli autori come per esempio l'*H-index*

## 2.3 La frode scientifica

Ora che abbiamo una visione un po' più ampia del sistema in cui si colloca la ricerca al giorno d'oggi vediamo quali siano gli effetti sociologici che producono questi meccanismi.

Dopo la Seconda guerra mondiale la ricerca acquista sempre più importanza e aumenta quindi anche la disponibilità di finanziamenti pubblici. Gli scienziati erano meno e ottenere fondi era più semplice. Nel quadro contemporaneo la situazione è ribaltata. Da una parte il numero di ricercatori è aumentato drasticamente, d'altra parte, però, i finanziamenti sono sempre gli stessi. I ricercatori per fare carriera devono spiccare in modo particolare dal mucchio attraverso la pubblicazione di articoli in riviste di alto profilo [10]. Si ritrovano a lavorare in un ambiente dominato dalla competitività in cui il proprio lavoro non è valorizzato secondo criteri di qualità, ma secondo indici di citazione, dove chi ha più citazioni è considerato uno scienziato migliore. Il "*paper*" (o meglio il *corpus di paper*) diventa così l'unità atomica della letteratura accademica su cui si basa la carriera di ogni ricercatore [3]. Per emergere, ma anche solo per sopravvivere professionalmente i ricercatori sono obbligati alla produzione di massa causando così un incremento nella possibilità di commettere errori e, purtroppo, anche nell'adottare atteggiamenti scorretti e fraudolenti. Questo fenomeno di pressione istituzionale è riassunto nell'imperativo *publish or perish*, che tra chi lavora nell'ambito della ricerca è diventato notevolmente popolare e che si traduce letteralmente con "pubblica o muori". Le conseguenze di questo paradigma si osservano attualmente nel mondo delle pubblicazioni. Le tradizionali motivazioni che muovevano i ricercatori a fare scienza, ovvero il progresso scientifico, la curiosità, la passione, il contributo all'umanità in generale, si sono trasformate. Spinti da incentivi individualistici, i ricercatori di oggi perdono di vista lo scopo vitale della scienza. Con l'obiettivo in mente di pubblicare il più possibile, producono errori più frequentemente e a volte, purtroppo, corrompono la veridicità dei loro operati [10].

### 2.3.1 I risultati negativi

Sappiamo che per questioni di pubblicabilità e appetibilità, risultati cosiddetti negativi, il più delle volte non vengono approvati per essere pubblicati. Questo fenomeno, in cui la pubblicazione di un lavoro è influenzata dai tipi di risultati ottenuti, prende il nome di *publication bias* [5].

I risultati negativi, però, rappresentano un'importante risorsa per il progresso scientifico. Da una parte si evita la ripetizione di esperimenti già effettuati e dall'altra parte i risultati ottenuti anche se negativi potrebbero essere la base di studi scientifici futuri. È evidente che pubblicare questi risultati possa contribuire in modo funzionale al progresso scientifico. Nonostante ciò, quando i ricercatori ottengono risultati negativi, tendono a non segnalare la propria ricerca e altri utilizzano stratagemmi per forzare l'esito positivo dei risultati contribuendo alla genesi della "cattiva scienza" [7].

### 2.3.2 La questione delle ritrattazioni

A partire dal nuovo millennio, si è riscontrato un aumento vertiginoso del numero di articoli ritrattati, soprattutto nelle più importanti riviste scientifiche internazionali. Si può notare dal grafico (figura 1) [11] che a partire dagli anni 2000 il numero di ritrattazioni subisce un'impennata e nel 2010 è aumentato di 20 volte rispetto agli anni '80.

Esistono vari motivi per cui un lavoro scientifico possa presentare degli errori. I ricercatori, essendo umani, possono sbagliare e non coscienti degli errori commessi, mandano i propri articoli in *submission*<sup>5</sup> alle riviste. L'ambito scientifico in cui questo fenomeno è maggiormente presente è la ricerca biomedica, dove la complessità e la variabilità di certi aspetti accrescono la probabilità di produrre risultati erronei. Nonostante questo, in uno studio condotto su un gruppo di ricercatori biomedici negli Stati Uniti nel 2012 [12] si attesta che il 9% degli scienziati indagati hanno ammesso di aver almeno una volta falsificato dei risultati. Si nota come la pressione istituzionale dettata dal paradigma *publish or perish* porti i ricercatori a commettere atti illeciti per rendere il proprio lavoro più "pubblicabile".

La causa delle ritrattazioni non è per forza sempre collegata ad atti di frode ma rimane certo che il fenomeno sia in parte causato dall'incentivazione dei ricercatori nel pubblicare il maggior numero di articolo possibili con l'obiettivo di far apparire il proprio nome nelle più prestigiose riviste scientifiche [10].

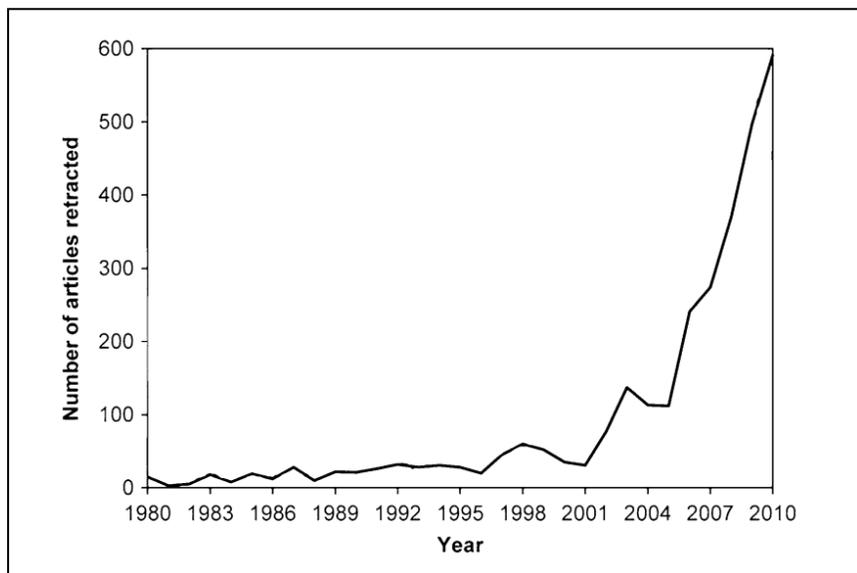


Figura 1: numero di articoli ritrattati nel tempo

<sup>5</sup> Con il termine *submission* si intende la sottomissione a valutazione tramite peer review di un articolo scientifico a una rivista

### 2.3.3 La crisi della riproducibilità

Nel mondo della ricerca e soprattutto nel mondo della ricerca medica/farmacologica riprodurre determinati esperimenti è un modo fondamentale per constatare l'effettiva validità di una scoperta scientifica. Per riproducibilità si intende la possibilità di ottenere gli stessi risultati di uno studio scientifico tramite l'applicazione dei dati e delle analisi riportate nella ricerca originale. Spesso, per scovare le false scoperte, l'unico modo davvero efficace è riprodurre gli esperimenti [7].

Quello che si è venuta a creare nel mondo della scienza è una vera e propria crisi della riproducibilità. Il *modus operandi* dei ricercatori, inaccurato, frettoloso e poco trasparente, porta alla produzione di studi poco affidabili culminando così nell'impossibilità di riprodurre le ricerche. In più la centralità del "paper" come nucleo di diffusione del sapere comporta la restrizione dei dati che vengono pubblicati, dati che risultano fondamentali nel momento in cui si vuole tentare la riproduzione di un esperimento [9]. La crisi coinvolge anche i metodi che si usano per determinare la significatività statistica di uno studio. Uno dei problemi rappresentati da questi indici statistici è la pratica del *p-hacking*<sup>6</sup> che consiste nell'omissione di quelle analisi dei dati che portano a un potenziale risultato negativo [13, 14].

## 2.4 Soluzioni open

Queste dinamiche che si sono consolidate nel tempo nel sistema di comunicazione scientifica mondiale risultano difficili da eradicare. Proporre metodi e procedimenti alternativi per risanare un modo di fare scienza che abbia come obiettivo primario il progresso scientifico è ritenuto più che mai fondamentale. L'Open Science è un fenomeno che a partire dal 21esimo secolo è in costante crescita e si ricollega fortemente alle innovazioni portate da Internet e dalle tecnologie connesse ad esso. La disponibilità di strumenti di condivisione e comunicazione ha mosso nella comunità scientifica un'intenzione molto forte, quella di rendere la ricerca accessibile a tutti. Le problematiche individuate nei paragrafi precedenti sono dannose sia per la comunità scientifica e per chi lavora nell'ambito della ricerca, ma anche più in generale per l'umanità. Una ricerca più accessibile e più trasparente permetterebbe a chiunque di interessarsi alle varie discipline e renderebbe la scienza più credibile ed affidabile anche agli occhi dei più scettici. In più, con la facilitazione dello scambio di informazioni è conseguenziale anche una velocizzazione del processo di scoperta [15].

### 2.4.1 Obiettivi dell'Open Science

Ciò a cui mira l'Open Science è rendere la Scienza un ambito più democratico e universalmente compreso. La volontà di ottenere più fiducia dalla collettività esterna all'attività

---

<sup>6</sup> Si chiama *p-hacking* perché va a influenzare il valore del *p-value*, indice che rappresenta la significatività statistica di un esperimento.

accademica ci conduce, dunque, a uno degli obbiettivi fondamentali dell'Open Science: l'affidabilità. Il tipo di ricerca promossa dal movimento open è caratterizzata dalla trasparenza, dall'accessibilità e dalla riproducibilità. Questi sono tutti elementi che permettono una migliore e più vasta circolazione della letteratura scientifica che di conseguenza contribuirebbe al superamento dell'era della post-verità in cui le fake news e il complottismo corrompono le menti dei più ingenui. Aprire la scienza significa anche condividere i risultati della ricerca che si ritrovano esclusi nel momento di stesura del *paper*. Avere a disposizione i dati, i materiali, il software completa il quadro della ricerca, rendendo tutto più comprensibile.

Più nello specifico, l'Open Science vuole attuare dei cambiamenti nel mondo delle pubblicazioni. Abbiamo visto come nel sistema attuale di divulgazione il blocco all'accesso alla letteratura scientifica non beneficia nessuno, tranne i proprietari delle riviste scientifiche che attraverso la privatizzazione dei contenuti traggono profitto mentre i metodi di valutazione, gestiti dagli editori, generano malumori nella comunità scientifica. Una delle caratteristiche centrali della filosofia *open* è che si applica a tutte le tappe del processo di ricerca, dalla definizione del problema iniziale all'operato finale. La trasparenza nelle fasi della ricerca è indispensabile per garantire la riproducibilità della ricerca. Stracke suggerisce come la filosofia del *Total Quality Management* si possa applicare anche all'Open Science (figura 2). Il TQM consiste nella distinzione tra potenziale, processo e risultati e nel loro continuo miglioramento in un ciclo iterativo. Queste tre fasi si influenzano a vicenda producendo un risultato sempre migliore. Lo stesso è applicabile all'Open Science: l'accesso ai risultati della progettazione, della fase di ricerca e della pubblicazione producono miglioramenti continui in ogni aspetto della ricerca [15].

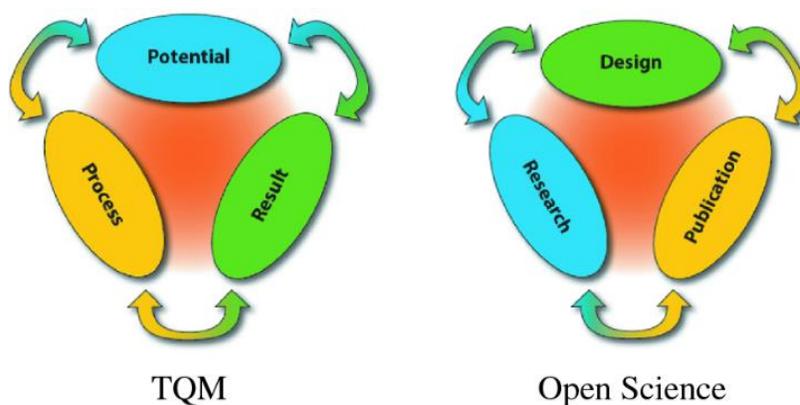


Figura 2

#### 2.4.2 I dati FAIR

L'*Open Data*, ovvero l'accesso libero ai dati della ricerca, è uno dei concetti subordinati al concetto più generico della Scienza Aperta. L'acronimo FAIR descrive come i dati devono essere per garantire la massima ottimizzazione della divulgazione del sapere. Nel mondo della

comunicazione scientifica è necessario implementare sistemi che incentivino la circolazione e il riutilizzo dei dati accademici. Nel 2014 durante un seminario (*Jointly Designing a Data FAIRport*) sono stati pensati e sviluppati i 4 principi fondamentali per i dati [16]:

- *Findable*: i dati devono poter essere facilmente ritrovabili
- *Accessible*: i dati devono poter essere accessibili
- *Interoperable*: i dati devono avere un formato che permetta la loro applicazione attraverso sistemi diversi da quelli dell'originale
- *Reusable*: i dati devono essere descritti in maniera chiara e comprensibile per permettere il loro riutilizzo

I principi FAIR non sono degli standard, ma delle linee guida che possono migliorare la qualità dei lavori accademici. Per far sì che i dati siano FAIR è importante avere una buona documentazione sui metodi, strumenti e procedimenti applicati, dei metadati che diano la capacità alle macchine di recuperare i dati in maniera automatica, usare degli identificatori persistenti<sup>7</sup> per intensificare l'interconnessione fra le risorse e, infine, utilizzare delle licenze per permettere il loro riutilizzo [17, 18].

### 2.4.3 Pratiche e strumentazioni

Un altro modo in cui possiamo intendere il movimento Open Science è come un insieme di metodologie che si contrappongono alle cattive pratiche adottate dai ricercatori condizionate dal mercato della letteratura scientifica mondiale. Per contrastare queste abitudini e quindi rendere accessibile la conoscenza in tutte le sue forme è necessario definire le modalità in cui farlo. Kramer e Bosman hanno sviluppato uno schema che mostra in maniera esplicativa le buone pratiche che si possono adottare a ciascun livello della ricerca per praticare l'Open Science (figura 3) [19]. Nel tempo sono state sviluppate strumentazioni e metodologie che aiutano i ricercatori a fare della ricerca che rispetti i principi FAIR.

Un modo per far sì che i propri sforzi nella ricerca non vengano chiusi dietro ad un abbonamento è di pubblicare il pre-print dell'articolo sui repository istituzionali. Un repository è un archivio di oggetti digitali che si possono accedere e scaricare gratuitamente. Di questi ne esistono di tutti i tipi, non solo istituzionali. Ci sono i repository di software, come *GitHub*, quelli per i dati come *Zenodo* o quelli per i protocolli come *MyExperiment*. Per ogni fase della ricerca ci sono implementazioni per rendere i risultati aperti per far sì che si arrivi al superamento dell'idea di articolo revisionato come unico prodotto scientifico di valore [9].

Per collaborare sui risultati della ricerca è necessario avere delle piattaforme di appoggio come, ad esempio, i Virtual Research Environment (VRE). Questi consentono non solo di condividere ma anche di generare discussioni e di collaborare tra le diverse comunità scientifiche. In questi

---

<sup>7</sup> Gli identificatori persistenti sono un tipo di codici identificativi attribuiti a oggetti digitali duraturi come DOI, ORCID, PURL, VIAF e altri. Per saperne di più: <https://www.openaire.eu/what-is-a-persistent-identifier>

contesti virtuali i ricercatori possono confrontarsi gli uni con gli altri sui dati, i software e le risorse caricate. Queste piattaforme sono rese disponibili anche da alcune biblioteche che creano dei VRE specifici per alcuni progetti. *Open Science Framework*<sup>8</sup> (OSF) è uno degli esempi di piattaforma collaborativa più noto.

Altro aspetto molto importante, vista soprattutto la centralità del paper come unità fondamentale del lavoro di uno scienziato, è poter lavorare a distanza e contemporaneamente sullo stesso documento. Per questo esistono piattaforme come *Overleaf* e *Authorea*, che consentono la scrittura collaborativa in tempo reale di un documento. Esistono anche strumenti come *Zotero*, software *open source* per la gestione dei riferimenti bibliografici che aiutano i ricercatori nell'organizzazione delle citazioni. Spesso i ricercatori vedono l'applicazione di queste pratiche e l'utilizzo di queste piattaforme come un intoppo e un impegno che porta via tempo al loro lavoro. Tuttavia, è importante evidenziare il fatto che queste strumentazioni sono tutt'altro che

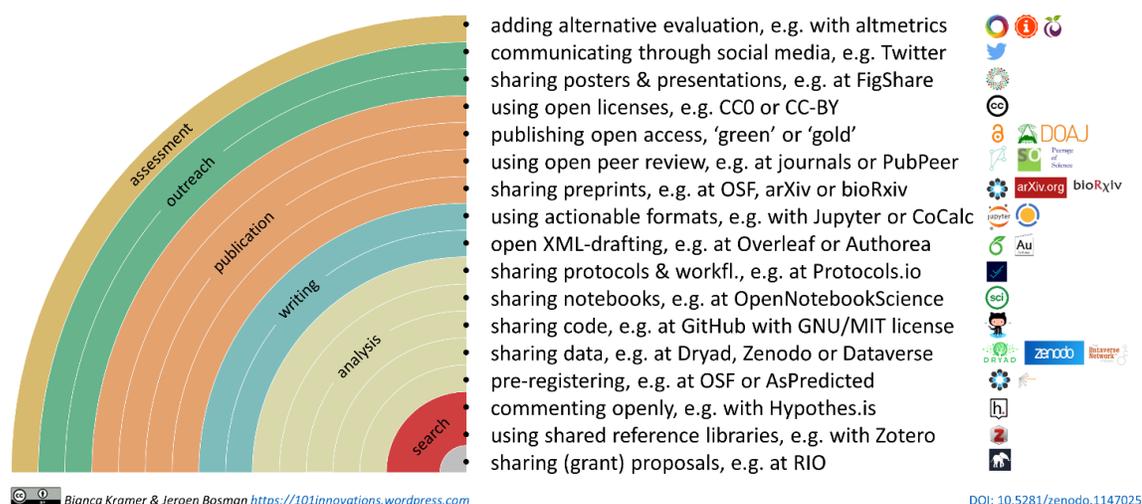


Figura 3: Rainbow of open science practices. Kramer & Bosman

limitanti, perché permettono una maggiore collaborazione e permettono di comunicare con chiarezza.

Le risorse per rispettare i principi FAIR devono poter essere riutilizzate. I ricercatori che pubblicano in repository Open Access sono invitati a utilizzare licenze aperte, come le licenze *Creative Commons*<sup>9</sup> che consentono il riuso dei dati secondo limiti impostati dall'autore. Le principali licenze CC sono sei e variano in base al livello di restrizioni e limiti stabiliti. Si riconosce in questo contesto anche l'utilità dei social media come luoghi in cui comunicare, condividere e divulgare la ricerca. Per concludere, il paradigma *open* incita all'uso di metriche di valutazione alternative con l'obiettivo di sottrarre il potere valutativo agli editori

<sup>8</sup><https://osf.io/>

<sup>9</sup> Per approfondire sulle licenze CC consultare il seguente link <https://creativecommons.org/licenses/>

commerciali. A un primo livello l'open peer review consente una valutazione della qualità del lavoro attraverso un approccio totalmente trasparente. Poi, a un secondo livello, sono stati discussi tanti modi per calcolare l'impatto attraverso diversi strumenti online, come ad esempio i salvataggi dei link, i *like* e le condivisioni [18].

#### 2.4.4 Benefici

Open Science non è soltanto una filosofia di fare scienza, ma si può anche vedere come la trasformazione che sta avvenendo nel tempo del modo di fare scienza e ricerca in seguito a una sempre maggiore globalizzazione e a un costante avanzamento tecnologico. Lo scopo è quello di far diventare l'apertura ai risultati della ricerca una norma. In questo cambiamento è anche fondamentale la collaborazione di enti governativi e istituzioni pubbliche, che attraverso la stesura di normative, possano cambiare il sistema di incentivazioni e prendere in mano il controllo della comunicazione scientifica. In questo contesto entrano in gioco le *policy*, che si confermano essere fondamentali per questo cambiamento [20].

È evidente che un approccio più trasparente ai dati e a tutti i risultati della ricerca possa produrre soltanto miglioramenti in termini di qualità. Sicuramente, i lavori svolti seguendo la filosofia open e attenendosi a produrre dati FAIR sono più attendibili. Condividendo tutte le fasi di una ricerca si ha una maggiore possibilità di riprodurlo e con una tale trasparenza si prevengono falsificazioni dei dati o altri tipi di frodi scientifiche. Inoltre, c'è una maggiore possibilità di essere pubblicati poiché anche le ricerche che producono risultati negativi, nonostante la mancanza di quelle caratteristiche che rendono un articolo appetibile alla pubblicazione, vengono prese in considerazione. Gli autori, di conseguenza, vengono anche più facilmente citati [21].

Universalmente tutti godrebbero di un approccio *open* in quanto questo si è dimostrato essere un acceleratore di scoperte. L'accessibilità ai dati e la loro massima divulgazione incrementano la potenzialità collaborativa tra comunità scientifiche anche appartenenti a realtà diverse. Il più esemplificativo dei casi di applicazione della filosofia *open* è il Progetto Genoma Umano che fu avviato nel 1990. L'ampissima e intensa condivisione dei dati ha permesso di decodificare il genoma in un lasso di tempo inferiore a 15 anni. Per permettere la massima concentrazione sulla collaborazione nei progetti di ricerca è necessario attuare delle modifiche nel sistema di incentivi e iniziare a premiare coloro che seguono le pratiche open science [20].

Infine, l'Open Science rappresenta un'opportunità per i singoli ricercatori di abbandonare il paradigma *publish or perish* e di lavorare in un ambiente più democratico e collaborativo. In più, genera vantaggi anche in termini professionali: una maggiore possibilità di pubblicare unita all'accesso libero ai lavori producono una crescita nel numero delle citazioni [21].

### 3 POLITICHE E INIZIATIVE

L'obiettivo dei sostenitori dell'Open Science è quello di raggiungere un'apertura tale al sapere accademico da rappresentare la totalità della produzione. È chiaro, che nonostante l'adeguarsi di molte istituzioni e di molti ricercatori alle pratiche *open*, la totale apertura ai risultati della ricerca è ancora lontana. Tuttavia, è evidente che all'interno del sistema si sta muovendo qualcosa e che si sta producendo un cambiamento, specialmente in Europa. I cambiamenti si possono notare nella volontà di varie istituzioni ed enti governativi di aderire a quelli che sono i principi dell'Open Science e definire delle procedure standard attraverso la stesura di policy. La Commissione Europea, portavoce di questa volontà di cambiamento, ha intrapreso un cammino verso la totale apertura della ricerca e negli ultimi cinque anni ha portato avanti varie iniziative al fine di promuovere i principi dell'Open Science. In questo capitolo si illustrano le politiche e alcune iniziative che hanno l'obiettivo di orientare la comunità scientifica verso un tipo di ricerca sostenibile e innovativo.

#### 3.1 Le politiche europee

Quando si parla di politiche per la Scienza Aperta si intendono quegli insiemi di procedimenti e strategie che servono a promuovere e a incentivare un modo di fare ricerca che rispetti i principi dell'Open Science. Se queste politiche, una volta, riguardavano soltanto l'accesso aperto alle pubblicazioni ora si estendono a tutte le pratiche dell'Open Science. Questi mandati possono essere emessi da enti di ricerca, enti finanziatori ed enti governativi e servono come base ai ricercatori per mettere in pratica i principi del paradigma *open* e quindi, anche, come incentivi che si scostano dal tradizionale approccio alle pubblicazioni [18].

La Commissione Europea, con l'obiettivo di ottimizzare la divulgazione della ricerca accademica prodotta attraverso finanziamenti pubblici ha sviluppato una politica olistica per la promozione delle buone pratiche dell'Open Science<sup>10</sup>. Queste politiche nascono con la speranza di creare nuovi incentivi per i ricercatori e di standardizzare le pratiche open in tutti i contesti della ricerca. Per l'Unione Europea la circolazione libera dei risultati della ricerca rappresenta una priorità e dunque vengono incoraggiate le istituzioni pubbliche delle varie nazioni a intraprendere un percorso per aderire a questa filosofia. La strategia si può riassumere in otto propositi descritti sul sito ufficiale della Commissione Europea:

- *Open Data*: la libera circolazione dei dati e la loro caratteristica FAIR come impostazione predefinita

---

<sup>10</sup> La piattaforma dedicata alle politiche Open Science dell'Unione Europea  
[https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en)

- *European Open Science Cloud (EOSC)*: un cloud europeo per l'archiviazione, la condivisione e il riuso di risorse digitali
- *New Generation Metrics*: lo sviluppo di nuovi indicatori per la valutazione e l'impatto della ricerca
- *Future of scholarly communication*: l'accesso libero alla letteratura accademica e l'incoraggiamento alla sua condivisione
- *Rewards*: il riconoscimento dell'applicazione di pratiche *open* nelle metriche di valutazione come incentivi
- *Research integrity & reproducibility of scientific results*: una ricerca più riproducibile e meno fraudolenta
- *Education and skills*: l'istruzione dei ricercatori alle pratiche open
- *Citizen science*: una ricerca che coinvolge e a cui possono contribuire tutti, non solo ricercatori

Il momento cruciale che ha funto da presa di coscienza sull'Open Science è stata la pubblicazione nel 2016 di *Open Innovation, Open Science, Open to the World* di C. Moedas, la visione dell'allora nuovo commissario europeo per la ricerca e l'innovazione. In questa pubblicazione viene esposta la visione del commissario: per la prima volta l'apertura ai risultati della ricerca vengono considerati come una priorità per la Commissione Europea [20]. Nel tempo sono stati sviluppati programmi per l'innovazione come *Horizon 2020*<sup>11</sup> e ora *Horizon Europe*<sup>12</sup> e sono state create infrastrutture tecnologiche che permettono la messa in pratica dell'Open Science. Alcune pratiche come l'Open Access alle pubblicazioni sono state rese obbligatorie per tutta quella ricerca prodotta tramite finanziamenti dell'Unione Europea.

### 3.1.1 Politiche nazionali

Le politiche europee sono necessarie come punto di partenza ai singoli stati per adattarsi alle pratiche open e renderle una norma. Tuttavia, si dimostra essere necessario attuare dei piani nazionali che, sulla scia della Commissione Europea, portino anche i singoli stati ad applicare i principi *open* con rigore e sistematicità. Alcuni paesi si sono attrezzati con delle politiche nazionali. Nella relazione bimestrale sul monitoraggio redatto da SPARC Europe<sup>13</sup> sono riportate le informazioni riguardo ai vari paesi europei e le relative politiche (se esistenti). Attualmente, gli stati membri dell'Unione Europea che possiedono dei piani nazionali che danno spazio alla Scienza Aperta sono il Belgio, il Cipro, la Repubblica Ceca, la Finlandia, la Francia, L'Irlanda, la Lituania, l'Olanda, la Slovacchia, la Slovenia, la Spagna e più recentemente l'Italia.

<sup>11</sup> <https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020>

<sup>12</sup> [https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe\\_en](https://ec.europa.eu/info/research-and-innovation/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en)

<sup>13</sup> L'ultima versione del report sulla situazione delle politiche in Europa di Sparc Europe si trova al seguente link <https://zenodo.org/record/4725817#.YMNhoPkzaUm>

Come paesi al di fuori dell'Unione che possiedono una politica nazionale per l'Open Access troviamo la Svizzera, la Norvegia, la Serbia e la Gran Bretagna. La maggior parte di queste politiche non si presentano come degli obblighi a cui ottemperare, bensì dei suggerimenti alle pratiche e riguardano sia l'Open Access delle pubblicazioni sia l'Open Data, quindi l'apertura ai dati. Una cosa da tenere presente è che per permettere una migliore collaborazione tra i paesi è necessario un allineamento a livello di regolamentazioni. Più le politiche sono simili tra di loro e danno indicazioni uniformi, più è facile lo scambio di idee e soprattutto la cooperazione tra ricercatori di differenti nazioni è agevolata. Un'altra cosa che risulta importante da inserire nelle politiche è una definita descrizione della tipologia di risultati e di dati che vanno condivisi. La chiarezza è un aspetto fondamentale per permettere ai ricercatori di lavorare nel migliore dei modi. La maggior parte delle politiche nazionali prese in considerazione nel report di SPARC presentano una descrizione accurata del tipo di output previsti. Inoltre, è importante anche incoraggiare i ricercatori a dare il giusto e legittimo riconoscimento agli autori dei lavori della ricerca, aspetto vitale per permettere un'attribuzione di merito corretta. Una questione importante ma assente in tutte le policy nazionali analizzate da SPARC, è promuovere i cosiddetti *data availability statements*, ovvero delle indicazioni sul dove e come poter accedere ai dati utilizzati in una determinata ricerca. Infatti, sarebbe utile incorporare all'interno delle proprie pubblicazioni delle dichiarazioni di questo tipo per aumentare la riproducibilità dei vari studi scientifici. Molti editori hanno incluso questo tipo di pratica all'interno delle loro policy, ma sembrano non essere presi in considerazione nelle politiche delle singole nazioni [22]. Infine, per fare in modo che la ricerca circoli nel migliore dei modi possibili le politiche sono tenute ad incoraggiare a rendere i dati riutilizzabili tramite l'uso di licenze aperte, alcune facendo riferimento a specifiche licenze come le licenze *Creative Commons*.

### 3.1.2 In Italia

Rispetto ad altri paesi europei, l'Italia è in netto ritardo per quanto riguarda l'adeguamento alle politiche per la Scienza Aperta e la promozione dell'accesso aperto ai risultati scientifici. Nonostante ciò, il 15 dicembre 2020 è stato approvato il Programma nazionale<sup>14</sup> per la ricerca che consiste in un piano di 7 anni che ha l'obiettivo di rivoluzionare alcuni aspetti della comunicazione scientifica e della ricerca italiana successivamente all'esperienza acquisita con la pandemia da Covid-19. Le ambizioni di questo programma prendono ispirazioni dalle linee d'azione presentate nel contesto di Horizon Europe. All'interno del programma è incluso il *Piano Nazionale per la Scienza Aperta*<sup>15</sup> in cui sono presentati gli obiettivi e le buone pratiche del paradigma *open*. Il piano si presenta strutturato secondo quattro assi di intervento:

- Le pubblicazioni scientifiche

---

<sup>14</sup> <https://www.mur.gov.it/it/aree-tematiche/ricerca/programmazione/programma-nazionale-la-ricerca>

<sup>15</sup> Alla sezione 6.2 del PNR: <https://www.mur.gov.it/sites/default/files/2021-01/Pnr2021-27.pdf>

- I dati della ricerca
- La valutazione della ricerca
- Il coinvolgimento dei ricercatori

L'adozione di un piano nazionale per la Scienza Aperta rappresenta un'importante presa di posizione da parte del ministero della ricerca italiano che dimostra il suo consenso ad un approccio alla ricerca che sia conforme alle proposizioni espresse dalla commissione europea. Nonostante la lentezza nel riconoscere l'Open Science come una priorità da parte del governo, in Italia già da alcuni anni esistono varie iniziative e associazioni che promuovono l'apertura ai risultati della ricerca, tra cui AISA<sup>16</sup> (Associazione Italiana per la promozione della Scienza Aperta), associazione nata a Trento che tenta di promuovere i principi della scienza aperta e

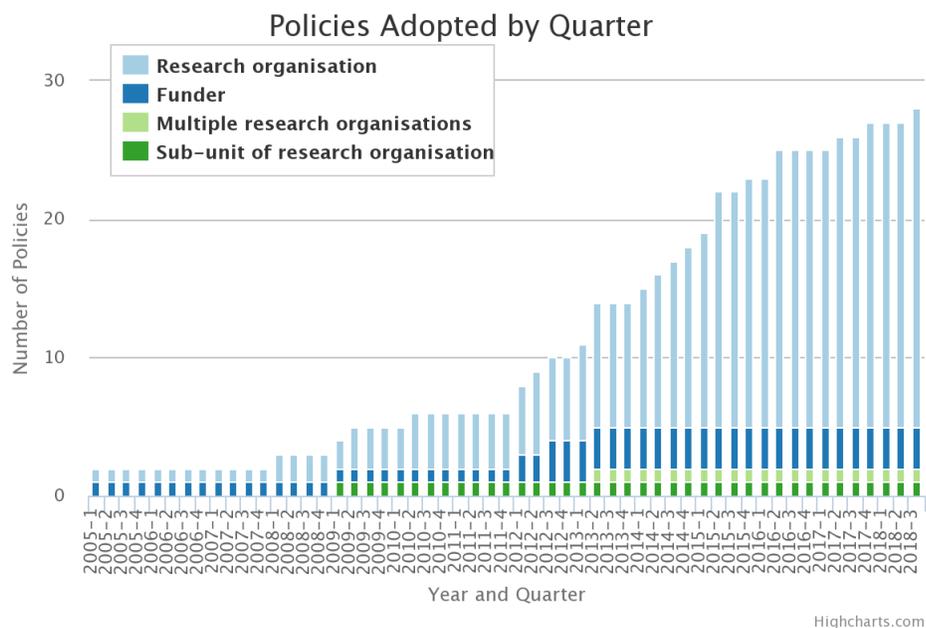


Figura 4: grafico sul numero di istituzioni che hanno una policy per l'Open Access in Italia. Fonte: ROARmap

IOSSG<sup>17</sup> (Italian Open Science Support Group), gruppo di lavoro transuniversitario nato con lo scopo di fornire supporto e strumentazione per la diffusione del paradigma *open*. È in fase di sviluppo il portale *open-science.it*, progetto organizzato dal CNR a cui ho attivamente partecipato, in particolare nella progettazione del catalogo che sarà ospitato sul sito. Sul portale saranno presenti materiali e novità riguardanti l'Open Science in Italia. Tra le altre cose, in Italia si sono mosse anche individualmente molte università, fondazioni e istituti che hanno preceduto il governo nella stesura di policy per l'accesso aperto. Il sito ROARmap<sup>18</sup>, *Registry of Open Access Repository Mandates and Policies*, che monitora il numero di istituzioni che hanno

<sup>16</sup> <https://aisa.sp.unipi.it/chi-siamo/>

<sup>17</sup> <https://sites.google.com/view/iossg/home?authuser=0>

<sup>18</sup> <http://roarmap.eprints.org/>

stabilito una politica sull'open access, conta 28 enti che hanno adottato delle policy per l'Open Access, 21 dei quali sono università. In questo contesto è interessante osservare dal grafico (figura 4) che si registra un aumento abbastanza marcato dal 2012 nel conteggio delle istituzioni in Italia che hanno definito una loro policy.

## 3.2 Iniziative

Per fare in modo che l'Open Science diventi facilmente applicabile e comprensibile a tutti nel tempo sono nate tante attività e iniziative, organizzate da enti ufficiali e non per la sua promozione e pianificazione. L'Europa, in particolare, riconosce l'importanza di costruire un tipo di ricerca, in quanto finanziata attraverso fondi pubblici, che possa tornare alla società. Queste iniziative servono ad accelerare un processo di cambiamento che vede l'accesso aperto al sapere accademico come una questione centrale e a creare uno spazio comune adatto per l'applicazione di queste pratiche.

### 3.2.1 Plan-S

Un'iniziativa che esprime questa volontà di cambiamento e soprattutto l'urgenza con la quale debba avvenire è quella portata avanti da cOAlition-S<sup>19</sup>, un gruppo nato il 4 settembre 2018, formato da vari enti finanziatori della ricerca con il sostegno della Commissione Europea e dell'*European Research Council*<sup>20</sup>, che nasce con lo scopo di rendere l'accesso aperto alle pubblicazioni scientifiche più in fretta possibile una realtà. Quest'iniziativa prende il nome di *Plan-S*, dove S sta per *speed*, ovvero "velocità" per sottolineare l'urgenza del cambiamento. Una prima versione di questo piano aveva come obiettivo che la ricerca finanziata dall'Europa fosse *open by default* entro gennaio del 2020, data che è stata spostata al 2021 per via delle reazioni contrastanti da parte degli editori e dei ricercatori, i quali hanno sollevato il problema riguardo alla libertà dei ricercatori di poter pubblicare dove e come preferiscono. In questo piano, infatti, viene richiesto che tutti i ricercatori che otterranno finanziamenti dai membri di Plan S pubblichino i propri articoli secondo dei procedimenti definiti e conformi ai principi dell'Open Access. Ci sono vari modi in cui i ricercatori, se vogliono aderire al piano, possono muoversi nel momento della pubblicazione. L'importante è che le riviste in cui vengono pubblicati gli articoli abbiano fatto, o stiano facendo, un processo per convertirsi a un modello Open Access, altrimenti devono permettere l'immediato accesso alle pubblicazioni tramite dei repository [23].

Plan-S si definisce su dieci principi:

---

<sup>19</sup> Su cOAlition-S: <https://www.coalition-s.org/about/>

<sup>20</sup> Informazioni su ERC: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/european-research-council>

- Gli autori o le loro istituzioni devono mantenere i diritti alle loro pubblicazioni, le quali saranno disponibili con licenza aperta, preferibilmente la *Creative Commons Attribution*<sup>21</sup> (CC BY) per soddisfare i requisiti definiti dalla Dichiarazione di Berlino<sup>22</sup> sull'accesso aperto alla letteratura scientifica.
- Gli enti finanziatori devono sviluppare criteri e requisiti solidi per i servizi che le riviste, le piattaforme e i repository Open Access devono fornire
- Nel caso in cui non esistessero riviste o piattaforme Open Access di alta qualità, i finanziatori sono tenuti a fornire incentivi per la loro creazione e il loro supporto
- Se previste, le spese per pubblicare in OA sono coperte dai finanziatori o dalle istituzioni di ricerca, non dai singoli ricercatori
- I finanziatori sostengono la varietà dei business models per le riviste e le piattaforme Open Access. Quando sono applicate tariffe per la pubblicazione in OA, queste devono essere proporzionate al servizio offerto e la struttura delle tariffe dev'essere trasparente, per permettere potenzialmente al mercato e ai finanziatori di standardizzarle
- I finanziatori incoraggiano gli enti governativi, le università, le istituzioni di ricerca, le biblioteche e le associazioni culturali a conformarsi alle strategie, politiche e pratiche per assicurare trasparenza
- I principi illustrati devono potersi applicare a tutti i tipi di pubblicazioni accademiche, è tuttavia comprensibile che la resa in accesso aperto come modalità per le monografie e i capitoli dei libri richieda più tempo e un processo separato
- I finanziatori non supportano la modalità "ibrida" di pubblicazione. Tuttavia, essendo un processo di transizione verso un totale accesso aperto in un tempo definito, i finanziatori possono contribuire al supporto economico sulla base di cosiddetti *transformative arrangements*<sup>23</sup>
- I finanziatori dovranno monitorare la fedeltà al piano dei beneficiari e sanzionare chi non risulta conforme alle disposizioni
- I finanziatori prenderanno decisioni sui finanziamenti in base al merito intrinseco del lavoro (senza considerare il canale di pubblicazione, il suo Impact Factor) o altre metriche di valutazione) o l'editore.

Osservando i principi è chiaro che uno dei principali obiettivi di Plan-S è di rivoluzionare il mondo dell'editoria scientifica e smantellare quel modello di mercato che prevede il pagamento di abbonamenti per accedere alla ricerca accademica. È anche evidente che a un primo sguardo

---

<sup>21</sup> La più libera delle licenze CC, permette l'uso dell'opera anche a fini commerciali:  
<https://creativecommons.org/licenses/by/4.0/>

<sup>22</sup> <https://openaccess.mpg.de/Berlin-Declaration>

<sup>23</sup> <https://www.coalition-s.org/faq/what-is-a-transformative-arrangement/>

le proposte possano risultare radicali e quindi meno appetibili a una parte della comunità scientifica.

### 3.2.2 EOSC

Abbiamo visto come per la Commissione Europea l'Open Science sia un argomento di interesse che occupa una posizione prioritaria nell'ottica di avanzare verso una tipologia di ricerca più efficiente e più collaborativa. È in questo contesto che nasce un progetto, l'*European Science Cloud* (EOSC)<sup>24</sup>. Si tratta di un ambiente virtuale dedito alla condivisione e al riutilizzo dei dati della ricerca che possa permettere ai ricercatori e alle comunità scientifiche di tante nazioni diverse di comunicare e collaborare. L'idea è nata dalla consapevolezza che il mondo della scienza e della ricerca stesse cambiando. La Commissione Europea già tra il 2014 e il 2015 aveva iniziato a parlare della *Science 2.0* e di come fronteggiare questa rivoluzione che stava spostando il mondo della ricerca verso una scienza *data-driven*<sup>25</sup>. Si era sviluppata l'idea che servisse una piattaforma che permettesse la gestione dei dati rispettando tre criteri fondamentali: la libera accessibilità, l'interoperabilità fra discipline e la salvaguardia della privacy. Il 12 giugno 2017 si è tenuto il primo *EOSC summit*, che ha visto le varie istituzioni riunite, in cui è stata definita la *EOSC Declaration* che racchiude tutte le proposizioni necessarie al fine di creare EOSC. Dal 2018 in poi lo sviluppo del Cloud subisce un'accelerazione per via di alcuni scandali riguardanti l'uso improprio di dati (scandalo di *Cambridge Analytica*, società che a inizio 2018 ha violato la privacy di molti utenti *Facebook* usando i loro dati personali per fare propaganda politica) che hanno messo in evidenza l'urgente necessità di avere un sistema per gestire i dati della ricerca prodotti dai fondi europei secondo termini rigorosamente definiti. Anche questo progetto è finalizzato alla trasformazione della comunicazione scientifica ponendo una particolare attenzione sull'importanza della gestione e del riutilizzo dei dati, i quali dovranno essere FAIR. L'intento è quello di avere una rete a cui la comunità scientifica possa appoggiarsi e dove la gestione e la ricerca dei dati sia facilitata grazie a servizi e strumenti offerti creando un ecosistema europeo dedicato allo sviluppo e alla promozione della Scienza Aperta [24].

### 3.3 Open Science e Covid-19

Nella situazione di emergenza globale venutasi a creare con la pandemia da Covid-19 le pratiche dell'Open Science sono state ampiamente attuate nella speranza di accelerare le procedure della ricerca e sviluppare un vaccino il prima possibile. In particolare, l'Open Access ha giocato un ruolo fondamentale in questo processo. L'unica possibilità di salvezza in un

---

<sup>24</sup> [https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en)

<sup>25</sup> Con il termine *data-driven* si intende un tipo di scienza che deriva conoscenza tramite l'analisi di grandi quantità di dati

momento di così grande disperazione è la ricerca, soprattutto quando ci si trova di fronte a un'incognita come il coronavirus.

### 3.3.1 Open Science applicata

Nell'ultimo decennio è stato ampiamente dimostrato che l'apertura ai risultati scientifici porti ad un'accelerazione della ricerca, specialmente in ambito biomedico. Durante la pandemia la Scienza Aperta è stata universalmente accolta per permettere una maggiore collaborazione nell'ambito della ricerca sul coronavirus. A gennaio 2020 è stata condivisa la prima sequenza del genoma del SARS-CoV-2 in un database Open Access al cui studio i ricercatori del mondo hanno potuto contribuire attivamente. Il 4 febbraio l'editoriale di *Nature* pubblica un articolo in cui si fa un appello ai ricercatori del mondo: “keep sharing, stay open”<sup>26</sup>. Nei mesi successivi la necessità di avere informazioni riguardanti al virus e di come combatterlo è sempre più urgente, così sono nate alcune iniziative che permettono alla comunità scientifica di ottenere informazioni in tempo reale e che hanno reso possibile una cooperazione a livello globale, eccone alcune elencate qui sotto:

- *Covid-19 data portal*: il portale realizzato dalla Commissione Europea che fornisce informazioni, novità e servizi per la condivisione dei dati sul virus
- *Global research Database on coronavirus disease*: creato dall'organizzazione mondiale della sanità che raccoglie tutte le più recenti scoperte e ricerche sul covid-19 che viene aggiornato giornalmente
- *CORD-19: Covid-19 Open Research Dataset*: iniziativa della Casa Bianca e altri gruppi di ricerca. È un dataset che contiene dati sull'andamento della pandemia
- *OpenAIRE COVID-19 Gateway*: piattaforma che raccoglie pubblicazioni, dati e altro sul virus da tutto il mondo lanciata da OpenAIRE
- *COVID-19 Molecular Structure and Therapeutics Hub*: repository che raccoglie dati, modelli e software per la modellistica molecolare del virus

Anche gli editori accademici hanno contribuito rendendo tutti gli articoli e le pubblicazioni sul Covid-19 ad accesso aperto. Tuttavia, l'Open Access a questi articoli è temporaneo. Infatti, le pubblicazioni sul coronavirus una volta conclusa l'emergenza globale saranno accessibili solo tramite il pagamento di un abbonamento. L'applicazione delle procedure dell'OS durante la pandemia porta a una riflessione sulla comunicazione scientifica attuale. È evidente che gli editori e i ricercatori siano consapevoli dei benefici che un approccio aperto porterebbe alla scienza, eppure si continua a perpetuare un sistema che attraverso l'uso di restrizioni commercializza il sapere [25, 26].

---

<sup>26</sup> <https://www.nature.com/articles/d41586-020-00307-x>

### 3.3.2 Aspetti da migliorare

Il potenziale beneficio dell'apertura alla ricerca in un momento così fragile per l'umanità è stato colto anche dai grandi editori come Elsevier, Springer Nature e altri, i quali hanno dimostrato la loro solidarietà aprendo l'accesso alle pubblicazioni relative al virus. Allo stesso tempo i ricercatori hanno spinto sulla pubblicazione dei loro preprint e parallelamente sono nate piattaforme per l'accessibilità ai dati del covid e alla loro analisi. Tuttavia, aprire i risultati della ricerca non sempre vuol dire applicare i principi dell'Open Science. Infatti, se da un lato l'OS è stata fondamentale per garantire una collaborazione a livello globale e per ottenere un vaccino in tempi record, alcune negligenze nell'applicazione corretta dei principi *open* sono state registrate. Ad esempio, un problema causato dalla mancata preregistrazione degli studi<sup>27</sup>, fase prevista dalle pratiche dell'OS, è stato lo spreco di sforzi impiegati in studi già esistenti. Nel processo di pubblicazione, invece, sono stati registrati molti casi in cui sono state attuate scorciatoie, specialmente nella fase di revisione, il che ha contribuito alla divulgazione di studi ancora non revisionati. Com'è stato illustrato nel primo capitolo, la fase di peer-review, anche se spesso lenta, è fondamentale per evitare casi di ritrattazione. Applicando soluzioni come l'Open Peer Review, che rende il processo di valutazione di una pubblicazione totalmente trasparente, si garantirebbero rigore e affidabilità e una velocizzazione sarebbe possibile senza dover sacrificare la qualità della valutazione. Poi, la comunicazione con il pubblico riguardo alle attività scientifiche nella ricerca sul coronavirus è stata gravemente afflitta da un uso improprio dei preprint. Infatti, se da un lato questi sono utilissimi per evitare duplicazioni e permettere di ottenere riscontri dalla comunità scientifica per migliorare il lavoro in corso, d'altra parte potrebbero contenere potenziali errori ed essere utilizzati dai media per diffondere notizie non ancora verificate. Si è ampiamente visto come potenziali studi non revisionati siano stati sfruttati da giornali e notiziari causando panico e confusione tra i cittadini. L'Open Science, incoraggia la pubblicazione dei preprint, ma è importante che si consolidi anche un uso rigoroso e cauto di questi. È risaputo che il giornalismo si affida alle comunicazioni stampa delle istituzioni di ricerca come fonti sulle notizie di ambito scientifico. Quindi è importante sottolineare da una parte la responsabilità che le istituzioni hanno nel fornire informazioni accurate, e dall'altra la necessità di un giornalismo più onesto che abbia come obiettivo quello di informare il pubblico e non di suscitare scandalo [27].

---

<sup>27</sup> Per preregistrazione si intende la pubblicazione degli obiettivi e delle metodologie che si andranno ad applicare nello studio prima che questo sia stato iniziato concretamente

## 4 I CAMBIAMENTI CONCRETI

Abbiamo visto che ad un livello ufficiale l'Open Science è stata riconosciuta come un'opportunità per il mondo di cambiare un sistema di comunicazione scientifica che permette la mercificazione del sapere accademico. Alcune istituzioni e molti enti governativi hanno aderito ai principi dell'OS tramite la stesura di policy e l'avvio di varie iniziative. Ma se a livello interno si sta vivendo un leggero cambiamento, a livello concreto che cos'è cambiato? Questo capitolo tenta di dare una risposta a questa domanda prendendo in considerazione alcuni parametri e disegnando un quadro generale della direzione in cui l'Open Science si sta sviluppando facendo un confronto con un periodo passato. Vedremo se le varie politiche e iniziative dedicate alla Scienza Aperta attuate dagli enti ufficiali stiano effettivamente avendo un effetto sul mondo della ricerca.

### 4.1 Open Access nel tempo

L'Open Science è un concetto ombrello che comprende sotto di sé molti altri concetti fondamentali. Quello che ha gettato le basi per lo sviluppo dei principi *open* è sicuramente l'Open Access. È un concetto che nasce più di un decennio prima dell'OS in un periodo di profonda crisi della letteratura accademica. Si è già accennato che dal '75 in poi i prezzi degli abbonamenti delle riviste accademiche sono aumentati sproporzionatamente. Negli anni '90 con l'avvento di Internet è nata la consapevolezza che la comunicazione potesse avvenire in maniera diretta e senza costi di produzione. Da una parte i bibliotecari, stufi di pagare eccessive quantità di denaro per gli abbonamenti alle riviste, unendo le proprie forze creano la *Scholarly Publishing and Academic Resources Coalition* (SPARC)<sup>28</sup> nel 1997, una coalizione che tutt'oggi promuove modi alternativi per fare divulgazione scientifica spingendo in particolar modo sull'Open Access. D'altra parte, invece, nascono le prime riviste ad accesso libero e i primi repository per l'auto-archiviazione, come arXiv.org<sup>29</sup> nel 1991, un archivio dedicato ai preprint di materia scientifica. Queste iniziative sono nate individualmente ma con un obiettivo simile, creare un tipo di divulgazione accademica che fosse democratica e aperta. Insieme queste iniziative confluiscono in quello che è il movimento Open Access. Come movimento possiamo collocare la sua nascita a febbraio del 2002, quando è stata pubblicata l'iniziativa di Budapest per l'accesso aperto, una dichiarazione al cui interno sono esposti i principi e gli obiettivi del movimento OA. Da questo momento in poi l'OA diventa una strada intrapresa da sempre più riviste, istituzioni e ricercatori [4].

L'Open Access si concentra principalmente sulle pubblicazioni. L'intento è quello di rendere liberamente e gratuitamente accessibile la letteratura accademica, eliminando quelle che sono

---

<sup>28</sup> <https://sparceurope.org/who-we-are/about-us/>

<sup>29</sup> <https://arxiv.org/>

le barriere poste dal tradizionale sistema di comunicazione scientifica. Per barriere possiamo intendere le restrizioni economiche, che limitano il numero di lettori e quindi anche l'impatto di un autore oppure i diritti d'autore, che riducono la possibilità di riutilizzo e di conseguenza anche l'efficacia di un lavoro di ricerca. L'Open Access promuove un accesso alle pubblicazioni gratuito e la possibilità di un libero riutilizzo dei contenuti [28].

#### 4.1.1 Le vie dell'Open Access

Per fare Open Access i ricercatori possono scegliere fra due modalità. Il Gold Open Access è la modalità che prevede il pagamento da parte dell'autore (di solito pagato dall'istituzione per cui lavora) per rendere l'accesso del proprio articolo aperto. In questo caso l'articolo una volta pubblicato dall'editore sarà liberamente accessibile senza barriere. Il vantaggio più grande in questo caso è che il *copyright* non viene ceduto all'editore ma viene mantenuto dall'autore dell'articolo, il quale può decidere la licenza da applicare. Il Green Open Access, invece, è la modalità che prevede l'auto-archiviazione. In questo caso si deposita il preprint, ovvero la versione che ancora non è stata sottoposta a peer-review, nei repository istituzionali. In questo caso il vantaggio è che non c'è bisogno di pagare, però spesso ci sono dei periodi di embargo che possono variare dai 6 ai 12 mesi [29].

Per quanto riguarda le riviste che permettono la via d'oro ne esistono di due tipi. Ci sono le riviste che hanno come unica modalità di pubblicazione quella ad accesso aperto, alcuni degli esempi più rappresentativi sono le sette riviste curate dalla *Public Library of Science* (PLOS)<sup>30</sup>. Poi ci sono gli ibridi, il cui modello di pubblicazione prende il nome di *Hybrid OA*, che sono riviste che hanno come modalità standard il pagamento degli articoli che pubblicano, ma che hanno alcune opzioni di Open Access [30].

#### 4.1.2 Sviluppi nel mondo delle pubblicazioni

Ora che abbiamo parlato di come funziona l'Open Access e dei modi in cui viene applicato, vediamo come questo abbia influenzato il modo di pubblicare. Innanzitutto, osserviamo con quale andamento il numero di riviste Open Access sia cambiato dal 2005. I dati sono stati raccolti attraverso la *Directory of Open Access Journals*<sup>31</sup> (DOAJ), il repertorio di riviste Open Access acceduto sia nel presente che nelle sue forme precedenti via *Internet Archive*<sup>32</sup>. Ad oggi DOAJ conta 16.488 riviste<sup>33</sup>. Osservando il grafico (figura 5) la prima cosa che si nota è sicuramente un calo improvviso nel 2016. Questo calo è dovuto a un evento avvenuto nel maggio di quell'anno, che ha visto la rimozione di circa 3000 riviste in seguito a un progetto del sito che prevedeva il rifacimento della domanda di inclusione. Le riviste rimosse sono quelle

---

<sup>30</sup> Per saperne di più su PLOS: <https://plos.org/about/>

<sup>31</sup> <https://doaj.org/>

<sup>32</sup> <https://archive.org/>

<sup>33</sup> Informazione reperita il 16 giugno 2021

che non hanno fatto domanda entro la scadenza prevista e sono quindi state eliminate dall'indice. Nonostante ciò, possiamo notare un incremento abbastanza costante negli anni. Specialmente tra il 2016 e il 2021 annualmente si contano circa tra i 1500 e i 2500 giornali in più. Comunque, DOAJ oggi contiene circa 10 volte il numero di riviste rispetto al 2005, il che è una dimostrazione di quanto l'OA, in particolare il Gold OA, sia diventato un modo molto utilizzato

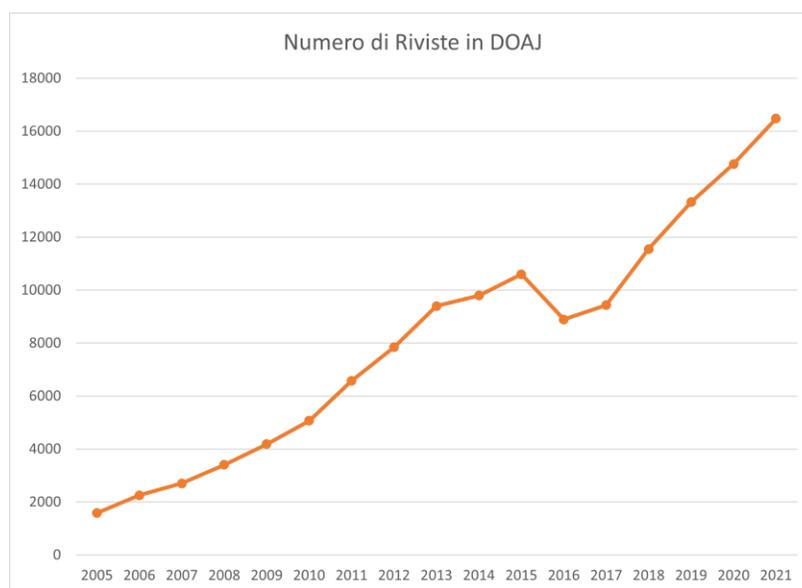


Figura 5: numero di riviste OA nel tempo

per pubblicare gli articoli. Le riviste per essere indicizzate in DOAJ devono essere di alta qualità, quindi devono essere riviste che sottomettono gli articoli alla revisione tra pari.

Per fare un ulteriore confronto tra come l'Open Access sia affrontato oggi rispetto a dieci anni fa, andiamo ad analizzare alcuni dati raccolti nell'ambito delle pubblicazioni. Per fare questo è necessario introdurre un ulteriore modello di pubblicazione in OA: il *Bronze OA*, che è una modalità di pubblicazione che consente di visualizzare il testo in modo gratuito ma a cui non sono associate licenze di riutilizzo. Per quanto riguarda gli articoli pubblicati in modalità standard, quindi in riviste a pagamento con *paywall*, la definizione che viene utilizzata nel mondo dell'OA è *Toll Access*. I dati sono stati raccolti dal database *Scopus*<sup>34</sup> attraverso il metodo di ricerca avanzato offerto dal sito che prevede l'utilizzo di *query*<sup>35</sup>. *Scopus* è un database che raccoglie un vasto insieme di *abstract* e citazioni di articoli accademici ed è messo a disposizione dalla rivista *Elsevier*. Per restringere il campo ho preso in considerazione solo gli oggetti che corrispondono a un determinato tipo di documento, l'articolo. Come prima cosa prendiamo in considerazione la quantità di articoli pubblicati in *Open Access* rispetto a quelli pubblicati in *Toll Access* in determinati anni. Nell'istogramma in basso (figura 6), in cui ho

<sup>34</sup> <https://www.scopus.com/>

<sup>35</sup> i dati sono stati raccolti il 18 giugno 2021

inserito i dati per le due categorie nel tempo a intervalli di 5 anni, possiamo osservare come la percentuale di articoli in Open Access sia aumentata particolarmente rispetto al 2005. Se prima la percentuale di articoli pubblicati in accesso aperto non arrivava nemmeno al 20%, ora (nel 2020) le pubblicazioni in OA, rappresentano quasi la metà della letteratura scientifica

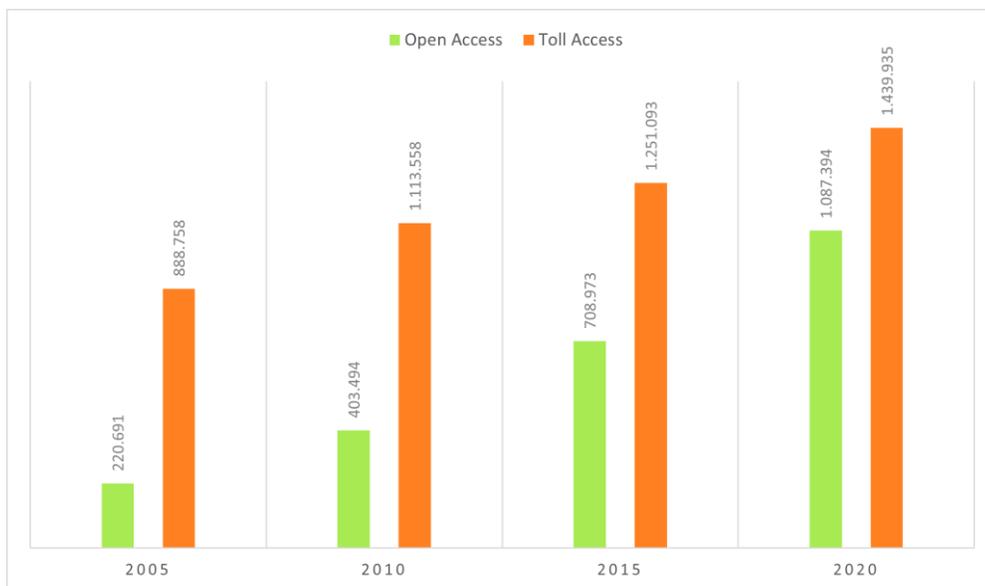


Figura 6: numero di articoli pubblicati in Open Access e numero di articoli pubblicati in Toll Access nei rispettivi anni

complessiva. Quindi possiamo dire che, come modalità di pubblicazione, l'accesso aperto sta prendendo sempre più piede.

Un altro aspetto che ho voluto analizzare è l'utilizzo dei vari modelli di pubblicazione OA per avere un'idea di quale fossero quelli più gettonati. La prima cosa che salta all'occhio dal grafico (figura 7) è sicuramente che il *Green OA* ha dominato di gran lunga fino al 2015. Nel 2020,

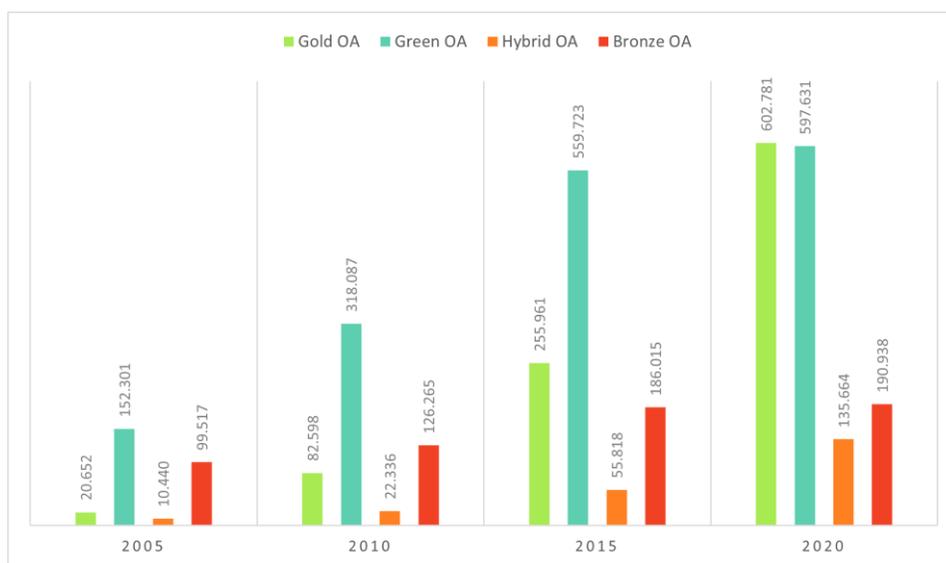


Figura 7: numero di articoli pubblicati nelle varie modalità quell'anno

invece, il *Gold Open Access*, che nel 2005 rappresentava una parte davvero minima degli articoli pubblicati in quell'anno, supera per poco il *Green OA*. Questo dato è significativo, soprattutto, se lo affianchiamo ai dati visti prima, ovvero il numero di riviste Open Access. La modalità *gold*, infatti, prevede la pubblicazione nelle riviste ad accesso totalmente aperto e abbiamo visto, infatti, come il numero di queste si sia innalzato particolarmente tra il 2015 e il 2020. Quindi, possiamo dedurre, che l'incremento di articoli pubblicati in *Gold OA* nell'ultimo anno sia in parte riconducibile all'aumento di riviste Open Access (e viceversa). Per quanto riguarda il modello *Bronze OA*, vediamo come in percentuale nel 2005 rappresentava una parte significativa dell'insieme di articoli analizzati, mentre nel 2021, si tratta di un numero molto meno importante. Il modello ibrido è, invece, quello che in percentuale rispetto agli altri è rimasto più costante, in tutti gli anni analizzati è stato quello meno utilizzato e con numeri decisamente più piccoli rispetto agli altri. Possiamo dire, quindi, che il *Green OA* sia stato e continua ad essere una modalità molto utilizzata dai ricercatori per la pubblicazione dei propri articoli, mentre il *Gold OA* è un modello che esiste già da molto, ma che ha iniziato a prendere maggiormente piede in tempi recenti.

Visti i successi del *Green OA* che continua ad essere tutt'ora la via scelta da molti ricercatori, per rendere apertamente accessibili gli articoli, è interessante anche esaminare l'andamento dei numeri di *repository* esistenti. I dati sono stati raccolti dal sito OpenDOAR<sup>36</sup>, una *directory* che contiene tutta una serie di *repository* ad accesso aperto, in particolare nella sezione dedicata alle

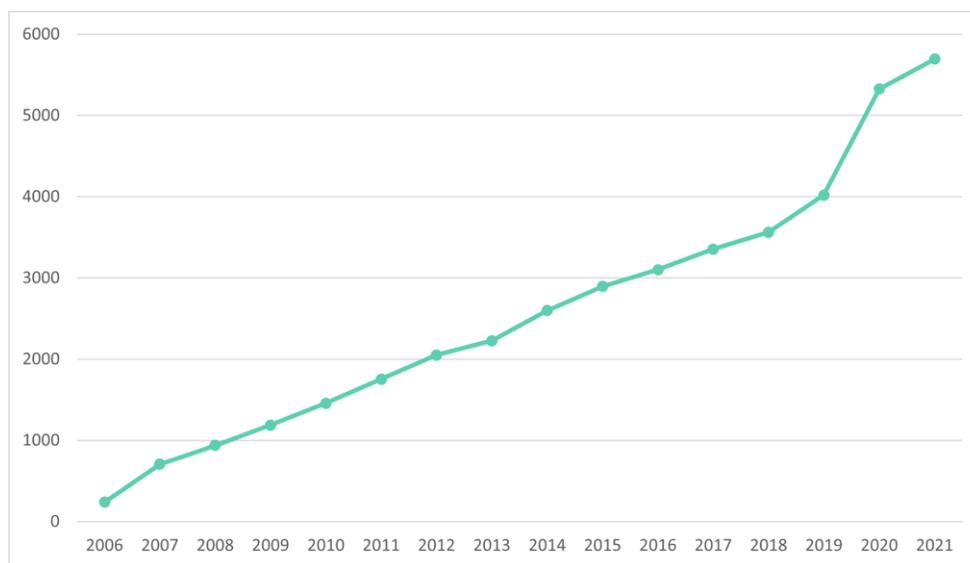


Figura 8: crescita del numero di repository su OpenDOAR

statistiche del sito<sup>37</sup>. Alla figura 8 si osserva la crescita del numero di repository all'interno di OpenDOAR. Si registra una crescita abbastanza costante con un aumento più marcato tra il 2019

<sup>36</sup> <https://v2.sherpa.ac.uk/pendoar/>

<sup>37</sup> Dati raccolti da OpenDOAR in data 18 giugno 2021

e il 2020. Quindi, è possibile dire, che, come nel caso delle riviste OA anche il mondo dei repository non si è mai fermato ma al contrario continua a crescere. Rispetto al 2006 nel 2021 OpenDOAR conta un totale che è quasi 24 volte più grande. Sul sito di OpenDOAR sono registrati anche altri aspetti interessanti, come ad esempio i paesi con più repository OA. L'Italia in questa classifica si posiziona dodicesima, con 141 repository, mentre al primo posto troviamo gli Stati Uniti con ben 907 repository. Altra informazione interessante è rappresentata dalla percentuale di ambiti disciplinari che coprono i repository. Più della metà degli archivi contenuti in OpenDOAR sono multidisciplinari, invece, come disciplina in sé quella che conta più repository è la medicina con ben 539 repository ad oggi.

Se da un lato il numero di archivi Open Access nel mondo è in costante crescita ora andiamo ad osservare le attività che si registrano al loro interno. Per avere un'idea di come le attività di deposito negli archivi OA si stiano evolvendo nel tempo, ho deciso di prendere come esempio uno degli archivi più conosciuti: arXiv, uno dei primi repository OA mai creati, attivo dal 1991 e dedicato alle pubblicazioni scientifiche. Come dato, andiamo ad osservare, il numero di preprint annui che vengono inseriti nell'archivio. Nel 2020 sono stati registrati più di 178mila

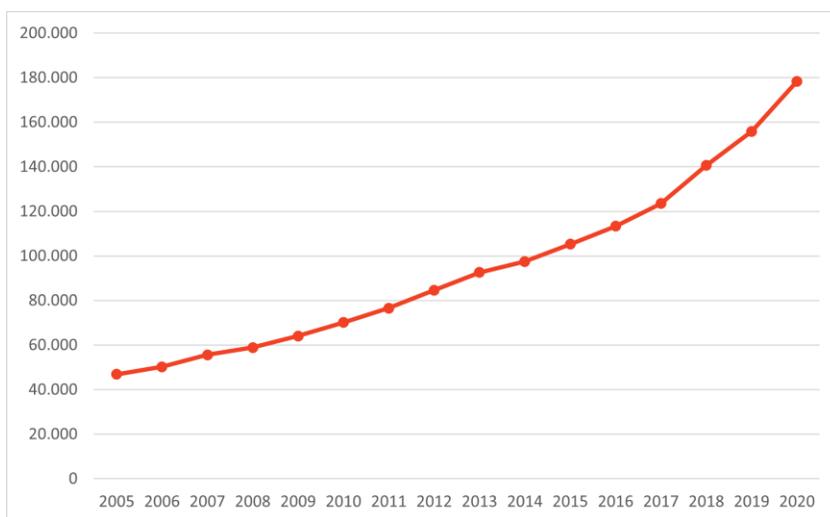


Figura 9: andamento attività di ArXiv

documenti mentre nel 2005 ne sono stati registrati 46mila. Quindi possiamo dire che l'attività di pubblicazione in arXiv ad oggi, rispetto a 15 anni fa, si sia quasi quadruplicata. Anche qui, dunque, si registra una costante crescita nel numero di pubblicazioni annue e in questo caso si registra un aumento più marcato dal 2017 in poi arrivando a un aumento del 14% tra il 2019 e il 2020.

Sempre nel contesto dei repository si osserva una cosa interessante in due archivi specialistici:

bioRxiv<sup>38</sup>, un archivio specializzato in biologia e scienze della vita e attivo dal 2013 e medRxiv<sup>39</sup>, un archivio specializzato in medicina e scienze della salute attivo dal 2019. I vari

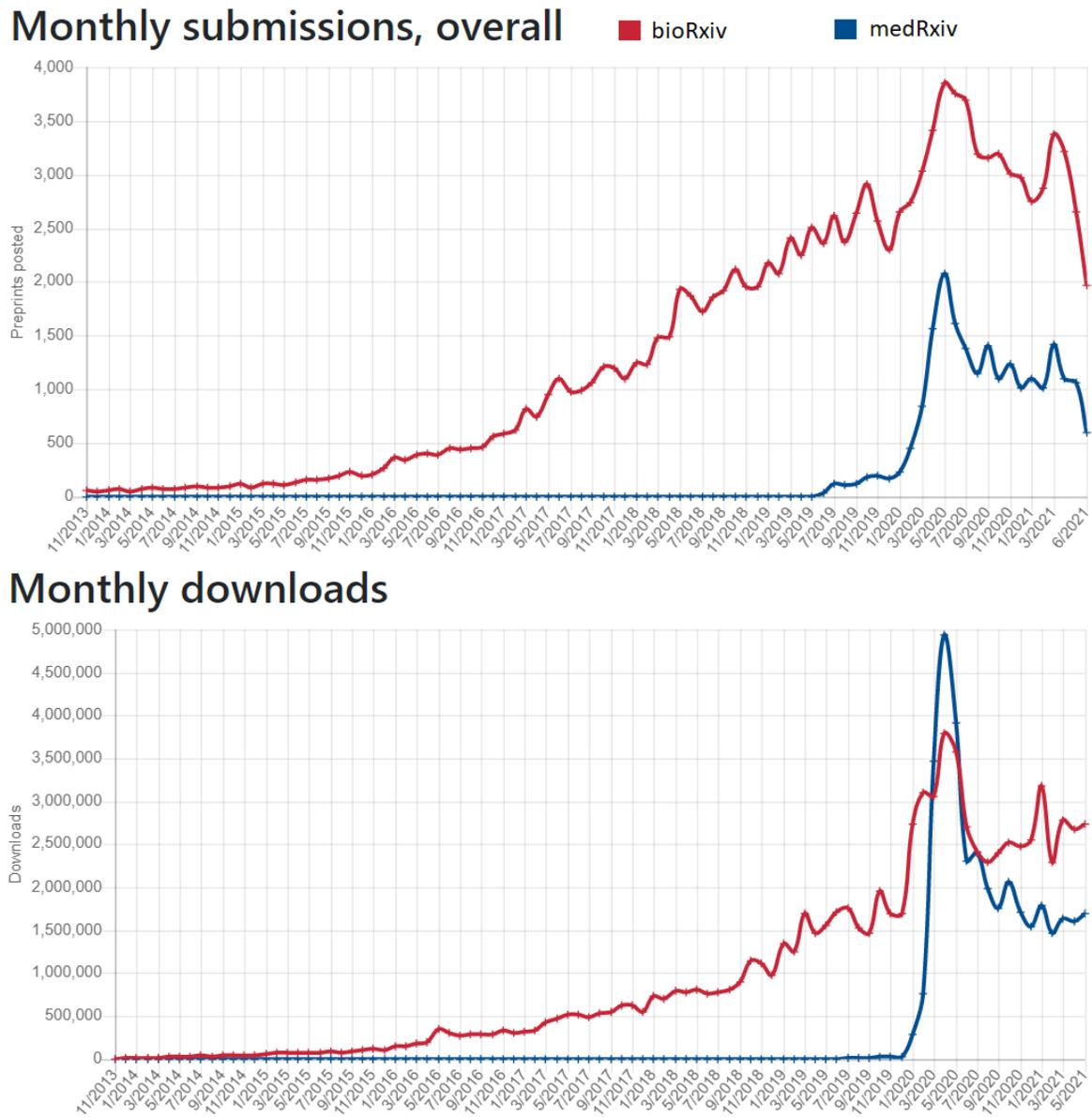


Figura 10: andamento delle attività di bioRxiv e medRxiv

paper che troviamo all'interno di questi due archivi sono indicizzati sul sito RXivist<sup>40</sup>, il quale riporta i dati sul loro andamento delle attività e questo mi sono sembrati degni di nota. I due grafici presenti su questa pagina raffigurano l'andamento mese per mese dei due archivi, il primo dal punto di vista delle nuove sottomissioni e il secondo dal punto di vista delle attività

<sup>38</sup> <https://www.biorxiv.org/>

<sup>39</sup> <https://www.medrxiv.org/>

<sup>40</sup> <https://rxivist.org/>

di *download* dal sito. Salta subito all'occhio un'improvvisa impennata da marzo 2020 in poi con un picco molto acuto nel mese di maggio. È evidente che ci troviamo nei primi mesi di pandemia globale e le numerose sottomissioni insieme ai tantissimi *download* rappresentano l'ondata di articoli Open Access pubblicati e scaricati in quei mesi. Questa emergenza globale ha significato per la comunità scientifica un urgente bisogno di informazioni sul virus che si è dunque tradotto in un'intensificazione delle attività di ricerca e soprattutto di quella Open Access. L'OA è sicuramente stato uno, insieme ad altri tanti aspetti, delle ancore di salvezza per la ricerca durante la pandemia.

## 4.2 L'Open data nel tempo

Introduciamo ora un altro dei pilastri portanti della Open Science, l'Open Data. La ricerca come la intendiamo nell'era contemporanea viene fatta grazie alla raccolta, l'analisi, il riutilizzo dei dati. Con l'avanzamento tecnologico portato dai calcolatori moderni, la potenza di calcolo è diventata tale da generare un incremento importante nella quantità di dati che si possono trattare e di conseguenza è nata anche la necessità di strumenti per gestirli. I dati rappresentano una grande risorsa per la scienza e saper sfruttarli al massimo potenziale è importante. Per avere una visione più chiara è necessario fare una distinzione. Quello di cui si parla in questa tesi è più precisamente l'Open Science Data, diverso dall'Open Government Science che è un movimento che nasce principalmente con l'obiettivo di rendere pubblici i dati prodotti da enti governativi e amministrativi. L'Open Science Data, invece, riguarda specificamente i dati prodotti dalla ricerca a cui d'ora in poi mi riferirò come Open Data (OD) [31]. È in questo contesto, infatti, che la gestione dei dati acquisisce una maggiore importanza. Secondo le linee guida della Commissione Europea, per una corretta applicazione dei principi dell'Open Research Data è necessario stilare un *Data Management Plan*, ovvero un piano che descriva l'intero ciclo di produzione dei dati. Questo piano permette da un lato ai ricercatori di avere un'organizzazione più precisa dei propri dati anche durante lo svolgimento della ricerca e da un altro di mettere a disposizione dei dati che, una volta condivisi con il resto della comunità, siano comprensibili e soprattutto riutilizzabili. La gestione dei dati è talmente centrale nel mondo dell'OD da aver generato un nuovo lavoro: il *data steward*, una figura lavorativa che ha il compito di controllare la qualità e l'organizzazione dell'insieme di dati prodotti [32].

Uno degli aspetti che l'OD pone al centro dei propri scopi è la possibilità di riutilizzo. Poter accedere ai dati non significa per forza poterli riutilizzare. È necessario che i dati rispettino alcune caratteristiche. Nel primo capitolo si è parlato di come l'acronimo FAIR (*Findable, Accessible, Interoperable, Reusable*) rappresenti la forma ottimale in cui i dati si possono presentare. L'utilizzo di licenze aperte è fondamentale, proprio per permettere il riuso dei dati e incoraggiare alla riproduzione degli studi [33].

Anche l'Open Data rappresenta uno degli obiettivi fondamentali della Commissione Europea ed è parte delle politiche per l'Open Science<sup>41</sup>. Nel progetto *Horizon 2020* i ricercatori che pubblicano degli articoli dovrebbero includere dei riferimenti ai *repository* in cui si possono trovare i dati che hanno portato ai risultati di un lavoro di ricerca. Ci sono comunque dei casi in cui i dati non vanno resi pubblici, ovvero quando questi comprendono dati sensibili o dati confidenziali. Il principio per l'Open Science, infatti, recita “*as open as possible, as closed as necessary*” proprio per sottolineare il fatto che quando si ha la possibilità è molto importante condividere i dati, ma che non sempre è lecito.

#### 4.2.1 Sviluppi nei repository di dati

Nel paragrafo precedente abbiamo potuto osservare come il mondo delle pubblicazioni OA si stia nel tempo evolvendo. Nell'ambito della letteratura accademica, anche se non interamente, una gran parte delle pubblicazioni sono ad accesso aperto. Andiamo ora a vedere come il mondo dei dati aperti si stia evolvendo nel tempo. Il lasso di tempo analizzato nell'ambito dell'Open Data è più corto perché gli strumenti e i siti da cui sono stati ricavati i dati sono nati in tempi più recenti.

Come parametro per la comprensione dell'andamento di attività nell'ambito dell'OD ho analizzato dei dati presi da Zenodo<sup>42</sup>, archivio *open* per la condivisione dei dati della ricerca lanciato nel 2013, gestito da CERN e parte del progetto OpenAIRE<sup>43</sup>. Per la raccolta di queste informazioni ho utilizzato delle *query* che permettono la ricerca avanzata<sup>44</sup>. In particolare, ho

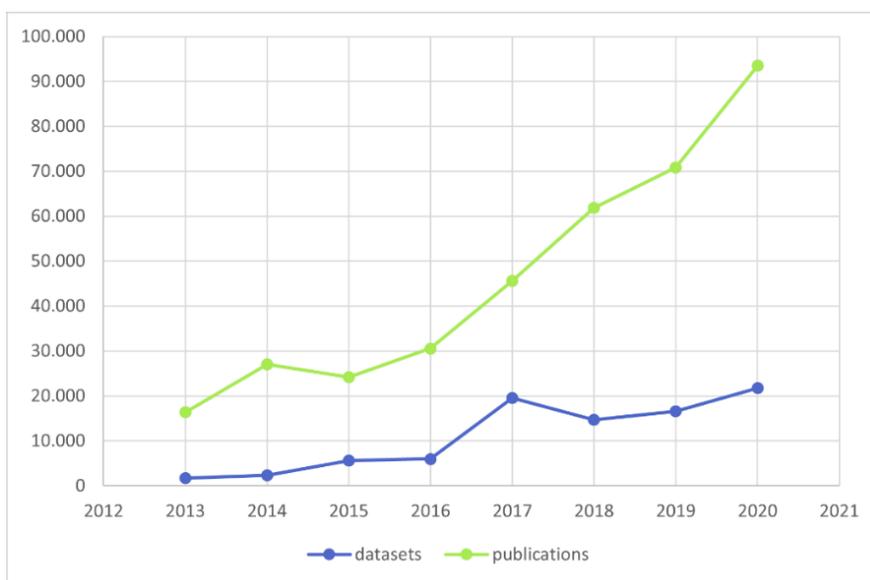


Figura 11: pubblicazioni annue di dataset e articoli accademici su Zenodo

<sup>41</sup> <https://open-research-europe.ec.europa.eu/for-authors/data-guidelines>

<sup>42</sup> <https://zenodo.org/>

<sup>43</sup> <https://www.openaire.eu/>

<sup>44</sup> Dati raccolti dal sito il 20 giugno 2021

cercato il numero di dataset e pubblicazioni che sono stati pubblicati annualmente sulla piattaforma, ovviamente prendendo in considerazione solo le risorse ad accesso aperto.

Innanzitutto, osservando la figura 11 è lampante la differenza nella quantità di risorse per ciascuna categoria. Le pubblicazioni viaggiano sin dall'apertura del sito a un numero parecchio più grande raggiungendo un picco nell'ultimo anno con un numero di pubblicazioni che è quasi cinque volte quello dei dataset. Nonostante ciò, si può comunque osservare un incremento nel numero dei dataset, soprattutto fra il 2016 e il 2017, in cui si è registrato un aumento di oltre 200%. In questo caso non ci troviamo di fronte a una crescita costante, vediamo che successivamente all'impennata nel 2017 c'è stato un lieve crollo, che poi negli anni seguenti riprende a crescere. Il numero di pubblicazioni, invece, subisce una crescita molto più rapida, specialmente nell'ultimo anno.

Prendiamo ora in considerazione un altro repository di dati: il *Dataverse*<sup>45</sup>. Questo è un repository aperto e gratuito in cui i ricercatori di qualsiasi disciplina possono condividere, citare, accedere e gestire i set di dati. Questa volta, però osserviamo l'andamento della totalità di dataset

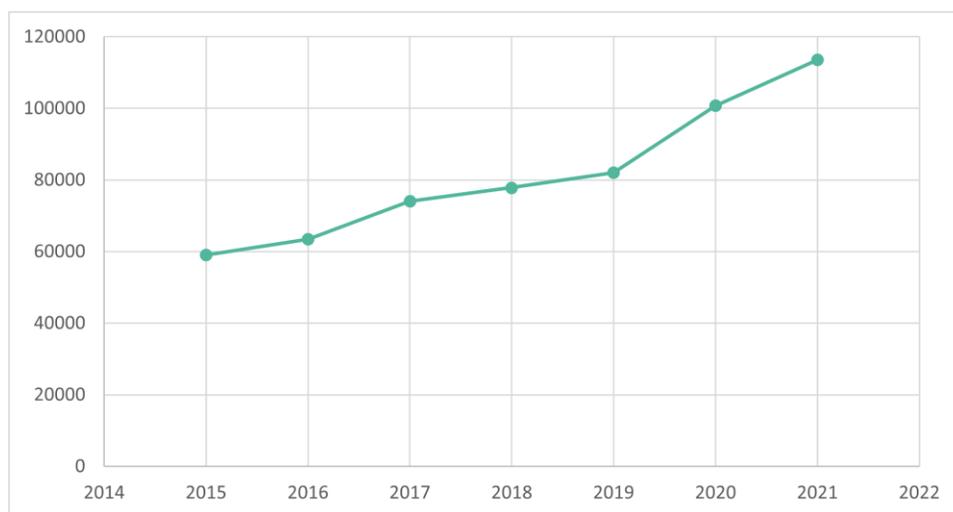


Figura 12: andamento del numero di dataset disponibile su Dataverse

contenuti nel repository. I dati che sono riuscita a raccogliere vanno dal 2015 al 2021<sup>46</sup>. Alla figura 12 vediamo l'andamento nel corso di questi 5 anni. Il numero di dataset disponibile è quasi raddoppiato e la crescita è stata costante. Ad oggi Dataverse conta 113600 dataset. Si registra comunque una crescita molto più lenta inizialmente, che dal 2019 in poi subisce un'impennata.

Ora, invece, prendiamo in considerazione un altro valore: il numero di repository esistenti. Per tenere traccia di tutti i repository di dataset esiste *re3data*<sup>47</sup>, un registro che permette ai

<sup>45</sup> <https://dataverse.harvard.edu/>

<sup>46</sup> Dati raccolti dal sito il 22 giugno 2021

<sup>47</sup> <https://www.re3data.org/>

ricercatori di avere un indice degli archivi internazionali per i dati della ricerca. Per avere una panoramica dell'insieme di questi repository e come essi si stiano sviluppando facciamo un confronto con i dati sul registro del 2015, monitorati in uno studio del DFG, la fondazione tedesca per la ricerca [34]. Il numero di repository di dati registrati alla fine del 2015 era 1381,

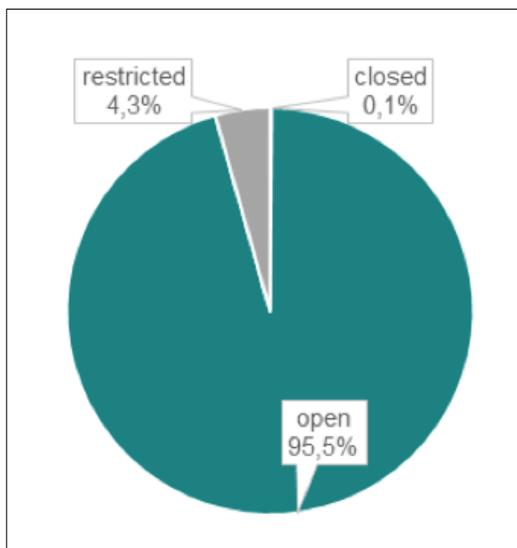


Figura 13: percentuale di archivi aperti, con restrizioni, chiusi in re3data nel 2015

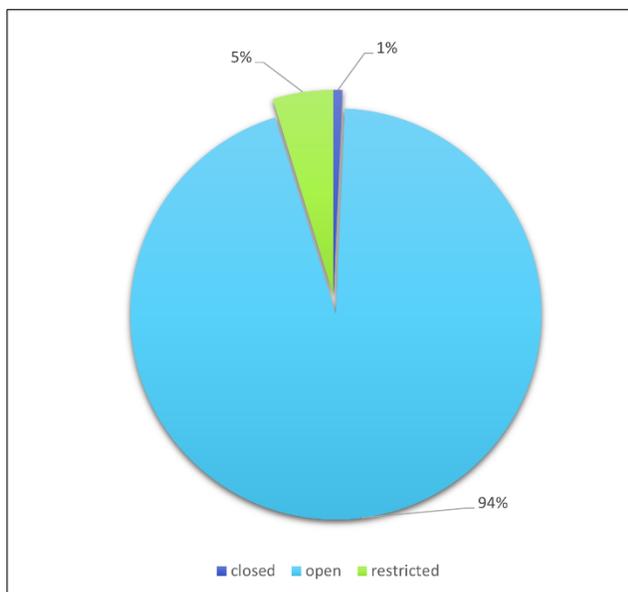


Figura 14: percentuale di archivi aperti, con restrizioni, chiusi in re3data oggi

mentre oggi (giugno 2021) in tutto se ne contano 2700, quindi nel giro di 5 anni e mezzo sono quasi raddoppiati. La percentuale di archivi ad accesso aperto all'interno del registro è rimasta quasi invariata e le figure 13 e 14 lo mostrano. Il numero di archivi con delle restrizioni per l'accesso sono lievemente cresciute in percentuale, ma non rappresentano un cambiamento significativo.

Osservando tutti questi dati possiamo dire che anche il mondo dell'OD (non con la stessa velocità dell'Open Access) stia gradualmente crescendo. Dal punto di vista delle policy sui dati, su ROARmap, la *registry* che conta le policy open access, è possibile vedere che su 1077 policy sull'open access soltanto 183 regolamentano anche i dati della ricerca, di cui la maggior parte, ovvero 137 sono europee. L'Open Access negli ultimi anni ha avuto sicuramente un impatto più prorompente rispetto all'Open Data. Tuttavia, anche l'apertura ai dati è una pratica che viene ogni anno sempre più utilizzata.

#### 4.2.2 Il punto di vista dei ricercatori

L'apertura ai dati della ricerca è uno degli aspetti fondamentali dell'Open Science. Nel rendere i dati condivisibili e quindi anche FAIR è necessario il totale coinvolgimento da parte dei ricercatori. Quindi la domanda sorge spontanea: come viene affrontato l'Open Data dai

ricercatori stessi? In uno studio<sup>48</sup> del 2017 condotto da Elsevier e il *Centre for Science and Technology Studies* dell'Università di Leiden su un campione di 1200 ricercatori sono emerse delle informazioni interessanti. Innanzitutto, un terzo dei ricercatori hanno dichiarato di non condividere i dati delle loro ricerche in nessun modo, mentre la maggior parte delle condivisioni dei dati avviene tramite l'aggiunta in forma di appendice al relativo articolo su rivista perché i ricercatori sostengono che in questo modo hanno una possibilità maggiore di collaborare e di essere citati. Un altro aspetto interessante che è emerso è che, tra coloro che condividono i propri dati soltanto in maniera diretta con i propri colleghi e collaboratori, solo una minima parte la condividono con persone che non conoscono direttamente, e questo è un indicatore del fatto che per i ricercatori la fiducia è un aspetto importante per determinare la possibilità di condivisione dei dati. Nello studio sono state definite due categorie che comprendono ciascuna diversi ambiti scientifici: ambiti in cui la condivisione di dati è una pratica consueta e ambiti in cui la condivisione dei dati è poco contemplata. Comunque, il beneficio della condivisione dei dati della ricerca è riconosciuto dalla stragrande maggioranza in entrambe le categorie. Tuttavia, molti ricercatori sono restii nel metterla in pratica per vari motivi, in particolare la mancanza di un addestramento tecnico appropriato alla gestione dei dati.

### **4.3 Il sito open-science.it**

Per quanto riguarda gli aspetti più pratici di questa tesi ho potuto dare il mio contributo all'Open Science e alla comunità scientifica partecipando allo sviluppo del sito open-science.it che sarà disponibile online entro la fine del 2021. Il portale, il cui sviluppo è affidato al personale del CNR, sarà un punto di incontro per ricercatori, enti finanziatori e istituzioni che vogliono avvicinarsi alle pratiche dell'Open Science in Italia in cui sarà possibile trovare materiali, novità e discussioni sul tema. La navigazione è mirata in base al punto di vista, il quale è determinato dal ruolo dell'utente. Selezionando il ruolo o posizione il sito mostra contenuti più specifici. I punti di vista sono 4: Ente Finanziatore, Ricercatore, Istituzione o Cittadino. La navigazione cambia anche in base alle tematiche selezionate che sono sette: Open Science, Gestione dei dati della ricerca, Regolamenti Open Science, Repository, Enti finanziatori e Open Access, Servizi, OS Progetti e Iniziative. Il portale avrà anche un catalogo di risorse digitali alla cui progettazione ho potuto partecipare attivamente, in particolare nella definizione dei tipi di risorse e dello schema di metadati a loro associato.

#### **4.3.1 Ricerche preliminari**

Per avere un'idea di come un catalogo di risorse sia strutturato ho compiuto una ricerca su dei cataloghi già esistenti. I cataloghi che principalmente ho preso come punto di riferimento

---

<sup>48</sup> Il report si può consultare a questo link: <https://www.elsevier.com/open-science/research-data/open-data-report>

sono quello di *Eosc-pillar*<sup>49</sup> che raccoglie materiale per il *training* dei ricercatori coinvolti nel progetto EOSC, e il catalogo di *D4science*<sup>50</sup>, che ospita una vasta gamma di risorse. Questi due cataloghi mi hanno più che altro dato un'idea di quale sarebbe stata la struttura del catalogo e a che livello i metadati avrebbero determinato la possibilità di navigazione. A ogni risorsa è associata una scheda in cui vengono elencati i metadati descrittivi e amministrativi. Esistono da un lato dei metadati obbligatori che non possono essere modificati e tramite i quali si naviga il sito utilizzandoli come dei filtri, e poi ci sono dei metadati aggiuntivi che servono per descrivere la risorsa. Lo schema di metadati può cambiare solo in base al tipo di risorsa selezionato, quindi a ogni tipo sarà associato uno schema distinto. Una volta identificate le due categorie di risorse per cui avrei definito lo schema di metadati, ovvero il tipo “Evento” e il tipo “Documento”, ho iniziato a consultare diversi siti (non solo cataloghi) per avere una comprensione delle informazioni che fossero necessarie per descrivere queste risorse. Per quanto riguarda gli eventi ho consultato siti con delle sezioni dedicate agli eventi. Mentre per i documenti, sapendo che la maggior parte dei documenti che verrebbero caricati sul sito sarebbero principalmente atti normativi o policy per l'OA, il sito *Roarmap*<sup>51</sup>, l'indice delle policy Open Access, mi è stato molto utile in quanto per ogni policy riporta una scheda con elencati tutti i metadati che la descrivono. Infine, prima di procedere alla realizzazione dello schema di metadati ho preso in esame anche degli standard di metadati, soprattutto per quanto riguarda i metadati che le due categorie di risorse avrebbero condiviso. Tra tutti quelli esaminati, i metadati del *Dublin Core*<sup>52</sup>, standard di metadati nato nell'ambito delle biblioteche digitali ma ampiamente utilizzato anche per la descrizione di altri tipi di risorse, mi sono sembrati i più adatti. \*

### 4.3.2 Lo schema di metadati

Oltre all'ideazione dei singoli metadati, mi sono anche occupata di definire quali fossero i loro valori, le possibili occorrenze, e laddove il valore fosse una stringa da vocabolario controllato, l'insieme di valori possibili. Prima ho definito i metadati base per tutte e due le categorie:

METADATO	VALORE	REPLICABILE	OBBLIGATORIO	descrizione
<b>ID</b>	id	no	si	id interno che identifica la risorsa all'interno del catalogo
<b>ALTERNATIVE IDENTIFIER</b>	stringa	si	no	ID esterni, per esempio DOI, PURL, ORCID ecc.
<b>AUTORE</b>	stringa	no	si	utente che ha creato la risorsa
<b>TITOLO</b>	stringa	no	si	nome con cui viene presentata la risorsa

<sup>49</sup> <https://eosc-pillar.d4science.org/web/eoscpillartrainingandsupport/catalogue>

<sup>50</sup> <https://services.d4science.org/catalogue-d4s>

<sup>51</sup> <https://roarmap.eprints.org/>

<sup>52</sup> Per saperne di più sullo standard Dublin Core visitare il seguente link: <https://dublincore.org/about/>

<b>DATA DI CREAZIONE</b>	data	no	si	data in cui viene creata la risorsa
<b>DESCRIZIONE</b>	stringa	no	si	campo di testo con descrizione della risorsa
<b>KEYWORD</b>	stringa	si	si	parole chiave che descrivono la risorsa
<b>GROUP</b>	stringa	si	si	riferimenti ad altre risorse, della stessa serie o collegati

E poi ho definito i metadati specifici per ciascun tipo. Quello degli eventi:

<b>METADATO</b>	<b>VALORE</b>	<b>REPLICA</b>	<b>OBBLIGATORIO</b>	<b>descrizione</b>
<b>DATA DELL'EVENTO</b>	data	si	si	data dell'evento formato YYYY-MM-DD ora inizio HH:MM e ora fine HH:MM
<b>DURATA</b>	numero intero	no	si	durata dell'evento in ore (possibilmente ricavandoli dal campo "data dell'evento")
<b>SETTORE DISCIPLINARE</b>	stringa(VC) <sup>53</sup>	no	si	Scienze naturali, Ingegneria e Tecnologia, Scienze mediche, Scienze agrarie e veterinarie, Scienze sociali, Scienze Umanistiche e arte
<b>TIPO DI EVENTO</b>	stringa(VC)	no	si	Corso, Seminario, Meeting, Workshop, Convegno o Conferenza, Tavola Rotonda, Conferenza Stampa, Altro
<b>SEDE</b>	stringa	no	solo se si svolge in presenza	dove ha luogo l'evento
<b>MODALITA'</b>	stringa(VC)	no	si	online, in presenza, in presenza e trasmesso
<b>URL</b>	URI	no	si	link al sito dove viene descritto/presentato l'evento
<b>ORGANIZZAT ORE</b>	stringa	si	si	ente che ha (co-)organizzato l'evento
<b>DATA AND RESOURCES</b>	URI	si	no	i file relativi all'evento
<b>TIPOLOGIA DI ACCESSO</b>	Stringa(VC)	no	si	ingresso libero, a pagamento, registrazione/accredito

<sup>53</sup> VC sta per vocabolario controllato, ovvero un insieme di stringhe che possono essere usati come valore per un metadato. Nella tabella troviamo l'insieme di stringhe per ogni VC nella descrizione

<b>INFORMAZIONI PER L'ACCESSO</b>	URI	no	se come tipologia di accesso è stato selezionato: a pagamento o registrazione/accredito	link alla pagina per l'iscrizione, registrazione, pagamento per accedere all'evento
<b>NUMERO DI PARTECIPANTI</b>	data + intero	si se l'evento ha più date	si (a posteriori)	il numero di persone che hanno partecipato (a posteriori) per ciascuna data
<b>NUMERO MAX DI PARTECIPANTI</b>	intero	no	si	numero massimo di persone che possono partecipare
<b>CERTIFICATO FINALE</b>	stringa(VC)	no	si	si, no

E quello dei documenti:

<b>METADATO</b>	<b>VALORE</b>	<b>REPLICABILE</b>	<b>OBBLIGATORIO</b>	<b>descrizione</b>
<b>TIPO DI DOCUMENTO</b>	stringa(VC)	no	si	Regolamento, Report o pubblicazione, atto normativo, raccomandazione, linee guida, altro
<b>URL DOCUMENTO</b>	URI	no	si	link al sito in cui si trova il documento
<b>AUTORE DEL DOCUMENTO</b>	stringa	no	si	Chi ha scritto il documento
<b>CORPORATE AUTHOR</b>	stringa	no	no	se il documento è stato rilasciato da un gruppo di lavoro/esperti
<b>TIPO DI ENTE DI RILASCIO</b>	stringa(VC)	no	si	Università, Ente di Ricerca, Fondazione, Istituto, Ente finanziatore, Policy maker, Altro
<b>ENTE DI RILASCIO</b>	stringa	no	si	il nome dell'ente che ha rilasciato il documento
<b>URL ALL'ENTE DI RILASCIO</b>	URI	no	si	link al sito dell'ente/istituzione che ha rilasciato il documento
<b>DATA DI APPROVAZIONE</b>	data	no	si	la data in cui è stato approvato il documento
<b>DATA DI APPLICAZIONE</b>	data	no	si	data di applicazione della normativa
<b>DATA DELL'ULTIMA REVISIONE</b>	data	no	no	data dell'ultima revisione del documento

<b>CONTATTI</b>	stringa	si	no	indirizzo e-mail o altro per eventuali domande riguardo al documento
-----------------	---------	----	----	--

Inizialmente il mio intento era quello di utilizzare anche delle sottocategorie per creare dei metadati con una specificità ancora maggiore. Tuttavia, la tecnologia utilizzata per lo sviluppo del catalogo non consentiva l'uso di sottocategorie, e quindi ho deciso di utilizzare il concetto di "sottotipo" semplicemente come metadato descrittivo. I metadati che ho derivato dallo schema standard del *Dublin Core* sono titolo, autore, descrizione, data e tipo (tipo di evento e tipo di documento). Il passo successivo è stato quello di definire i metadati attraverso un XML Schema<sup>54</sup>. Per ciascuna categoria sono stati creati due schemi distinti, quindi anche due file distinti. Per ogni metadato è stato definito l'elemento tramite il tag "<xs:element>". L'attributo "type" va a specificare il tipo di valore, mentre con l'attributo "name" si esprime il nome dell'elemento. Gli elementi di tipo complesso, che vanno espressi con il tag "<xs:complexType>" all'interno dell'elemento, in questo caso sono quelli con attributi o quelli che contengono altri elementi. L'unico elemento per cui è stato definito un attributo è "Data and Resources", elemento che dovrà contenere i file relativi a un *item* del catalogo, al quale ho aggiunto l'attributo (tramite tag "<xs:attribute>") di nome "src", al cui interno viene specificata la URI del file che si vuole associare alla risorsa. Mentre l'unico elemento che può contenere altri elementi, è "data dell'evento", nello schema degli eventi, che contiene in questo ordine: "data", "ora inizio" e "ora fine", l'ordine è specificato perché questi elementi sono definiti all'interno del tag "<xs:sequence>". Per quanto riguarda gli elementi i cui valori sono determinati da vocabolario controllato, ho specificato i vari valori che il metadato può assumere inserendoli con il tag "<xs:enumeration>" all'interno del tag "<xs:restriction>". Per esprimere la natura opzionale di alcuni metadati, ho utilizzato l'attributo "minOccurs", che definisce il numero minimo di occorrenze dell'elemento all'interno dello schema, con valore 0, mentre per quelli che possono esseri ripetuti all'infinito ho usato l'attributo "maxOccurs", che invece esprime il numero massimo di ricorrenze dell'elemento, con valore "unbounded". I due schemi scritti in XML Schema si possono trovare nell'appendice di questa tesi. Il lavoro, una volta completato è stato poi consegnato al personale tecnico del CNR che ha proceduto alla configurazione del catalogo.

### 4.3.3 Inserimenti di prova

Per concludere il lavoro e testare l'efficacia dei metadati ho potuto contribuire al popolamento di risorse all'interno del catalogo facendo degli inserimenti. Come layout per la navigazione, alla pagina principale troviamo alla sinistra dei filtri. I metadati che vengono usati come filtri sono quelli di default e non si possono cambiare. Tutte le risorse, invece, sono

<sup>54</sup> XML schema è un linguaggio, definito in XML, che descrive i contenuti (elementi e attributi) di un file XML

elencate in ordine di rilevanza di *default*, ma c'è un menu a tendina che ci permette di cambiare il tipo di ordinamento. Quando si procede all'inserimento è necessario compilare un modulo e riempire i campi che corrispondono ai vari metadati, che saranno visibili una volta caricata la risorsa nella scheda individuale di ogni item. Come primo inserimento di prova ho creato la risorsa per il documento della policy per l'Open Access dell'Università di Pisa e trattandosi di un documento *pdf* questo viene caricato sul sito tramite il metadato *Data and Resources*. È possibile vedere la scheda dedicata alla relativa risorsa alla figura 15, dove il cerchio rosso evidenzia l'elenco di metadati. Il lavoro si è concluso con alcuni inserimenti e lo schema da me definito si è rivelato funzionale agli obiettivi prefissati.

The screenshot shows the item page for 'Regolamento OA - Università di Pisa'. The page includes a sidebar with 'Followers' (0), 'Rating' (5 stars, 0 reviews), 'Your rating' (5 stars, no rating given), 'Organisation' (Open-Science.it), and 'License' (Creative Commons Attribution 4.0). The main content area shows the item title, description, tags ('Open Access', 'Regolamento', 'Unipi'), and a 'Data and Resources' section with a PDF icon and the label 'regpromo\_Unipi'. Below this is the 'Item URL' and a QR code. The 'Additional Info' section is highlighted with a red circle and contains the following table:

Field	Value
Data di applicazione	2015-06-11
Data di approvazione	2015-06-11
Ente di rilascio	Università di Pisa
Referente	Pavone, Gina, info@open-science.it
Tipo di Documento	Regolamento
Tipo di Ente	Università
system:type	Documento

Below the 'Additional Info' table is the 'Management Info' section with the following table:

Field	Value
Author	ALICE BERGONZINI
Maintainer	Alice Bergonzini
Version	1
State	active
Last Updated	24 giugno 2021, 16:20 (UTC+02:00)
Created	23 giugno 2021, 11:58 (UTC+02:00)

Figura 15: scheda descrittiva della risorsa "Regolamento OA -Università di Pisa" sul catalogo di open-science.it

## 5 CONCLUSIONI

I benefici dell'apertura ai risultati della ricerca sono ormai quasi del tutto universalmente riconosciuti. Tuttavia, il sistema di comunicazione scientifica tradizionale fatica a essere sradicato e gli oligopoli di questo mercato, nonostante siano disposti a proporre dei modelli ibridi per l'accesso aperto, continuano a perpetuare questo sistema ed a sfruttarlo a proprio beneficio. Nonostante questo, l'Open Science e tutti i concetti compresi sotto di esso, sono in continua crescita. Da un punto di vista istituzionale c'è una grande volontà, soprattutto nel territorio europeo, di rendere l'accesso aperto la norma e di creare incentivi per realizzare questo cambiamento. La Commissione Europea propone in maniera chiara la Scienza Aperta come uno degli obiettivi fondamentali della strategia europea per l'innovazione e la ricerca e riconosce il potenziale di questo approccio non solo per la comunità scientifica, ma anche per l'intera società. L'Open Science, infatti, già in passato ha dimostrato di essere un acceleratore della scoperta in molte occasioni. Basti pensare a tutta la ricerca compiuta durante la pandemia globale. L'intenso scambio di dati e la collaborazione fra ricercatori di tutto il mondo ha decisamente contribuito al rapido ottenimento di un vaccino. Tuttavia, le modalità in cui l'OS è stata applicata durante la pandemia ha anche dimostrato che non basta rendere aperti i dati, ma anche saperli gestire in maniera opportuna. Ed è proprio per questo che risulta indispensabile seguire le linee guida dell'Open Science e applicare i principi FAIR quando si decide di condividere i propri studi.

Anche da un punto di vista applicativo il mondo dell'Open Science, specialmente nella realtà dell'Open Access, sta subendo una rapida crescita. Esistono sempre più riviste ad accesso aperto, sempre più repository e la percentuale di articoli e dati pubblicati in accesso aperto diventa anno per anno sempre più grande. Alcuni ricercatori sono, comunque, ancora restii a convertirsi alla filosofia *open* perché implica uno sforzo in più e gli incentivi dal punto di vista lavorativo sono ancora pochi. Infatti, il prestigio rappresentato dalle riviste è ancora molto rilevante quando si parla di carriera nel mondo della ricerca. Inoltre, è ancora poco compreso il fatto che aprire l'accesso non significa fare Open Science. Alcune pratiche che sono previste dal paradigma, come l'Open Peer Review o la preregistrazione degli studi, non vengono applicate con frequenza come dovrebbero. Nonostante questo, in molte realtà è stato avviato un cammino verso la totale apertura dei risultati accademici finanziati da enti pubblici e questo è sicuramente sintomo di una volontà di cambiamento all'interno della comunità scientifica.

Aver partecipato allo sviluppo del catalogo per il futuro sito di open-science.it, mi ha consentito di aggiungere un tassello a quella che con l'avanzare del tempo sta diventando un'infrastruttura globale fatta di servizi e strumenti per la promozione della Scienza Aperta. Inoltre, l'aver contribuito a questo progetto mi ha permesso di applicare alcune delle mie conoscenze acquisite durante il percorso di studi.

## 6 APPENDICE

### 6.1 XML Schema per lo schema di metadati “Documenti”

```
<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:element name="Documenti e normative">
    <xs:complexType>
      <!-- metadati per tutti i tipi di risorsa -->
      <xs:element name="id" type="xs:ID"/>
      <xs:element name="Alternative Identifier" type="xs:string"
maxOccurs="unbounded" minOccurs="0"/>
      <xs:element name="Autore" type=""/> <!-- come identificare
l'utente, nome o id? -->
      <xs:element name="Titolo" type="xs:string"/>
      <xs:element name="Data di Creazione" type="xs:date"/> <!-- Data
della crezione della risorsa -->
      <xs:element name="Descrizione" type="xs:string"/>
      <xs:element name="Keyword" type="xs:string"
maxOccurs="unbounded"/>
      <xs:element name="Group" type="xs:string"
maxOccurs="unbounded"/>
      <xs:element name="Data and Resources" maxOccurs="unbounded"
minOccurs="0">
        <xs:complexType>
          <xs:attribute name="src" type="anyURI"
use="required"/>
        </xs:complexType>
      </xs:element>
      <!-- metadati per il tipo: Documenti e normative -->
      <xs:element name="Tipo di documento">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:enumeration value="Regolamento"/>
            <xs:enumeration value="Report o
pubblicazione"/>
            <xs:enumeration value="Atto normativo"/>
            <xs:enumeration value="Raccomandazione"/>
            <xs:enumeration value="Linee Guida"/>
            <xs:enumeration value="Altro"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:element>
      <xs:element name="URL Documento" type="xs:anyURI"/>
      <xs:element name="Atuore del documento" type="xs:string"/>
      <xs:element name="Corporate Author" type="xs:string"
minOccurs="0"/>
      <xs:element name="Tipo di ente di rilascio">
        <xs:simpleType>
          <xs:restriction base="xs:string">\
```

```

        <xs:enumeration value="Università"/>
        <xs:enumeration value="Ente di ricerca"/>
        <xs:enumeration value="Fondazione"/>
        <xs:enumeration value="Istituto"/>
        <xs:enumeration value="Ente finanziatore"/>
        <xs:enumeration value="Policy Maker"/>
        <xs:enumeration value="Altro"/>
    </xs:restriction>
</xs:simpleType>
</xs:element>
<xs:element name="Ente di Rilascio" type="xs:string"/> <!-- il
nome dell-ente che ha rilasciato il documento -->
<xs:element name="URL dell'Ente di rilascio" type="xs:anyURI"/>
<xs:element name="Data di approvazione" type="xs:date"/>
<xs:element name="Data di applicazione" type="xs:date"/>
<xs:element name="Data dell'ultima revisione" type="xs:date"/>
<xs:element name="Luogo di deposito" type="xs:string"/>
<xs:element name="Contatti" type="xs:string"
maxOccurs="unbounded" minOccurs="0"/>
</xs:complexType>
</xs:element>
</xs:schema>

```

## 6.2 XML Schema per lo schema di metadati “Eventi”

```

<?xml version="1.0"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <xs:element name="Evento">
        <xs:complexType>
            <!-- metadati per tutti i tipi di risorsa -->
            <xs:element name="id" type="xs:ID"/>
            <xs:element name="Alternative Identifier" type="xs:string"
maxOccurs="unbounded" minOccurs="0"/>
            <xs:element name="Autore" type=""/> <!-- come identificare
l'utente, nome o id? -->
            <xs:element name="Titolo" type="xs:string"/>
            <xs:element name="Data di Creazione" type="xs:date"/> <!-- Data
della creazione della risorsa -->
            <xs:element name="Descrizione" type="xs:string"/>
            <xs:element name="Keyword" type="xs:string"
maxOccurs="unbounded"/>
            <xs:element name="Group" type="xs:string"
maxOccurs="unbounded"/>
            <xs:element name="Data and Resources" maxOccurs="unbounded"
minOccurs="0">
                <xs:complexType>
                    <xs:attribute name="src" type="anyURI"
use="required"/>
                </xs:complexType>
            </xs:element>
        </xs:complexType>
    </xs:element>

```

```

        <!-- metadati per il tipo: Evento -->
<xs:element name="Data dell'evento" maxOccurs="unbounded">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="data" type="xs:date"/>
      <xs:element name="Ora inizio"
type="xs:time"/>
      <xs:element name="Ora fine" type="xs:time"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<xs:element name="Durata" type="xs:decimal"/>
<xs:element name="Settore disciplinare">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Scienze naturali"/>
      <xs:enumeration value="Ingegneria e
Tecnologia"/>
      <xs:enumeration value="Scienze mediche"/>
      <xs:enumeration value="Scienze agrarie e
veterinarie"/>
      <xs:enumeration value="Scienze sociali"/>
      <xs:enumeration value="Scienze umanistiche e
arte"/>
    </xs:restriction>
  </xs:simpleType>
</xs:element>
<xs:element name="Tipo di Evento">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Corso"/>
      <xs:enumeration value="Seminario"/>
      <xs:enumeration value="Meeting"/>
      <xs:enumeration value="Workshop"/>
      <xs:enumeration
value="Convegno/Conferenza"/>
      <xs:enumeration value="Tavola
rotonda"/>
      <xs:enumeration value="Conferenza
stampa"/>
      <xs:enumeration value="Altro"/>
    </xs:restriction>
  </xs:simpleType>
</xs:element>
<xs:element name="Sede" type="xs:string"/> <!-- campo esiste
solo se il campo modalità è in presenza o in presenza e trasmesso -->
<xs:element name="Modalità">
  <xs:simpleType>
    <xs:restriction base="xs:string">
      <xs:enumeration value="Online"/>
      <xs:enumeration value="In presenza"/>

```

```

        <xs:enumeration value="In presenza e
trasmesso"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="URL" type="xs:anyURI"/> <!-- il link alla
pagina web dell'evento -->
<xs:element name="Organizzatore" type="xs:string"
maxOccurs="unbounded"/>
<xs:element name="Tipologia di accesso">
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:enumeration value="Ingresso
libero"/>
            <xs:enumeration value="A pagamento"/>
            <xs:enumeration
value="Registrazione/Accredito"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
<xs:element name="Informazioni per l'accesso" type="xs:anyURI"/>
<!-- link alla pagina web per l'iscrizione/registrazione o pagamento, campo esiste
solo se Tipologia di accesso è a pagamento o registrazione/Accredito -->
<xs:element name="Numero di partecipanti" maxOccurs="unbounded">
    <xs:complexType>
        <xs:sequence>
            <xs:element name="data" type="xs:date"/>
            <xs:element name="Numero partecipanti"
type="xs:integer"/>
        </xs:sequence>
    </xs:complexType>
</xs:element> <!-- numero di persone che hanno partecipato (a
posteriori) -->
<xs:element name="Numero max di partecipanti"
type="xs:integer"/>
<xs:element name="Certificato Finale"> <!-- se viene rilasciato
un certificato o meno -->
    <xs:simpleType>
        <xs:restriction base="xs:string">
            <xs:enumeration value="Si"/>
            <xs:enumeration value="No"/>
        </xs:restriction>
    </xs:simpleType>
</xs:element>
</xs:complexType>
</xs:element>
</xs:schema>

```

## BIBLIOGRAFIA

- [1] R. Caso, «La scienza aperta contro la mercificazione della ricerca accademica?», Trento Law and Technology Research Group, Trento, 2016.
- [2] A. Posada e G. Chen, «Inequality in Knowledge Production: The Integration of Academia Infrastructure by Big Publishers», ELPUB, Toronto, 2018.
- [3] G. Origgi, G. B. Ramello e F. Silva, «Publish or Perish. Cause e conseguenze di un paradigma», *Economia Politica Industriale*, vol. 4, pp. 9-25, 2014.
- [4] J.-C. Guédon, Open Access. Contro gli oligopoli nel sapere, Pisa: Edizioni ETS, 2009.
- [5] B. Brembs, K. Button e M. Munafò, «Deep Impact: unintended consequences of journal rank», *Frontiers in Human Neuroscience*, 2013.
- [6] T. Ross-Hellauer, «What is open peer review? A systematic review», F1000 Research, 2017.
- [7] P. E. Smaldino e R. McElreath, «The natural selection of bad science», Royal Society Open Science, 2016.
- [8] F. C. Fang e A. Casadevall, «Retracted Science and the Retractive Index», *Infection and Immunity*, 2011.
- [9] E. Giglia, «Open Access e Open Science. Per una scienza più efficace», *Journal of Biomedical Practitioners*, vol. 1, n. 1, pp. 7-28, 2017.
- [10] E. Carafoli, «Scientific misconduct: the dark side of science», *Rendiconti Lincei*, vol. 26, pp. 369-382, 2015.
- [11] M. L. Grieneisen e M. Zhang, «A Comprehensive Survey of Retracted Articles from the Scholarly Literature», *PLoS ONE*, vol. 7, n. 10, 2012.
- [12] L. K. John, G. Loewenstein e e. al, «Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling», *Psychological Science*, vol. 23, n. 5, pp. 524-532, 2012.
- [13] M. L. Head, L. Holma, R. Lanfear, A. T.Kahn e M. D. Jennions, «The Extent and Consequences of P-Hacking in Science», *PLoS Biology*, vol. 13, n. 3, 2015.
- [14] D. Marasini, P. Quatto e E. Ripamonti, «Una nota sulla riproducibilità nella ricerca scientifica», *Il Mulino*, Vol. %1 di %21-2, pp. 335-350, 2019.
- [15] C. M. Stracke, «Open Science and Radical Solutions for Diversity, Equity and Quality in Research: A Literature Review of Different Research Schools, Philosophies and Frameworks and Their Potential Impact on Science and Education», in *Radical Solutions and Open Science*, Springer, 2020, pp. 17-37.

- [16] M. D. M. A. I. Wilkinson e e. al., «The FAIR Guiding Principles for scientific data management and stewardship,» *Scientific Data*, vol. 3, n. 160018, 2016.
- [17] E. Lazzeri, E. Giglia e G. Pavone, *Open Science e finanziamenti europei: come ottemperare agli obblighi nei progetti H2020 e in Horizon Europe*, Zenodo, 2021.
- [18] S. Bezjak, P. Conzett, P. L. Fernandes, E. Görögh, K. Helbig, B. Kramer, I. Labastida, K. Niemeyer, F. Psomopoulos, T. Ross-Hellauer, R. T. J. Schneider, E. Verbakel, A. Clyburne-Sherin, L. Heller e H. Brinken, *Il Manuale per Formatori della Scienza Aperta*, Zenodo, 2019.
- [19] B. Kramer e J. Bosman, *Rainbow of open science practices*, Zenodo, 2018.
- [20] J.-C. Burgelman, C. Pascu, K. Szkuta, R. Von Schomberg, A. Karalopoulos, K. Repanas e M. Schoupe, «Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century,» *Frontiers in Big Data*, vol. 2, n. 43, 2019.
- [21] C. Allen e D. M. A. Mehler, «Open Science challenges, benefist and tips in early career and beyond,» *PLoS Biology*, vol. 17, n. 5, 2019.
- [22] L. M. Federer, C. W. J. D. J. Belter, A. Livinski, Y.-L. Lu, L. N. Snyders e H. Thompson, «Data sharing in PLOS ONE: An analysis of Data Availability Statements,» *PLoS ONE*, vol. 13, n. 5, 2018.
- [23] S. A. Moore, «Open Access, Plan S and 'Radically Liberatory' Forms of Academic Freedom,» *Wiley Online Library*, 2021.
- [24] J.-C. Burgelman, «Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story),» *Data Intelligence*, vol. 3, n. 1, pp. 5-19, 2021.
- [25] P. Castellucci e C. Barillari, «Il cerchio si allarga intorno al mondo: open access contro la pandemia,» *AIB Studi*, vol. 60, n. 1, 2020.
- [26] K. T. Kadakia, A. L. Beckman, J. S. Ross e H. M. Krumholz, «Leveraging Open Science to Accelerate Research,» *NEJM*, vol. 384, n. 17, p. e61, 2021.
- [27] L. Besançon, N. Peiffer-Smadja, C. Segalas, H. Jiang, P. Masuzzo, C. Smout, E. Billy, M. Deforet e C. Leyrat, «Open Science Saves Lives: Lessons from the COVID-19 Pandemic,» *BMC Med Res Methodol*, vol. 1, n. 21, 2021.
- [28] P. Suber, «Open Access Overview,» in *Open Access - Unrestricted Access to Published Research*, 2006, pp. 7 - 12.
- [29] B.-C. Björk, M. Laakso, P. Welling e P. Paetau, «Anatomy of green open access,» *Journal of the Association for Information Science and Technology*, n. 65, pp. 237-250, 2014.
- [30] M. Laakso e B.-C. Björk, «Hybrid Open Access - A longitudinal study,» *Journal of Infometrics*, vol. 10, n. 4, pp. 919-932, 2016.

- [31] P. Murray-Rust, «Open Data in Science,» *Serials Review*, vol. 34, n. 1, pp. 52-64, 2008.
- [32] D. Orru', «Open Data Steward: bibliotecari e alfabetizzazione ai dati aperti,» *AIB STUDI*, vol. 60, n. 2, pp. 311-323, 2020.
- [33] P. Huston, V. L. Edge e E. Bernier, «Reaping the benefits of Open Data in public Health,» *Can Commun Dis Rep*, vol. 45, n. 10, pp. 252-256, 2019.
- [34] M. Kindling, H. Pampel, S. van de Sandt, J. Rücknagel, P. Vierkant, G. Kloska, M. Witt, P. Schirmbacher, R. Bertelmann e F. Scholze, «The Landscape of Research Data Repositories in 2015: A re3data Analyzsis,» *D-Lib Magazine*, vol. 23, n. 3/4, 2017.