

Una panoramica di Open Data, Open Science e Open Access

Sara Di Giorgio
Crete, May 2017

1ITA

L'accesso aperto ai dati e alla letteratura scientifica, sono elementi fondamentali della cosiddetta 'Scienza Aperta'. Come astronauti su Marte, siamo impegnati ad esplorare un nuovo panorama scientifico in corso di trasformazione, con nuove sfide da affrontare.

Il paradigma della 'Scienza Aperta' sta apportando profondi cambiamenti nei metodi di ricerca, di collaborazione dei ricercatori, di condivisione e organizzazione delle conoscenze. La Scienza Aperta si avvale delle tecnologie digitali ed è trainata dalla crescita esponenziale dei dati, dalla globalizzazione e dall'allargamento delle comunità scientifiche.

La Scienza Aperta ha come logica conseguenza la trasformazione in corso del modo in cui si esegue e si organizza la ricerca e propone nuovi modelli di collaborazione tra i ricercatori, basati sulla condivisione.

Ha infatti un impatto sull'intero processo della ricerca e sulla sua organizzazione, dal punto di avvio, alla pubblicazione dei risultati ottenuti.

L'Open Science significa apertura dei dati della ricerca, dei metodi, dei risultati e della loro pubblicazione.

Per sviluppare la Scienza Aperta è perciò necessario concentrarsi sui dati aperti, sull'accesso aperto e sull'affidabilità della ricerca.

2ITA

Il paradigma dell'Open Access sta di fatto ridefinendo le condizioni di sviluppo della scienza puntando sull'innovazione (Open) e incoraggiando una circolazione più dinamica della conoscenza.

La Commissione europea è sostenitrice di azioni politiche per il rafforzamento della scienza aperta puntando alla circolazione delle conoscenze.

La Scienza Aperta è infatti una delle tre aree prioritarie della politica europea in materia di ricerca, scienza e innovazione. La politica è stata presentata nel 2015 da Carlos Moedas - commissario europeo alla ricerca e all'innovazione, in un discorso intitolato 'Un nuovo inizio per l'Europa: apertura all'ERA dell'innovazione' alla Conferenza 'Open Innovation, Open Science, Open to the World' (Bruxelles 22-23 giugno 2015).

"L'apertura definirà il nostro secolo", afferma il Commissario Carlos Moedas.

L'apertura è un principio fondamentale della scienza e grazie all'avvento di nuove tecnologie digitali e della rete web, è possibile realizzare un incremento significativo della produzione

scientifica rendendo possibile la collaborazione remota degli scienziati e lo sviluppo di iniziative innovative grazie al riuso dei risultati della ricerca resi accessibili in modalità aperta. 'La Scienza Aperta dipende da menti aperte', ha detto nel video, Neelie Kroes, l'ex commissario europeo.

Ma Open Science non significa "scienza libera". È essenziale garantire che la proprietà intellettuale sia protetta prima di rendere pubbliche le conoscenze per attirare gli investimenti in grado di tradurre i risultati della ricerca in innovazione. In considerazione di ciò, l'accesso sempre più ampio alle pubblicazioni scientifiche e ai dati di ricerca può contribuire ad accelerare l'innovazione.

La condivisione delle informazioni e l'allargamento delle reti di collaborazione sono le chiavi per il successo dei nuovi obiettivi fissati dalla ricerca.

L'UE ha messo a punto l'Open Science Monitor, un'iniziativa che propone una serie di indicatori per l'analisi e il monitoraggio delle attività legate allo sviluppo dell'Open Science in Europa. E' stata elaborata un grafico in cui sono presenti le caratteristiche e gli indicatori validi sia per l'accesso aperto alle pubblicazioni scientifiche che alla comunicazione scientifica.

3ITA

'La scienza viene condotta e comunicata in modo da consentire ai ricercatori e agli studiosi di contribuire, collaborare e aggiungere nuovi dati, nuovi risultati e protocolli, liberamente disponibili nelle diverse fasi del processo di ricerca". Questa definizione del Research Information Network del Regno Unito, coglie il carattere innovativo che sta emergendo rispetto al modo in cui la scienza e la ricerca sono state condotte negli ultimi cinquant'anni: assistiamo al passaggio dalle pratiche standard di pubblicazione dei risultati della ricerca nelle pubblicazioni scientifiche alla condivisione e all'utilizzo di tutte le conoscenze disponibili nel processo di ricerca.

Quindi l'Open Science è molto più della mera pubblicazione in Open Access, come si può vedere nel diagramma di Andreas Neuhold del Digital Curation Centre. Infatti include anche Open Data, Open Source che significa condividere sia il codice che gli strumenti per consentire ad altri di riprodurre il lavoro di ricerca. E poi metodi aperti, per documentare e condividere i flussi di lavoro e utilizzare gli strumenti disponibili in rete per facilitare la collaborazione e l'interazione con il mondo esterno.

Open Science sta modificando l'intero "modello economico" di fare scienza e ricerca - dalla selezione dei temi della ricerca, alla sua esecuzione e all'uso e ri-uso dei risultati ottenuti - nonché sta ridefinendo il ruolo di tutti gli attori coinvolti, come università ed editori.

Gli scienziati, i cittadini, gli editori, gli istituti di ricerca, i finanziatori di ricerca pubblici e privati, gli studenti e i professionisti nel campo dell'istruzione nonché le imprese di tutto il mondo possono partecipare e condividere un ambiente virtuale aperto.

4ITA

La ricerca ha una lunga storia in Europa. Negli ultimi decenni, si è gradualmente sviluppata la collaborazione tra i paesi europei che integra innanzitutto le attività di ricerca e successivamente le politiche in questo campo.

Il primo programma quadro europeo per la ricerca è stato lanciato 30 anni fa per riunire esperienze provenienti da tutta la Comunità europea e per rendere l'Europa più competitiva in ambito tecnologico. Il finanziamento europeo della ricerca si è evoluto dal finanziamento di programmi distinti per diventare oggi uno dei componenti principali della ricerca e dell'innovazione in Europa.

I programmi quadro sono diventati una parte fondamentale della cooperazione nella ricerca in Europa. Si è partiti con l'obiettivo di sostenere la collaborazione transfrontaliera in materia di ricerca e tecnologia, arrivando oggi allo sviluppo di azioni di coordinamento europeo delle attività e delle politiche. La ragione di questo è semplice: la ricerca, la tecnologia e l'innovazione sono al centro dell'economia europea e sono elementi di sviluppo nella nostra società.

Oggi Orizzonte 2020, l'ottavo programma quadro, è ancora più importante e ambizioso con un budget di circa 80 miliardi di euro.

Con Orizzonte 2020, la ricerca e l'innovazione assumono un ruolo fondamentale nella politica della Commissione europea.

5ITA

Dall'Open Science all'European Science Cloud: la Commissione ha lanciato una nuova iniziativa 'European Open Science Cloud' che offrirà ai ricercatori europei una piattaforma comune dove poter immagazzinare, gestire, analizzare e riutilizzare i dati prodotti dalla ricerca in Europa, superando le barriere disciplinari e frontaliere, integrando e consolidando le piattaforme di infrastrutture digitali. Questa piattaforma riunirà le infrastrutture scientifiche di ricerca esistenti e supporterà lo sviluppo di servizi basati sul Cloud. Una scelta che nasce dalla consapevolezza che l'Europa è il più grande produttore mondiale di dati scientifici, che oggi sono difficili da utilizzare e coordinare proprio a causa della mancanza di una piattaforma comune.

Questa iniziativa sarà sostenuta da un'infrastruttura di dati europea, che si sviluppa grazie alla rete di banda ultra larga, a servizi di conservazione del digitale su larga scala e alla capacità di super calcolo necessaria per accedere ed elaborare efficacemente i grandi set di dati memorizzati nel Cloud. Questa infrastruttura di livello mondiale garantirà all'Europa la partecipazione allo sviluppo globale del calcolo ad alte prestazioni in linea con il suo potenziale economico e di conoscenza.

Inizialmente ci si focalizzerà sulle comunità della ricerca scientifica, ma nel tempo la base di utenti verrà ampliata nel settore pubblico e nell'industria. Questa iniziativa fa parte di un pacchetto di misure volte a rafforzare il posizionamento dell'Europa nei processi di innovazione guidati dalla produzione dei dati, per migliorare la competitività e la coesione e contribuire a creare un mercato unico digitale in Europa.

La European Cloud Initiative sbloccherà il valore dei Big Data fornendo, entro il 2020, capacità di super calcolo, connettività ad alta velocità, insieme a dati e servizi software avanzati per la scienza, l'industria e il settore pubblico.

Con questa iniziativa la scienza sarà più efficiente e produttiva e consentirà a milioni di ricercatori di condividere e analizzare i dati in un ambiente affidabile, attraverso diverse tecnologie, discipline e frontiere.

La European Cloud Initiative semplificherà ai ricercatori e agli innovatori l'accesso e il riuso dei dati, ridurrà i costi di archiviazione e permetterà analisi ad alte prestazioni. Rendere apertamente disponibili i dati della ricerca contribuirà così a migliorare la competitività dell'Europa attraverso la creazione di start-up, le PMI e l'innovazione basata sui dati.

6ITA

Il progetto Open Definition di Open Knowledge Foundation così definisce dati (e contenuti) aperti: 'un contenuto o un dato è aperto se chiunque è in grado di utilizzarlo, ri-utilizzarlo e ridistribuirlo, soggetto, al massimo, alla richiesta di attribuzione e condivisione allo stesso modo'.

Secondo tale definizione, sono dati di tipo aperto, i dati che presentano le seguenti tre caratteristiche:

Disponibilità e accesso: sono resi disponibili nel loro complesso gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

Riutilizzo e redistribuzione: sono disponibili secondo i termini di una licenza che ne permette l'uso e la distribuzione da parte di chiunque, in formato disaggregato. I dati devono essere leggibili anche in modalità automatica.

Partecipazione universale: chiunque deve essere in grado di utilizzare, riutilizzare e ridistribuire i dati – senza alcuna discriminazione nei confronti di iniziative, persone o gruppi. Ad esempio, non sono consentite restrizioni "non commerciali" che impediscano l'uso "commerciale" dei dati, o restrizioni di uso per determinati scopi (ad esempio solo per la didattica).

Quali sono le tipologie di dati aperti?

Ci sono molti tipi di dati aperti che hanno potenziali usi e applicazioni:

Cultura: dati relativi a opere culturali e manufatti - ad esempio titoli e autori - e generalmente raccolti e conservati da gallerie, biblioteche, archivi e musei.

Scienza: i dati che vengono prodotti dalla ricerca scientifica.

Economia: dati relativi a bilanci pubblici (spese e entrate) e informazioni sui mercati finanziari (azioni, azioni, obbligazioni, ecc.).

Statistica: dati prodotti da uffici statistici come ad esempio relativi a iniziative di censimento o indicatori socio-economici.

Clima: informazioni utilizzate per conoscere e prevedere condizioni meteorologiche e climatiche.

Ambiente: informazioni relative all'ambiente naturale, come ad esempio dati relativi alla presenza di inquinanti, alla qualità dei fiumi e dei mari.

Queste sono le licenze più diffuse per i dati aperti.

Liberatoria per il Pubblico Dominio: è fortemente raccomandata la liberatoria esplicita dei dati nel dominio pubblico tramite PDDL o CCZero che garantisce la conformità sia al Science Commons Protocol for Implementing Open Access Data che alla definizione di dato aperto dell'Open Knowledge Initiative. In questo modo l'informazione o il contenuto è completamente libero da diritti, può essere utilizzato da terzi anche per fini commerciali senza obbligo di citare la fonte;

CC BY, i dati possono essere utilizzati da terzi anche per fini commerciali, ma con obbligo di citare la fonte;

CC BY SA, i dati possono essere utilizzati con obbligo di citare la fonte, e debbono essere condivisi allo stesso modo, cioè se il materiale viene remixato o trasformato o ci si basa su di esso, è necessario distribuire i propri contributi con la stessa licenza del materiale originario.

7ITA

Oltre ad essere Aperti i dati dovrebbero essere disponibili secondo il paradigma FAIR, un'iniziativa sviluppata dal gruppo Force 11 e ora adottata dalla Commissione Europea per i progetti H2020.

I dati saranno perciò:

Reperibili (Findable):

- F1. I (Meta)dati hanno un identificatore globale e persistente
- F2. I Dati vengono descritti attraverso metadati ricchi.
- F3. I (Meta)dati vengono registrati o indicizzati in quanto risorsa ricercabile.
- F4. I Metadati specificano l'identificatore dei dati.

Accessibile (Accessible):

- A1 I (Meta)dati sono reperibili attraverso il proprio identificatore, attraverso un protocollo standard.
- A1.1 Il protocollo è aperto, gratuito e universalmente implementabile.
- A1.2 Il protocollo consente una procedura di autenticazione e autorizzazione, ove necessario.
- A2 I metadati sono accessibili, anche quando i dati non sono più disponibili.

Interoperabile (Interoperable):

- I1. I (Meta)dati utilizzano una lingua formale, accessibile, condivisa e ampiamente applicabile per la rappresentazione della conoscenza.
- I2. I (Meta)dati utilizzano vocabolari che seguono i principi FAIR.
- I3. I (Meta)dati includono riferimenti qualificati ad altri (meta)dati.

Riutilizzabile (Re-usable):

- R1. I (Meta)dati presentano una pluralità di attributi accurati e rilevanti.
- R1.1. I (Metadati vengono rilasciati con una licenza chiara e accessibile per l'utilizzo dei dati.
- R1.2. I (Meta)dati sono associati alla loro provenienza.
- R1.3. I (Meta)dati soddisfano gli standard comunemente utilizzati in un determinato dominio disciplinare.

8ITA

Come fare i dati aperti? Puoi seguire la linea guida dell' Open Knowledge Foundation che fornisce un valido aiuto nel processo:

1. Scegliere il set di dati

- Quali dati aprire? Potrebbe essere necessario rivedere la scelta iniziale se si verificano dei problemi lungo il processo di apertura. In primo luogo si consiglia di chiedere alla comunità di utenti. Saranno loro infatti ad accedere e utilizzare i dati, perciò probabilmente hanno una buona comprensione dei dati che potrebbero essere utili.

2. Applicare una licenza aperta

- E' necessario determinare l'esistenza di eventuali diritti d'autore. Quindi si applica una licenza adatta, ad esempio CC-BY. Anche nei casi in cui l'esistenza dei diritti è incerta, è importante applicare una licenza, semplicemente per motivi di chiarezza.

3. Rendere disponibili i dati

Fornire i dati in un formato adeguato. Utilizzare standard di metadati rilevanti nel proprio settore. Utilizzare repository per distribuire i dati. I dati aperti devono essere tecnicamente aperti e legalmente aperti. In particolare, i dati devono essere disponibili in un formato leggibile dalle macchine. Per assicurarsi che i dati possano essere compresi dagli utenti, i ricercatori dovrebbero creare i metadati e mettere a disposizione una documentazione adeguata per renderli comprensibili a tutti. I Metadati sono informazioni descrittive che servono per identificare e comprendere la struttura dei dati, come ad es. Titolo, Autore. La documentazione fornisce un contesto più ampio. È utile condividere la metodologia, il flusso di lavoro, il software e tutte le informazioni necessarie per comprendere i dati, come ad es. la spiegazione di abbreviazioni o acronimi.

4. Rendere i dati ricercabili

Ci sono molti standard che possono essere usati. Il progetto PARTHENOS ha avviato un catalogo di metadati disciplinari.

9ITA

5. Utilizzare formati di file appropriati

Se desideri che i tuoi dati siano riutilizzati e vuoi garantirne la sostenibilità a lungo termine, utilizzerai formati aperti e non proprietari.

10ITA

Open Access è il termine comunemente utilizzato per indicare un movimento che promuove la libera circolazione e l'uso non restrittivo dei risultati della ricerca e del sapere scientifico. La

letteratura Open Access (OA) è digitale, online, libera da costi per il lettore, libera dalla gran parte delle restrizioni di copyright, e non ha barriere date da costi o autorizzazioni. Le licenze che si applicano all'Open Access permettono la libera circolazione dei dati e del sapere, garantendo eventualmente l'attribuzione appropriata dell'autore.

Ciò che rende possibile l'Accesso Aperto è la rete Internet e il consenso dell'autore o del titolare del copyright.

Il web offre nuovi metodi di pubblicazione: infatti rende la circolazione della ricerca più facile, ampia, veloce, e spesso meno costosa, favorendo in questo modo un progresso più rapido della scienza.

Il web offre nuovi sbocchi e metodi per la condivisione e l'utilizzo dei risultati della ricerca e per supportare la conoscenza, proponendo un nuovo paradigma di accesso alla conoscenza che consente ai docenti e alle università di trarre pieno vantaggio dai repository istituzionali o disciplinari per la ricerca.

Sfruttando le potenzialità offerte dalla rete, gli articoli sono gratuitamente resi accessibili agli utenti senza restrizioni e barriere. La disseminazione dell'informazione garantisce un impatto efficace: più un articolo è liberamente scaricabile, più è letto, più è citato. Questo favorisce la condivisione del sapere e quindi un più rapido avanzamento della conoscenza, senza barriere, in tutto il mondo.

Si sono affermate due strategie principali per garantire pubblicazioni ad accesso aperto:

La via verde ("green road") che prevede l'autoarchiviazione ("self-archiving") da parte degli autori, di copie dei loro articoli, prima, contemporaneamente o dopo la pubblicazione in repository on-line istituzionali o disciplinari, o ancora nei loro siti personali. Alcuni editori richiedono che l'accesso aperto sia concesso solo dopo che è trascorso un periodo di embargo.

La via d'oro ("gold road") cioè la pubblicazione di riviste in cui gli articoli sono direttamente ed immediatamente accessibili ad accesso aperto. Tali pubblicazioni vengono chiamate riviste ad accesso aperto. In questo modello, il costo le spese di pubblicazione non costituisce più una barriera all'accesso, perché è previsto all'origine sotto forma di abbonamento. La scelta nell'Open Access è di far ricadere i costi (minimi) su chi produce la ricerca invece che sui lettori, per assicurare la massima disseminazione.

I costi editoriali, denominati *Article Processing Charges* (APCs), sono generalmente a carico dell'università o dell'istituto di ricerca o dell'ente che finanzia la ricerca. La tendenza a livello internazionale è perciò sempre più quella di ricomprendere i costi per la pubblicazione Open Access nel budget iniziale della ricerca.

11ITA

I vantaggi di Accesso Aperto sono:

- I ricercatori hanno un accesso immediato e gratuito ai dati e ai risultati della ricerca
- Garanzia di trasparenza e riproducibilità della ricerca
- Maggiore visibilità, impatto e diffusione e del lavoro di ricerca
- Sviluppo di nuove collaborazioni e partnership di ricerca

- Accesso a lungo termine ai dati e alla conoscenza
- Accelerazione della ricerca grazie ad una maggiore efficienza

Alcune comunità di ricerca hanno constatato il vantaggio nel condividere i dati, ottenendo in modo più rapido dei risultati. Ad esempio la comunità dei ricercatori impegnati nello studio dell'Alzheimer hanno condiviso immediatamente i dati raggiungendo più facilmente e rapidamente importanti scoperte scientifiche sulla malattia e la sua cura.

- Aumento esponenziale del numero di citazioni per i singoli ricercatori. E' oramai evidente che i dati condivisi ricevono il 9% di citazioni in più. Gli articoli liberamente disponibili in rete sono infatti più citati.

11ITA

Se vuoi saperne di più consulta il sito web www.whyopenresearch.org dove troverai statistiche, informazioni sull'apertura dei dati e dei risultati della ricerca, per renderli valutabili e riutilizzabili da tutti. Lo slogan che viene qui adottato è "L'accesso alla conoscenza è un diritto umano e come tale dovrebbe essere trattato".

12ITA

La Commissione Europea, in quanto principale finanziatore della ricerca in Europa, sostiene lo sviluppo dell'Open Access con l'obiettivo di rendere ad accesso aperto entro il 2020 tutti i lavori pubblicati grazie a finanziamenti europei. PERCHE L'ACCESSO APERTO ALLE PUBBLICAZIONI E AI DATI IN HORIZON 2020?

- Sviluppare la ricerca basandosi sui risultati precedenti (miglioramento della qualità dei risultati)
- Incoraggiare la collaborazione ed evitare le duplicazioni (maggiore efficienza)
- Accelerare l'innovazione (accelerazione del progresso verso il mercato per favorire una crescita più rapida)
- Partecipazione dei cittadini e della società (miglioramento della trasparenza del processo scientifico)

La Commissione europea considera fondamentale il seguente assunto: "La visione della Commissione europea è che le informazioni finanziate con fondi pubblici non debbano essere pagate nuovamente ogni volta che sono rese accessibili e utilizzabili e che le imprese e i cittadini europei ne dovrebbero beneficiare in pieno."

13ITA

Carlos Moedas, Eommissario europeo per la ricerca, la scienza e l'innovazione, ha lanciato la sfida per la ricerca in Europa: "Abbiamo bisogno di un accesso più aperto ai risultati della ricerca e ai dati collegati. La pubblicazione tramite Accesso Aperto è già oggi un requisito previsto nell'ambito dei progetti finanziati dal programma Orizzonte 2020, ma ora dobbiamo guardare seriamente ai dati aperti [...] Vogliamo rendere l'Europa aperta all'innovazione, aperta alla scienza e al mondo".

Questo auspicio possiamo assumerlo come conclusione di questa presentazione, perché ancora oggi la maggior parte dei dati è chiusa nella letteratura e nei documenti e bisogna sbloccare questi dati.

Troverete una visione per lo sviluppo della Scienza in Europa nel libro **The Three Os – Open Innovation, Open Science, Open to the World a European Open Science Cloud.**