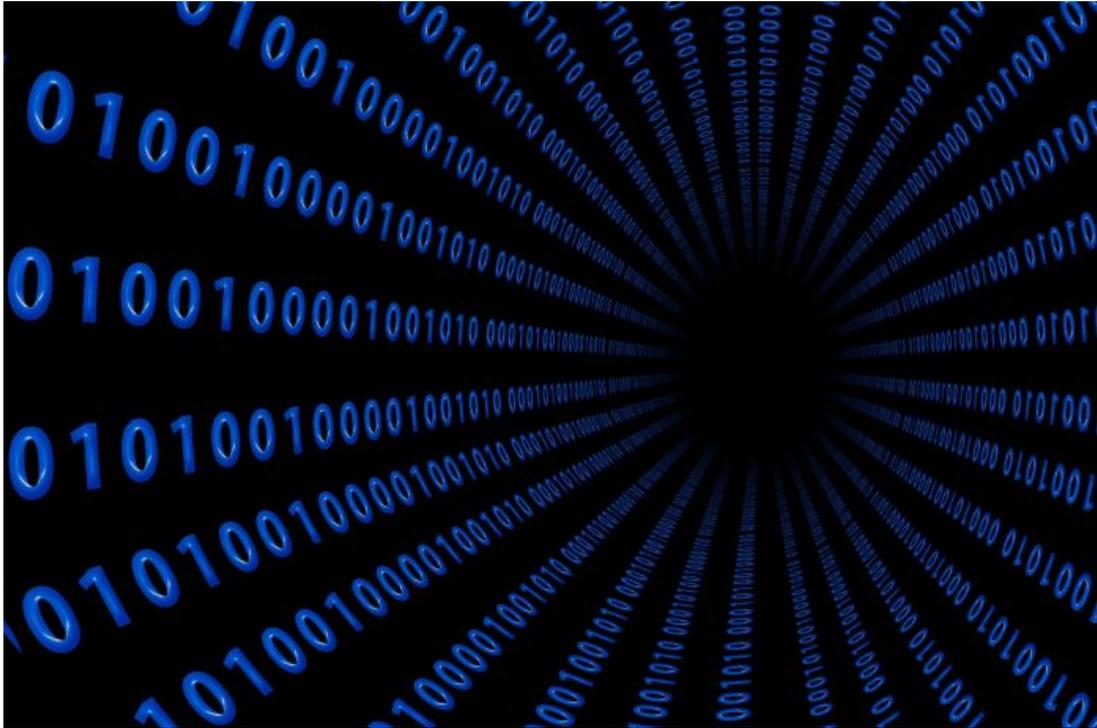


Dietro di noi un deserto digitale?

By **Federico Ruggieri** - 20/11/2019



*"Dietro di noi un deserto digitale, un altro Medioevo. **Se tenete a una foto, stampatela**".* così si raccomandava **Vint Cerf**, uno dei padri di internet in occasione di uno dei meeting della American Association for the Advancement of Science. Aggiungendo: *"A meno che il concetto di **conservazione digitale** non entri alla svelta nei nostri cervelli"*.

Monito che fa riflettere molto e da cui vorrei partire per condividere con voi quello che è emerso, insieme a qualche mia riflessione, nel panel sulla **conservazione e riuso dei beni digitali** che si è tenuto durante la Conferenza **GARR** di quest'anno, a cui sono intervenuti Elena Borgi dell'Accademia delle Scienze, Massimo Cocco di INGV, Walter Ricciardi dell'Università Cattolica del Sacro Cuore e Riccardo Smareglia di INAF.

Effettivamente il concetto di conservazione digitale è un argomento molto complesso perché negli ultimi anni ci siamo trovati d'avanti a scenari completamente nuovi. Noi stessi produciamo un'enorme mole di dati tracciabili, con l'Internet of Things anche gli oggetti producono dati, non esiste più documento analogico senza il suo corrispettivo digitale e, grazie ai dispositivi mobili, tutti questi dati sono immediatamente accessibili.

Ed ecco che si aprono diversi interrogativi a cui dare una risposta: **come conserviamo tutti questi dati** affinché possano essere accessibili e riutilizzabili oggi e negli anni a venire?

A ciascuna disciplina la sua digitalizzazione

A seconda delle discipline di ricerca il problema ha varie sfaccettature e criticità.

Le biblioteche per prime si trovano a fronteggiare un cambiamento davvero enorme, a cui non erano del tutto preparate. **I libri digitalizzati sono cresciuti in maniera esponenziale**, ma mancano procedure condivise, tanto care ai bibliotecari di una volta, che permetterebbero una fruizione agevole e condivisa da tutti e la stessa gestione delle immagini e dei metadati presenta ancora molte criticità. Inoltre le biblioteche si trovano ad operare scelte importanti su cosa digitalizzare per non entrare in concorrenza con leader dell'industria come Google books ed essere invece in grado di valorizzare ciò che sono le loro unicità in quanto depositarie di documenti dal valore storico inestimabile.

Ulteriore complessità è dovuta ad un **pubblico che sta cambiando nel tempo** dato che, con l'abbondanza dei dati in rete, il principale pubblico di riferimento delle biblioteche ovvero i ricercatori, oggi tendono ad utilizzare sempre meno le fonti primarie delle biblioteche e sempre più quanto già rielaborato da altri negli ultimi anni.

Questo porta alla necessità di **reinventarsi, immaginare e individuare nuovi pubblici di riferimento**, puntando ad esempio su percorsi personalizzati e fruizioni agevolate per aree tematiche o fasce d'età.

Nella **sanità**, un progetto integrato di digitalizzazione potrebbe davvero avere un impatto dirompente nelle nostre vite quotidiane. Se pensiamo, infatti, alla possibilità di poter conservare in digitale tutta la storia clinica di ciascun cittadino, emerge chiaramente come l'analisi di tutti questi dati, adeguatamente supportata dalle modernissime tecniche di machine learning che abbiamo a disposizione, porterebbe a prevedere lo stato di salute della persona negli anni a venire così da consentire di attuare nei tempi giusti misure di prevenzione per evitare futuri stati di malessere. Ovviamente ciò si tradurrebbe in un netto miglioramento della qualità della vita della persona e dei suoi familiari oltre ad un risparmio notevole di risorse pubbliche. È evidente che in questo caso la grande sfida che si presenta è quella di garantire la privacy del paziente.

Altre discipline scientifiche, come ad esempio **l'astronomia**, di contro non sono legate al GDPR e questo comporta una maggiore facilità nel trattare dati pubblici e privati. Resta comunque il problema della proprietà intellettuale, in quanto pur utilizzando fondi pubblici, la proprietà intellettuale resta dello scienziato. Come fare quindi per garantirla dando la possibilità ad altri di utilizzare questi dati? Ancora, le scienze della Terra hanno una tradizione di lunga durata sulla raccolta dei dati, ad esempio nella sismologia i dati relativi ad uno specifico terremoto possono essere condivisi in accesso aperto e anonimo e in tempo reale da ogni scienziato in qualsiasi parte del mondo, ma le criticità in questo caso sono legate sia alla governance, che garantisce la gestione condivisa di tutta la catena di generazione del dato scientifico (dall'acquisizione, alla qualificazione e all'archiviazione attraverso controllo di qualità e standard comuni), sia agli aspetti legali ed economici associati alla long-term preservation. Tutti problemi che si traducono con un'unica parola: [sostenibilità](#).

Il problema della sostenibilità

Una volta digitalizzato, dove si conserva il dato? Gli stessi supporti tecnologici cambiano, sono soggetti a degrado e ad obsolescenza tecnologica e questo richiede un'attività periodica di registrazione dei dati su supporti nuovi. La stessa creazione dei metadati (dalla definizione della sua provenienza e tracciabilità per dare il giusto riconoscimento a chi ha prodotto il dato fino alla definizione delle licenze da mettere sul dato stesso) è un processo che richiede tempo e risorse economiche.

E ancora, **a chi spetta il compito di conservare questi dati nei millenni?** Le biblioteche potrebbero essere il soggetto più adeguato a farlo ma occorrono diverse figure professionali perché a seconda della disciplina occorrono diverse competenze per la creazione dei metadati. È importante però essere consapevoli che, sebbene non ci sia adesso la soluzione per i prossimi 500 anni, abbiamo il dovere di conservare i dati per le generazioni future perché un dato è in grado di rappresentare un fenomeno nella sua unicità, quale può essere un terremoto o l'esplosione di una supernova o un accadimento storico e potrà essere usato in futuro per creare scienza e conoscenza. Dati su cui dovrà essere sempre operato un controllo di qualità per non incorrere nel rischio di minare le basi della riproducibilità di un esperimento sulla quale si regge tutta la nostra cultura scientifica. Tutto questo potrà essere fatto solo se non cediamo alla tentazione di guardare principalmente agli accessi al dato in termini quantitativi e considerare la key performance indicator il parametro essenziale da cui farci guidare nelle nostre scelte più importanti.

EOSC per la scienza aperta

Nell'aprile 2016 la Commissione europea ha lanciato una nuova iniziativa per il [Cloud europeo per la scienza aperta](#) che va proprio in questa direzione e che, come si auspica, metterà a disposizione un [ambiente](#) federato per l'archiviazione, la gestione, l'analisi e il riutilizzo dei dati della ricerca.

A livello italiano GARR anima il tavolo di [lavoro](#) ICDI (Italian Computing and Data Infrastructure), che coinvolge i rappresentanti di alcune tra le principali Infrastrutture di Ricerca e Infrastrutture Digitali italiane con l'obiettivo di promuovere sinergie a livello nazionale al fine di ottimizzare la partecipazione italiana a EOSC.

Riprendendo le parole di Massimo Cocco, Executive Director di EPOS ERIC e membro dell'Executive Board di ICDI: *"EOSC può rappresentare un'importante opportunità per sostenere la scienza aperta e la condivisione di dati e prodotti scientifici attraverso servizi sostenibili. Affinché questo avvenga, è fondamentale che si riesca a preservare il lavoro svolto negli ultimi decenni dalle infrastrutture di ricerca europee e nazionali per la raccolta e la standardizzazione di dati e metadati e ad armonizzare il piano di costruzione pan-europeo con le priorità e le strategie nazionali. A tal proposito Il tavolo di lavoro ICDI (Italian Computing and Data Infrastructure), è stato creato anche per contribuire a questi due obiettivi e far in modo di avere come Paese un ruolo chiave nella sua definizione. Senza dimenticarci che, per avere dei **dati FAIR (Findable, Accessible, Interoperable, Re-usable)** ovvero rintracciabili, accessibili, interoperabili e riusabili, occorre sì la condivisione di principi ma soprattutto la condivisione di pratiche basate sulla consapevolezza delle risorse e dei tempi necessari per affrontare e vincere queste sfide."*

Facebook Comments

Federico Ruggieri

Laureato in Fisica e dirigente di ricerca dell'INFN, è direttore del GARR dal 2015. Ha lavorato nel campo della Fisica delle particelle partecipando a diversi esperimenti al CERN e ai Laboratori Nazionali INFN di Frascati. I suoi interessi di ricerca sono da sempre concentrati sugli aspetti del calcolo, delle reti di trasmissione dati e dei sistemi di acquisizione e processamento dei dati. È stato Chairman del Comitato Europeo per il Calcolo nella Fisica delle Alte Energie (HEPCCC) e Direttore del Centro di calcolo INFN-CNAF. Tra i pionieri della tecnologia Grid e Cloud, ha avviato il primo progetto europeo DataGRID ed ha lavorato per portare le infrastrutture digitali nei Paesi del Mediterraneo, del Medio Oriente ed in Cina. Attualmente è membro del Consiglio direttivo della rete della ricerca europea Géant. Appassionato di tecnologie audio e fumetti, ha una collezione di albi bonelliani di tutto rispetto.