

QUADERNI di GEOFISICA

Open Science
per il futuro della Ricerca europea



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

148

Direttore Responsabile

Silvia MATTONI

Editorial Board

Luigi CUCCI - Editor in Chief (luigi.cucci@ingv.it)
Raffaele AZZARO (raffaele.azzaro@ingv.it)
Christian BIGNAMI (christian.bignami@ingv.it)
Mario CASTELLANO (mario.castellano@ingv.it)
Viviana CASTELLI (viviana.castelli@ingv.it)
Rosa Anna CORSARO (rosanna.corsaro@ingv.it)
Domenico DI MAURO (domenico.dimauro@ingv.it)
Mauro DI VITO (mauro.divito@ingv.it)
Marcello LIOTTA (marcello.liotta@ingv.it)
Mario MATTIA (mario.mattia@ingv.it)
Milena MORETTI (milena.moretti@ingv.it)
Nicola PAGLIUCA (nicola.pagliuca@ingv.it)
Umberto SCIACCA (umberto.sciacca@ingv.it)
Alessandro SETTIMI (alessandro.settimi1@istruzione.it)
Andrea TERTULLIANI (andrea.tertulliani@ingv.it)

Segreteria di Redazione

Francesca DI STEFANO - Referente
Rossella CELI
Barbara ANGIONI

redazionecen@ingv.it

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

© 2014 INGV Istituto Nazionale di Geofisica e Vulcanologia
Rappresentante legale: Carlo DOGLIONI
Sede: Via di Vigna Murata, 605 | Roma



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

QUADERNI di GEOFISICA

Open Science
per il futuro della Ricerca europea

*Open Science
for the future of European Research*

Grazia Pavoncello¹, Mario Locati², Stefano Cacciaguerra³

¹MIUR | Ministero dell'Istruzione, dell'Università e della Ricerca

²INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Milano

³INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna

Accettato 21 settembre 2018 | Accepted 21 September 2018

Come citare | How to cite Pavoncello G. et al., (2018). Open Science per il futuro della Ricerca europea. Quad. Geofis., 148: 1-60.

In copertina Kirigami rappresentativo della condivisione dati tra realtà scientifiche europee | Cover Kirigami representative of data sharing between European scientific realities

148

INDICE

Riassunto	7
Abstract	7
Introduzione	8
1. I cardini dell'Open Science	9
1.1 Open Science: definizione e opportunità	9
1.2 Dati prodotti dalla ricerca scientifica	11
1.3 Integrità ed etica scientifica (<i>Research Integrity and Ethics</i>)	12
1.4 Politiche di gestione dei dati (<i>Legal framework 1/2</i>)	13
1.5 Tipologie e compatibilità tra le licenze (<i>Legal framework 2/2</i>)	15
1.6 Terminologie condivise e collegamenti tra i dati (<i>Taxonomy, Metadata, Ontologies</i>)	17
1.7 Riconoscimento dei risultati della Ricerca (<i>Identification, citation</i>)	18
1.8 Tracciabilità (<i>Traceability, Provenance, Lineage</i>)	20
1.9 Standard per l'interoperabilità di dati e servizi (<i>Interoperability</i>)	21
1.10 Conservazione e sostenibilità (<i>Preservation, Sustainability</i>)	21
1.11 Infrastrutture tecnologiche	23
2. Infrastrutture Tecnologiche di supporto alla Ricerca	24
2.1 Analisi delle principali infrastrutture tecnologiche nazionali	25
2.1.1 GARR - la rete nazionale della ricerca	25
2.1.2 CINECA - il centro di supercalcolo nazionale	25
2.1.3 CNAF - la grid nazionale	26
2.2 Analisi delle principali e-Infrastructure	26
2.2.1 GÉANT - la rete della ricerca europea	27
2.2.2 PRACE ed EuroHPC per il supercalcolo europeo	27
2.2.3 EGI - la Grid europea	28
2.2.4 EUDAT - la gestione delle banche dati europee	28
3. Analisi dei casi studio	29
3.1 Standing Working Group on Open Science & Open Innovation	29
3.2 European Plate Observing System (EPOS)	30
3.3 European Multidisciplinary Seafloor and water-column Observatory (EMSO)	33
3.4 Emilia-Romagna Big Data Community	35
3.4.1 Il Mandato della Regione Emilia-Romagna per la creazione della piattaforma	36
3.4.2 Applicazioni della piattaforma	37
4. Le sfide dell'Open Science	37
4.1 Nuovi approcci per il coinvolgimento, valutazione e incentivazione	37
4.2 Nuove professionalità	39
4.3 Nuove soluzioni per la sostenibilità	41
4.4 Ricerca, Università e Pubblica Amministrazione	42
4.5 European Open Science Cloud (EOSC)	43
4.6 Modelli di business delle e-Infrastructure (intervista a Sanzio Bassini)	46



Conclusioni	48
Bibliografia	49
Sitografia	55

Open Science per il futuro della Ricerca europea

Questo lavoro presenta i risultati di una ricerca condotta dagli autori nell'ambito delle attività della quarta edizione del "Master in Management of Research, Innovation and Technology" tenutosi presso la Fondazione MIP del Politecnico di Milano da febbraio 2016 a luglio 2017. Oggetto dell'indagine è la mappatura del sistema complesso che ruota attorno al nuovo paradigma dell'Open Science e al mondo dei Big Data, con particolare riferimento alla realtà italiana e alle attività in cui è coinvolto l'Istituto Nazionale di Geofisica e Vulcanologia. Il master, oltre a nozioni teoriche e pratiche per la gestione dell'innovazione e del trasferimento tecnologico, ha offerto l'opportunità di confrontarsi con persone che si occupano di questi argomenti all'interno di realtà istituzionali nazionali ed europee come università, istituti di ricerca e aziende private. Gli autori del presente testo sono uno specchio di questa eterogeneità, con una persona che si occupa di infrastrutture di ricerca a livello ministeriale, un tecnologo INGV che si occupa della gestione e valorizzazione di banche dati sismologiche e un altro tecnologo INGV che si occupa della gestione di un Centro Elaborazione Dati (CED) per il supporto e lo sviluppo di servizi informatici per la vulcanologia, la geofisica e l'oceanografia. Le diverse competenze hanno permesso di indagare aspetti complementari dell'Open Science e dei Big Data sotto punti di vista diversi, toccando aspetti normativi, scientifici, gestionali, informatici e infrastrutturali. Ciascun aspetto è stato analizzato tramite un'approfondita ricerca bibliografica, l'esperienza personale degli autori maturata nelle loro attività professionali e grazie al supporto di esperti del settore, cercando di volta in volta di individuare opportunità e criticità, e tentando una mediazione continua tra linguaggi e punti di vista a volte molto diversi. Il risultato è un'analisi interdisciplinare che potrà offrire spunti di riflessione e, si auspica, possa aiutare a permettere una più efficiente pianificazione delle attività.

Open Science for the future of European Research

This paper illustrates the results of a research conducted during a Master's degree entitled "Master in Management of Research, Innovation and Technology", organised by the Graduate School of Business, run by the Politecnico di Milano, between February 2016 and July 2017. The research goal was to map the complex system supporting the "Open Science" paradigm and the so-called "Big Data", with a particular focus on the Italian research environment, where INGV (Istituto Nazionale di Geofisica e Vulcanologia) operates. In addition to transferring the necessary expertise on the management of the innovation and the technology transfer, the Master's degree offered the unique chance to encounter and exchange experiences with people operating at different levels, fields and environments, such as universities, research centres, ministries, and private companies. The three authors of this paper reflect the heterogeneous audience of the Master's courses, with a lawyer working at the Italian Ministry of Education, Universities and Research active in the field of European Research Infrastructures, and two technologist working at INGV, one managing seismological data, and the other working as a chief engineer of the data centre in Bologna. The -very- different backgrounds of the three authors, allowed the exploration of the Open Science paradigm and the

use of the Big Data in modern Science, mastering regulatory, scientific and technological aspects. To achieve a complete mapping, the authors underwent an extensive bibliographical search, and took advantage of their respective experience matured in their professional activities. The result is a comprehensive, interdisciplinary analysis, meant to give food for thought and clear suggestions to decision makers, supporting them in a more proactive, efficient, planning.

Introduzione

Parallelamente alla realizzazione dei quattro perni del *Single Market*, ovvero la libera circolazione 1) delle persone, 2) dei servizi, 3) delle merci e 4) dei capitali, il programma delle principali istituzioni europee (i.e. Consiglio, Commissione, e Parlamento) e dei governi nazionali auspica il consolidamento del *Digital Single Market*. Il *Digital Single Market* si basa a) sul libero accesso a servizi e prodotti online, b) l'incentivazione della crescita massiccia delle reti di comunicazione e dei servizi innovativi basati su di essi, e c) il supporto alla crescita di un'economia basata sul digitale. I nuovi strumenti e le nuove infrastrutture del *Digital Single Market* saranno funzionali anche alla realizzazione dell'*European Research Area*, una piattaforma che aumenterà l'attrattività e la competitività della Ricerca Scientifica condotta in Europa. La realizzazione della *European Research Area* mira alla libera circolazione della conoscenza [Potočnik, 2007] identificata come la quinta libertà, che andrà ad aggiungersi alle quattro su cui si basa il *Single Market*. Per queste finalità, è stata ad esempio posta come condizione ai progetti H2020 la pubblicazione in forma di *Open Data* dei risultati ottenuti.

L'*Open Science*, il nuovo paradigma di Scienza 2.0, promuove un nuovo approccio alla ricerca basato su una più veloce condivisione dei dati per favorire nuovi scenari di Ricerca che promette di essere più efficiente, efficace, responsabile, trasparente ed inclusivo. La Ricerca guidata dai dati (*data driven*) sfrutterà pienamente l'utilizzo di moderne infrastrutture di *Cloud Computing* e di analisi di *Big Data*, favorendo attività collaborative e interdisciplinari per processare dati provenienti anche da discipline diverse fra loro. La multidisciplinarietà diventa una delle sfide di questo scenario e l'interoperabilità tra servizi e dati ne permetterà una piena attuazione. Per poter raggiungere questi obiettivi è importante considerare tutti i processi che gravitano intorno al dato, per crearlo, mantenerlo, condividerlo, tracciarne l'uso e riconoscerne la paternità.

La tematica complessa dell'*Open Science* è oggetto di politiche mirate del Ministero dell'Istruzione dell'Università e della Ricerca, una tematica che investe "ambiti, competenze e problemi assai vasti e articolati che coinvolgono direttamente la governance sia degli Enti di ricerca e delle Università sia del Ministero stesso; coinvolgono le politiche della ricerca, le scelte strategiche nell'istruzione e formazione, la valutazione, i modelli e le politiche dell'informazione e comunicazione scientifica" [MIUR, 2016].

Scopo della ricerca presentata in questo testo è quello di identificare i processi necessari per l'attuazione dell'*Open Science* e l'analisi delle loro interconnessioni, sia a livello nazionale che europeo, individuando le possibili azioni che il sistema italiano dovrebbe adottare per essere in linea con le migliori prassi internazionali valorizzando al contempo le proprie specificità. Per arrivare al risultato atteso ci siamo avvalsi di un'ampia ricerca bibliografica, dell'analisi delle fonti normative vigenti, si è partecipato a progetti, conferenze e seminari nazionali ed europei come lo *Standing Working Group (SWG) on Open Science and Innovation*, l'*European Plate Observing System (EPOS)*, l'*European Multidisciplinary Seafloor and water-column Observatory (EMSO)*, l'*Emilia-Romagna Big Data Community*, e lo *European Open Science Cloud (EOSC)* e infine si sono intervistate personalità rappresentative delle infrastrutture di ricerca europee come Massimo Cocco di EPOS, Laura Beranzoli di EMSO e Sanzio Bassini di PRACE. Per la realizzazione dell'*Open Science* le istituzioni europee hanno stabilito di integrare le infrastrutture tecnologiche esistenti tra di loro, per questo nella ricerca sono state analizzate le infrastrutture nazionali come GARR, CINECA e CNAF che partecipano a quelle paneuropee come GÉANT, PRACE, EGI e EUDAT.

Il nuovo scenario prospettato dall'*Open Science* necessita oltre che di un diverso approccio culturale, anche di nuovi metodi e strumenti per la valorizzazione delle carriere, di nuove professionalità, di nuovi modelli di sostenibilità per le infrastrutture e di una soluzione all'annoso problema dell'appartenenza di Università ed Enti Pubblici di Ricerca alla Pubblica Amministrazione che, con i suoi vincoli, limita lo sviluppo competitivo della ricerca italiana, tutte problematiche che dovranno essere affrontate dal progetto *European Open Science Cloud pilot* (EOSC). Il programma Horizon 2020 offre infatti per la realizzazione dell'EOSC e lo sviluppo delle tematiche relative come *European Cloud Initiative*, *European Data Infrastructure*, *High Performance* e *Quantum Computing*, uno stanziamento di circa 2 miliardi di euro, a fronte di un contributo degli Stati Membri di circa 4,7 miliardi di Euro. In una prima fase, il piano previsto per l'attuazione dell'EOSC, vedrà il coinvolgimento della sola componente scientifica, solo successivamente sarà inclusa quella privata al fine di creare il mercato unico digitale.

La ricerca condotta dimostra come la realizzazione e l'adozione dello scenario prospettato dalla *European Research Area* e dall'*Open Science* sia ancora in una fase embrionale, e tale rimarrà fino a che non verranno individuate delle valide soluzioni di *governance* senza cui non si potrà procedere a uno sfruttamento effettivo delle potenzialità offerte dal binomio *Big Data* e *Cloud Computing*, il cui potenziale, secondo diversi pareri autorevoli, dovrebbe dare un vantaggio competitivo all'economia di tutta l'Unione Europea.

La tesi che presentiamo fornisce una panoramica organica sufficientemente semplificata e documentata, tale da permettere di individuare quali siano i nodi ancora da sciogliere, e, di conseguenza, dove è necessario concentrare i maggiori sforzi di intervento.

1. I cardini dell'Open Science

Le tre principali istituzioni europee, la Commissione, il Consiglio e il Parlamento Europeo, hanno intrapreso da tempo la costruzione dell'ERA, l'*European Research Area* [Comunicazione CE COM(2000) 6], consolidandone le basi sia con accordi tra gli Stati Membri dell'Unione Europea (Trattato di Lisbona del 2007) che successive comunicazioni della Commissione. ERA, definita come "*uno spazio di ricerca unificato aperto al mondo e fondato sul mercato interno, nel quale i ricercatori, le conoscenze scientifiche e le tecnologie circolano liberamente e grazie al quale l'Unione e gli Stati membri rafforzeranno le loro basi scientifiche e tecnologiche, nonché la loro competitività e la loro capacità di affrontare collegialmente le grandi sfide*" [Comunicazione CE COM(2012) 392]. L'avvento di più efficienti strumenti collaborativi ha permesso di ipotizzare nuovi approcci alla ricerca scientifica che ben si adattano alle finalità di ERA, e che potrebbero dare un vantaggio competitivo al Sistema della Ricerca nel suo complesso a livello Europeo oltre a una potenziale ricaduta economica sulla società [EDP, 2015; 2017].

La ricerca condotta ha permesso di ridurre l'enorme complessità dei processi in corso ad alcuni cardini su cui verte l'attuazione dell'Open Science. Lo scopo è quello di cercare di capire il grande piano che sta dietro a un'impressionante quantità di attività e documenti che si intrecciano tra loro, attività condotte da un altrettanto impressionante quantità di soggetti diversi che operano a tutti i livelli, dalla singola università o istituto di ricerca, a network di istituzioni, ciascuno con una sua rilevanza, da quella strettamente locale, alla nazionale, all'europea o internazionale.

1.1 Open Science: definizione e opportunità

L'*Open Science* è un nuovo approccio all'attività di ricerca che promuove la condivisione della conoscenza scientifica in ogni sua forma attraverso le nuove tecnologie digitali e i nuovi strumenti collaborativi [Nielsen, 2011]. Questo nuovo approccio impatta sui tempi e sui modi

che caratterizzano il tradizionale ciclo della ricerca [Cox A.M. and Ting Tam, 2018], in cui le pubblicazioni scientifiche svolgono ancora un ruolo fondamentale e in cui è marginale l'accesso ai dati oggetto delle pubblicazioni. I nuovi strumenti tecnologici interattivi orientati alla collaborazione permettono la condivisione immediata e a basso costo di ogni passaggio e risultato dell'attività scientifica, promuovendo la verificabilità e riproducibilità delle tesi presentate [Thanos, 2017; Stewart Lowndes et al., 2017] e lo sviluppo di interazioni anche al di fuori del proprio ristretto ambito disciplinare [OECD, 2015].

La semplicità con cui oggi è possibile accedere alla produzione scientifica ha come effetto l'abbassamento delle barriere all'ingresso, allargando la base di persone che potrebbe potenzialmente contribuire con idee, relazioni e servizi, generando così nuovi modelli di lavoro, nuove relazioni sociali e modificando in profondità il rapporto tra Ricerca e Società. Fare Ricerca interdisciplinare sarà molto più semplice, non solo tra discipline appartenenti allo stesso ambito scientifico (**disciplinarietà interna**), ma anche tra discipline di ambiti molto diversi tra loro (**disciplinarietà esterna**) [ESFRI, 2016].

Le istituzioni preposte all'indirizzo politico, economico e sociale dell'Unione Europea, hanno colto le potenzialità offerte dal paradigma dell'*Open Science* [EC DG-Research, 2016 a; b], e, forti del primato nella produzione di dati scientifici a livello globale, vorrebbero innalzarne impatto ed efficacia. Tramite l'aggregazione, i risultati della ricerca scientifica forniscono informazioni cruciali per finalizzare decisioni strategiche in ambito scientifico e tecnologico, nella proposizione e valutazione di nuovi progetti, nelle attività di educazione e formazione, nelle decisioni sugli investimenti o sulla scelta degli investitori da coinvolgere.

Nonostante la crescente semplicità nella condivisione dei dati e l'indubbio vantaggio per la società, molti ricercatori tendono a non condividere i propri dati. Una interessante indagine condotta dall'editore Wiley [2014] ha evidenziato le principali ragioni per cui i ricercatori sono pro o contro la condivisione dei dati, il risultato dell'inchiesta è riportato in Tab.1.

Ripensare l'attuale sistema della Ricerca Europea non significa solo mettere a disposizione nuovi strumenti, ma significa anche coinvolgere attivamente tutti i soggetti del settore. Già nel 2015 EGI, l'*European Grid Infrastructure* (vedi art. 2.2.3), comprendendo le necessità e la complessità di un sistema che supporti l'*Open Science*, ha provato a identificare le principali caratteristiche necessarie alla sua piena attuazione definendo le cosiddette **Open Science Commons** [EGI, 2015], ovvero la descrizione del *framework* basato su quattro pilastri: a) **i dati**, b) le **e-Infrastructure**, c) gli **strumenti scientifici**, e d) il **sapere scientifico**. Secondo EGI questi quattro pilastri dovrebbero rispettare sei principi per permettere l'*Open Science*: 1) **Shared resources**, tutti i quattro pilastri dovrebbero condividere le stesse risorse, semplificando il coordinamento degli sforzi, 2) **Access rights**, permettere l'accesso in modo indiscriminato a ciascun pilastro a qualunque membro dell'ERA, 3) **Policies**, la politica di accesso e uso dovrebbe essere la stessa per tutti i pilastri, 4) **Management**, la gestione amministrativa dovrebbe essere trasparenti e finalizzati a mantenere l'accesso e la qualità nel lungo periodo, 5) **Governance**, la politica di gestione dovrebbe coinvolgere direttamente tutte le parti coinvolte, dai finanziatori alle comunità di ricercatori, i gestori delle infrastrutture tecnologiche regionali, nazionali ed europee, 6) **Stewardship**, sistemi e soluzioni che possano garantire la sostenibilità finanziaria adeguata a permettere una pianificazione di lungo termine.

Nel 2016 la *League of European Research Universities* (LERU), ha reso pubbliche delle linee guida [LERU, 2016] per realizzare una serie di buone pratiche di gestione dei dati della ricerca, che, nonostante non facciano esplicito riferimento alle *Open Science Commons*, ne applicano lo spirito. A riprova della convergenza delle visioni dei diversi soggetti europei coinvolti che individuano come prioritaria la creazione di un *framework* di base a supporto delle infrastrutture di ricerca nell'ottica *Open Science*, anche nella *roadmap* dello *European Strategy Forum on Research Infrastructures* [ESFRI, 2016] viene ripreso questo concetto chiamandolo **e-Infrastructure Commons**.

Motivi a favore della condivisione		Motivi a sfavore della condivisione	
57%	La condivisione dei dati è diffusa nella comunità	42%	Problemi con proprietà intellettuale o riservatezza
55%	Incrementa l'impatto e la visibilità della ricerca	36%	Il finanziatore o l'istituzione non richiede di condividere i dati
50%	Pubblica utilità	26%	Paura che la ricerca venga copiata
42%	Requisito della rivista scientifica	26%	Paura che i dati siano interpretati e/o usati male
37%	Trasparenza e riutilizzo	23%	Preoccupazioni etiche
30%	Fiducia in chi richiede i dati	22%	Preoccupazione relativa a ottenere una citazione/attribuzione corretta
25%	Rendere trovabili e accessibili i dati	21%	Non si sa dove condividere i dati
23%	Requisito del finanziatore	20%	Risorse e/o tempo insufficienti
18%	Requisito della mia istituzione	16%	Non si sa come condividere i dati
13%	Richiesta di libertà d'informazione	12%	Non ci si sente in dovere di condividere
13%	Conservazione del dato	12%	I dati non vengono considerati importanti
2%	Altro	11%	Mancanza di fondi
		7%	Altro

Tabella 1 Inchiesta sui motivi a favore o contro la condivisione dei dati [Wiley, 2014].

Table 1 Survey on the pros and cons of the data sharing [Wiley, 2014].

1.2 Dati prodotti dalla ricerca scientifica

Una **definizione di dato** chiara e largamente condivisa non esiste, pertanto si evidenzieranno alcune alternative. Si riporta qui la definizione di dato tratta dall'Enciclopedia di Filosofia dell'Università di Stanford [Floridi, 2015]: *“Diaphoric Definition of Data - A datum is a putative fact regarding some difference or lack of uniformity within some context within the real world, between two physical states (science) or between two symbols (humanities – linguistics). Data are neutral – they have no meaning without context. Data are relational entities, information can consist of different types of data, there can be no information without data representation and data can have meaning independently of whoever reports it”*.

La seguente è la definizione di *dati della ricerca scientifica* adottata dal programma H2020 [EC DG-Research, 2017a]: *“Research data refers to information, in particular facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form”*.

Infine, la *Research Data Alliance* (RDA), l'organizzazione di riferimento per la trattazione dei dati della ricerca scientifica, adotta la definizione riportata dal dizionario del *Consortia Advancing Standards in Research Administration Information*, che richiama la definizione di Landry et al. [1973]: *“Facts, measurements, recordings, records, or observations about the world collected by scientists and others, with a minimum of contextual interpretation. Data may be in any format or medium taking the form of writings, notes, numbers, symbols, text, images, films, video, sound*

recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, work flow charts, equipment descriptions, data files, data processing algorithms, or statistical records”.

Un’alternativa, definito dal Research Information Network, propone di classificare i dati della ricerca attraverso lo scopo per cui sono stati generati e secondo le procedure seguite:

- **da osservazioni** (*observational*): i dati sono raccolti in tempo reale e sono generalmente insostituibili, come quelli registrati dai sensori, da campagne di indagine, o da campioni;
- **sperimentali** (*experimental*): dati da strumenti di laboratorio, spesso riproducibili, a volte ad alto costo, come sequenze di geni, cromatogrammi, bobine per campi magnetici toroidali;
- **da simulazioni** (*simulation*): dati generati da modelli, dove i modelli e i metadati sono solitamente più importanti dei dati ottenuti, come quelli da modelli climatici o economici;
- **derivati o compilati** (*derived, compiled*): dati riproducibili e costosi, come quelli ottenuti dalle procedure di data mining su big data, dalla compilazione di un database o dalla realizzazione di un modello 3D;
- **riferimenti o canonici** (*reference, canonical*): una raccolta (statica od organica) di insiemi di dati più piccoli oggetto di *peer-review*, solitamente pubblicati e revisionati, come sequenze di geni, archivi di dati, strutture chimiche o portali di dati geografici.

Una trattazione più completa di sistemi di classificazione dei dati è fornita da Schöpfela [2017] che col termine dato identifica semplici fatti. Quando i dati vengono processati, organizzati, strutturati o presentati e contestualizzati così da poterne fare uso, divengono **informazione**. Solitamente i dati sono codificati in modi leggibili dai computer, mentre le informazioni sono solitamente codificate in un testo discorsivo. Al fine dell’attuazione dell’*Open Science*, sono necessari i cosiddetti **Open Data** [Open Definition] che dovrebbero auspicabilmente essere pubblicati dai ricercatori rispettando i “*FAIR data Principles*” [Force11, 2014a; Wilkinson et al., 2016; Hodson et al., 2018a, 2018b] che prevedono dati 1) **recuperabili**, 2) **accessibili**, 3) **interoperabili** e 4) **riutilizzabili**.

1.3 Integrità ed etica scientifica (*Research Integrity and Ethics*)

Al fine di condividere all’interno di ERA un comune approccio alle buone pratiche della ricerca scientifica, la Commissione Europea ha pubblicato la “**Carta europea dei ricercatori**” [Raccomandazione CE 2005/251/EC], un documento importante che formalizza pratiche di buon senso e regole non scritte. Sono fissati **12 principi riguardanti l’attività dei ricercatori**: libertà di ricerca, principi etici, responsabilità professionale, comportamento professionale, obblighi contrattuali e legali, responsabilità finanziaria, buona condotta nel settore della ricerca, diffusione e valorizzazione dei risultati, impegno verso l’opinione pubblica, rapporti con i supervisor, doveri di supervisione e gestione, sviluppo professionale continuo.

In questo contesto, sono evidenziati i **19 principi per i datori di lavoro e i finanziatori**: 1) riconoscimento della professione, 2) non discriminazione, 3) ambiente di ricerca, 4) condizioni di lavoro, 5) stabilità e continuità dell’impiego, 6) finanziamento e salari, 7) equilibrio di genere, 8) sviluppo professionale, 9) valore della mobilità, 10) accesso alla formazione alla ricerca e alla formazione continua, 11) accesso all’orientamento professionale, 12) diritti di proprietà intellettuale, 13) coautore, 14) supervisione, 15) insegnamento, 16) sistemi di valutazione, 17) reclami e ricorsi, 18) partecipazione agli organismi decisionali, 19) assunzione.

Anche *All European Academies* (ALLEA), un’organizzazione europea nata nel 1994 che vede più di 60 università europee consorziate, mantiene aggiornato un documento dedicato alle

buone pratiche della Ricerca [ALLEA, 2017] che è un riferimento nei *Grant Agreement* per i progetti H2020.

Secondo questo documento le buone pratiche dovrebbero essere vincolate a quattro principi:

1. **Affidabilità** nell'assicurare la qualità della ricerca, affidabilità che si rispecchi nella progettazione, nelle metodologie adottate, nelle procedure di analisi e nell'uso delle risorse disponibili;
2. **Onestà** nello sviluppo, nelle iniziative, nelle revisioni, nella documentazione e nella comunicazione della ricerca, attività portate avanti in modo trasparente, equo, completo e imparziale;
3. **Rispetto** per i colleghi, per chiunque partecipi alle attività di ricerca, della società, dell'ecosistema, del patrimonio culturale e dell'ambiente;
4. **Responsabilità** in tutte le fasi della ricerca, dall'ideazione alla pubblicazione, nella sua gestione e amministrazione, nell'insegnamento, nella supervisione e nel tutoraggio, prendendo in considerazione il potenziale impatto dei risultati ottenibili.

1.4 Politiche di gestione dei dati (*Legal framework 1/2*)

In riferimento alle politiche vigenti di gestione dei dati, è auspicabile avviare una riflessione in merito a come rimodellare la disciplina del diritto d'autore affinché promuova la libera circolazione della conoscenza e dell'innovazione all'interno del mercato comune. Ogni eventuale restrizione a questa libera circolazione dovrebbe venire adeguatamente giustificata e motivata. Sotto questo punto di vista, la realtà italiana è poco soddisfacente e spesso si rilevano limitazioni ingiustificate alla conoscenza. La situazione a livello nazionale non imputabile solamente alla normativa vigente in materia di diritto d'autore e purtroppo il movimento dell'*Open Access* italiano non ha la forza necessaria a definire né tantomeno a risolvere tutte le imperfezioni di mercato [Vezzoso, 2008].

Per migliorare la situazione nazionale non si tratterebbe "semplicemente" di intervenire ex-post in caso di violazione del diritto all'accesso ai dati scientifici, ma di indirizzare fin da subito le scelte di politica ritenute più opportune in materia di proprietà intellettuale. Una soluzione è stata proposta dal legislatore con l'art.52 comma 2 del D.Lgs. 7/3/2005 n.82 che recita: "*I dati e i documenti che le amministrazioni titolari pubblicano, con qualsiasi modalità, senza l'espressa adozione di una licenza di cui all'articolo 2, comma 1, lettera h), del decreto legislativo 24 gennaio 2006, n. 36, si intendono rilasciati come dati di tipo aperto ai sensi all'articolo 68, comma 3, del presente Codice, ad eccezione dei casi in cui la pubblicazione riguardi dati personali del presente Codice. L'eventuale adozione di una licenza di cui al citato articolo 2, comma 1, lettera h), è motivata ai sensi delle linee guida nazionali di cui al comma 7*".

Dalla norma, si evince che **i dati pubblicati dalla PA senza una specifica licenza sono considerati Open Data**. Tuttavia questa soluzione, per quanto auspicabile, non ha trovato adeguata applicazione. Sarebbe, pertanto, doverosa una modifica nella normativa del diritto d'autore, al fine di disciplinare l'obbligatorietà di apporre una licenza. Un segnale positivo giunge dall'accordo per la libera condivisione delle pubblicazioni scientifiche e dei dati, siglato nel 2013 tra CRUI e i rappresentanti degli EPR [CRUI, 2013a].

Specifiche situazioni in cui una licenza obbligatoria può essere istituita sono settate nella legislazione di ogni sistema brevettuale e variano da sistema a sistema.

Alcuni esempi in cui una licenza obbligatoria può essere concessa includono:

- invenzioni finanziate dal governo;
- fallimento o impossibilità di un titolare di soddisfare una domanda per un prodotto brevettato e in cui il rifiuto di concedere una licenza porta alla incapacità di sfruttare un importante progresso tecnologico, o di sfruttare un ulteriore brevetto.

La normativa vigente prevede che una licenza obbligatoria debba essere preceduta da una richiesta al titolare del brevetto volta ad ottenere una cosiddetta licenza volontaria e comunque, una volta stabilita, deve essere soggetta a adeguati ricompensi al titolare del brevetto. Tale *modus operandi* risulta **solo parzialmente estendibile all'ambito dei dati scientifici in considerazione del fatto che da una parte il dato, non generando valore economico direttamente, non giustifica in alcun modo l'onerosità prevista dalle procedure di brevettazione, dall'altra invece l'obbligatorietà della licenza favorirebbe l'Open Access**, in quanto garantirebbe il riconoscimento scientifico in capo all'autore (o "creator"). L'attuazione dell'*Open Access* dovrebbe anche poter passare attraverso un processo sistematico, in cui tutti i soggetti coinvolti siano rivestiti di specifiche responsabilità: legislatore, governo, soggetti finanziati con fondi pubblici e ricercatori. Senza investimenti economici e organizzativi, senza lo sviluppo di una cultura rivolta all'apertura della conoscenza scientifica (che richiede innanzitutto impegno sul piano della divulgazione e della formazione) e senza regolamentazioni di dettaglio, il cammino dell'*Open Access* rischia di arrestarsi o di rallentare sempre più. Infatti, si rende necessario avviare un processo che consenta alla politica di metter mano alle norme formali, non dimenticando però che il definitivo successo dell'*Open Access* passa attraverso un mutamento radicale della comunità scientifica e accademica, un mutamento che è prima di tutto etico e investe le norme informali che governano la ricerca. In particolare, risulta imprescindibile:

- inserire l'*Open Access* nel processo di anagrafe e valutazione della ricerca;
- porre nei bandi del MIUR (Prin e Furb) l'obbligo di pubblicazione *Open Access* all'interno di appositi archivi istituzionali;
- regolamentare a livello di istituzione finanziata, gli obblighi di deposito e pubblicazione sugli archivi *Open Access* rendendoli compatibili con il diritto d'autore;
- attuare la Raccomandazione UE con riferimento allo sviluppo e incoraggiamento di riconoscimenti in termini di carriera ai ricercatori che sposino la cultura dell'*Open Science* nonché di nuovi indicatori e criteri che valorizzino le caratteristiche delle pubblicazioni in *Open Access*;
- elaborare una politica istituzionale di apertura dei dati scientifici.

L'intervento dei promotori dell'*Open Access* si è dipanato lungo molteplici strade. A questo riguardo si è soliti distinguere, tra due diverse strategie volte ad assicurare l'accesso aperto alla conoscenza scientifica (Fig. 1):

- **Green open access** (*self-archiving*), quando il ricercatore provvede a depositare il proprio articolo già pubblicato (in genere dopo il periodo di embargo imposto dall'editore) o il manoscritto finale *peer-reviewed* presso gli archivi telematici della ricerca
- **Gold open access** (*open access publishing*), quando l'editore pubblica immediatamente l'articolo, una volta accettato, in maniera aperta, ovvero visibile a tutti gratuitamente.

L'Italia ha creato un profondo scostamento rispetto a quanto indicato dalle Raccomandazioni UE che parlano espressamente della necessità di una pianificazione finanziaria. Si tratta di uno dei punti più deboli dell'intervento legislativo sui quali occorre intervenire in senso correttivo. La via praticata a livello sistemico da Università ed Enti Pubblici di Ricerca italiani è al momento la *Green open access*.

Dal 2012, la **Commissione Europea ha incoraggiato gli Stati membri a rendere pubblici i risultati della ricerca finanziata pubblicamente** [Direttive CE 2012/417/EU; 2013/37/EU], incrementando l'*Open Access*. L'obiettivo della politica europea per l'*Open Access* risiede nell'ottimizzazione dell'impatto della ricerca finanziata pubblicamente, sia per il Settimo programma quadro (FP7), sia di Horizon 2020 che a livello di singoli stati membri. L'*Open Access* contribuisce, infatti, a rafforzare l'impatto economico e la competitività europea attraverso la diffusione della conoscenza, nel rispetto dei diritti di proprietà intellettuale, della sicurezza e

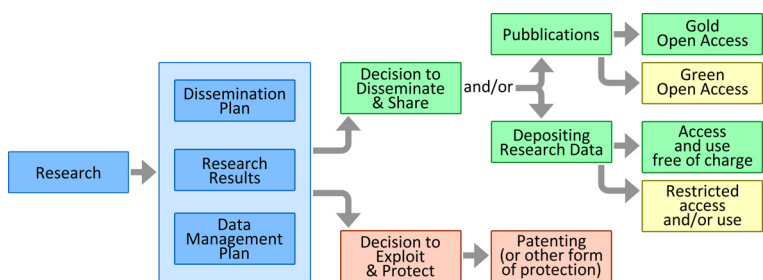


Figura 1 Schema che rappresenta i possibili approcci gestionali alla disseminazione e utilizzo delle pubblicazioni scientifiche e i dati della Ricerca (EC DG-Research 2017a).

Figure 1 Open access to scientific publication and research data in the wider context of dissemination and exploitation (EC DG-Research 2017a).

della privacy. Un maggiore accesso alle pubblicazioni e ai dati scientifici permette di:

- costruire sulla base di precedenti risultati di ricerca (migliore qualità dei risultati);
- incoraggiare la collaborazione ed evitare di duplicare gli sforzi (maggiore efficienza);
- accelerare l'innovazione (un accesso più rapido al mercato significa una crescita più veloce);
- coinvolgere i cittadini e la società (maggiore trasparenza del processo scientifico).

Carlos Moedas¹, commissario europeo per la Ricerca, la Scienza e l'Innovazione, ha recentemente affermato che entro il 2020 le pubblicazioni scientifiche dovranno essere accessibili a tutti gratuitamente e che gli editori dovranno trovare nuove forme di business. L'appello vale come un orientamento politico per i 28 Stati membri, concordi sul rendere accessibili gratuitamente e riutilizzabili i risultati della ricerca scientifica finanziata pubblicamente. In esito all'introduzione di tali politiche è necessario che:

- sia assicurato il prima possibile *Open Access* alle pubblicazioni prodotte nell'ambito di attività di ricerca finanziate con fondi pubblici;
- i sistemi di concessione in licenza contribuiscano ad assicurare in maniera equilibrata *Open Access* alle pubblicazioni scientifiche prodotte nell'ambito di attività di ricerca finanziate con fondi pubblici, fatta salva la legislazione applicabile sul diritto d'autore e nel rispetto della stessa e incoraggino i ricercatori a mantenere il diritto d'autore pur concedendo licenze agli editori;
- il sistema delle carriere in Università ed EPR sostenga e premi i ricercatori che aderiscono a una cultura di condivisione dei risultati delle proprie attività di ricerca, in particolare, assicurando *Open Access* alle loro pubblicazioni e ai dati nonché sviluppando, incoraggiando e utilizzando nuovi modelli alternativi di valutazione delle carriere, nuovi criteri di misurazione e nuovi indicatori;
- sia migliorata la trasparenza, in particolare, informando il pubblico in merito agli accordi conclusi tra enti pubblici ed editori per la diffusione dell'informazione scientifica. A questo riguardo, dovrebbero essere inclusi gli accordi riguardanti le offerte cumulative di abbonamenti che permettono di accedere sia alla versione elettronica, sia alla versione stampata delle riviste a prezzo scontato.

1.5 Tipologie e compatibilità tra le licenze (*Legal framework 2/2*)

Sulle problematiche legate al diritto d'autore, cessione dei diritti e tipologie di licenze esiste una letteratura sterminata. Al fine di semplificare e standardizzare le tipologie di licenze utilizzate dagli

¹ Speech by Commissioner Carlos Moedas in Amsterdam, NL: "Open science: share and succeed", Amsterdam, 4 April 2016. http://europa.eu/rapid/press-release_SPEECH-16-1225_en.htm

autori, nel 2001 nacquero le licenze *Creative Commons* (CC) ad opera di Lawrence Lessig della *Stanford Law School*.

Le licenze CC, nate per tutelare opere poste al di fuori del mondo scientifico, sono state efficacemente trapiantate dal 2005 anche nel settore delle pubblicazioni scientifiche e rappresentano uno strumento trasversale al rapporto tra l'autore, l'editore e gli utenti finali.

La novità del modello introdotto con le CC è dovuto a una **integrazione tra il modello intermedio e il copyright tradizionale, per cui "tutti i diritti sono riservati", e il "public domain" per cui "nessun diritto riservato"**. Le CC sono una leva potente per la definizione a monte dei diritti e la diffusione dell'*Open Science*, poiché è l'autore a stabilire i termini di utilizzo del proprio lavoro, senza cedere il controllo di tali aspetti agli editori. A supporto di quanto appena affermato, nelle linee guida per i progetti H2020 si dice infatti che *"the Commission encourages authors to retain their copyright and grant adequate licences to publishers. Creative Commons offers useful licensing solutions. This type of licence is a good legal tool for providing open access in its broadest sense"* [EC DG-Research, 2017a]. Nei limiti della normativa vigente, la scelta della specifica tipologia di licenza in base alla quale concedere il proprio documento resta nella discrezionalità dell'autore. Ai fini della libera condivisione dei dati in un'ottica di *Open Science*, **sarebbe proficuo che gli EPR e le Università emanassero delle linee guida in cui vengano meglio specificati i meccanismi di apposizione della licenza da parte dei pubblici dipendenti**, e qualche primo esempio è oggi disponibile [EPOS, 2016; INGV, 2016; 2017a]. In caso contrario, il rischio paventato, sarebbe quello di rendere problematica la *governance* nella gestione dell'*Open Science* a livello istituzionale, a causa dell'eventuale frammentazione e incompatibilità tra le licenze apposte (Tab. 2). In questo senso, è utile citare i principi e le linee guida applicative in materia di interoperabilità legale proposte da RDA [RDA-CODATA, 2016].

Sul versante delle riviste *Open Access*, si possono citare come esempi, l'esperienza della *Public Library of Science* e quella di *BioMed Central*. Entrambe queste riviste hanno adottato licenze CC che richiedono la sola citazione della fonte, liberalizzando l'utilizzo e la redistribuzione delle pubblicazioni. **I vantaggi riconoscibili derivanti dalla promozione della logica di *Open Access* alla conoscenza scientifica sono individuabili in una serie di effetti positivi di cui si avvantaggerebbero tutti i soggetti coinvolti**: gli autori, i quali riscuoterebbero un maggior impatto sul pubblico, poiché i loro contributi, liberamente disponibili sui siti web, sono accessibili da un ventaglio potenzialmente illimitato di utenti; i lettori degli articoli scientifici, i quali hanno in tal modo un accesso illimitato e gratuito alle pubblicazioni di loro interesse, al di là delle limitazioni contrattuali e tecnologiche imposte dalle licenze proprietarie, nonché le stesse istituzioni che finanziano la ricerca le quali otterrebbero un maggior sfruttamento dei risultati delle scoperte scientifiche e delle ricerche da esse stesse sostenute e, quindi, un più rapido progresso della scienza, reale beneficio, questo, per tutta la collettività.

Sembra quindi di tutta evidenza come finora sia mancato all'appello della «chiamata all'*Open Access*» il legislatore, il quale non ha ancora provveduto ad introdurre nella normativa sul diritto d'autore norme *ad hoc* (fermo restando il D.Lgs. 24/01/2006 n.36 che recepisce la Direttiva 2003/98/CE), preso atto della specificità delle esigenze proprie alla ricerca scientifica che favoriscano il raggiungimento di un nuovo e più equilibrato bilanciamento tra gli interessi della comunità scientifica e quello di coloro che sfruttano economicamente le pubblicazioni scientifiche.

L'approccio *top-down*, sicuramente indispensabile, deve necessariamente essere supportato e coadiuvato da un approccio *bottom-up*, cioè da strumenti normativi privati, quali, ma non solo, i contratti: si pensi agli Addenda ai contratti di edizione, alla *License to Publish* e, infine, alle *University License*, che si propongono quali modelli di un'efficiente ed equilibrata allocazione di interessi ai quali autori ed editori possano aderire. I modelli contrattuali menzionati sono un segnale positivo della diffusa consapevolezza delle esigenze sottese all'*Open Access*, portando con sé il vantaggio di consentire ai singoli la scelta –tra tutte le opzioni possibili– di quella che meglio soddisfa le proprie esigenze e favorendo, attraverso una fisiologica competizione tra modelli, l'emersione di quello migliore; per altro, vi è il rischio che l'attuale proliferazione degli stessi strumenti negoziali pregiudichi

la loro effettività, creando incertezza tra gli utenti, innalzando i costi transattivi e disperdendo il significativo coefficiente di potere contrattuale che essi sono in grado di attribuire agli autori.

	Public domain	CC0	CC BY	CC BY-SA	CC BY-NC	CC BY-ND	CC BY-NC-SA	CC BY-NC-ND
Public domain	✓	✓	✓	✓	✓	✗	✓	✗
CC0	✓	✓	✓	✓	✓	✗	✓	✗
CC BY	✓	✓	✓	✓	✓	✗	✓	✗
CC BY-SA	✓	✓	✓	✓	✗	✗	✗	✗
CC BY-NC	✓	✓	✓	✗	✓	✗	✓	✗
CC BY-ND	✗	✗	✗	✗	✗	✗	✗	✗
CC BY-NC-SA	✓	✓	✓	✗	✓	✗	✓	✗
CC BY-NC-ND	✗	✗	✗	✗	✗	✗	✗	✗

Tabella 2 Matrice di compatibilità tra le diverse tipologie di licenze Creative Commons.

Table 2 Compatibility matrix of various Creative Commons licenses.

1.6 Terminologie condivise e collegamenti tra i dati (*Taxonomy, Metadata, Ontologies*)

Ogni ramo della ricerca scientifica adotta una terminologia coerente al suo interno. Purtroppo, molti termini sono in comune tra diverse discipline, nonostante il significato ad essi associato sia spesso diverso. Al fine di risolvere eventuali problemi di incomprensione è necessaria l'adozione di una **tassonomia condivisa**, ovvero una classificazione e un ordinamento della terminologia utilizzata e associata a un significato univoco. *Research Data Alliance* (RDA) ha istituito un gruppo di lavoro denominato *Data Foundation and Terminology* che si occupa di affrontare queste problematiche e che ha stilato delle raccomandazioni sulla costruzione di archivi per gestire la terminologia [RDA, 2015].

Se la tassonomia è soprattutto utile alla comprensione tra le persone che lavorano sul piano interdisciplinare, i **metadati sono utili nel momento in cui è necessario cercare e processare i dati con l'ausilio di strumenti informatici**, strumenti sempre più in auge grazie all'enorme quantità di dati disponibili. La compilazione di metadati è fondamentale per permettere a chiunque di poter accedere correttamente, utilizzare, comprendere ed eventualmente processare i dati, utilizzare un servizio o eseguire una procedura codificata. Lo spirito con cui approcciare la compilazione di metadati è guidato dal fine ultimo di massimizzare l'usabilità di ciò che viene descritto.

Esistono innumerevoli standard per la codifica di metadati, ma al fine di ridurre le barriere tra le diverse discipline scientifiche, è bene adottare standard riconosciuti a livello internazionale e

interdisciplinare. Grazie all'adozione di questi standard si favorisce una maggiore condivisione tra chi produce e chi usa il dato o servizio, indipendentemente dal tipo di tecnologia utilizzata per il trasferimento dell'informazione. A livello europeo lo standard di riferimento per la codifica dei metadati relativi alla ricerca scientifica è il *Common European Research Information Format* (CERIF). Questo standard è basato su un modello di dati relazionali per l'archiviazione e l'interoperabilità di informazioni relative alla ricerca nel suo complesso, e permette la descrizione di un'ampia gamma di aspetti, tra cui le persone coinvolte, le istituzioni, i finanziatori, le pubblicazioni, i set di dati, i brevetti, i prodotti della ricerca, aspetti relativi alla misurazione delle performance (*bibliometrics, impact factor*), strumenti di laboratorio, e molti altri ancora.

L'importanza dello standard CERIF è testimoniato da uno studio commissionato dal Parlamento Europeo [STOA, 2014] finalizzato a stabilire il grado di fattibilità per la realizzazione di un sistema europeo capace di raccogliere e monitorare le prestazioni dei dati della ricerca. Lo studio conclude con la seguente risposta affermativa: *"A European Integrated Research Information e-Infrastructure is technically feasible and in full alignment with the current policy context in the European Union. The development of this e-infrastructure should not constitute a major technical endeavour, thanks to the recent technological developments and especially the maturity of the European CERIF standard"*.

Oltre all'ausilio di metadati, è auspicabile che chi rende disponibili dati o servizi renda disponibili anche delle pubblicazioni scientifiche che descrivano più o meno estesamente le metodologie adottate e il contesto in cui è stata condotta la ricerca. La disponibilità di questi testi descrittivi permette agli utenti di cogliere quegli aspetti procedurali che non possono essere codificati nei metadati, favorendo la consapevolezza degli utenti finalizzata ad un utilizzo corretto dei dati. Parallelamente, l'esigenza di contestualizzare i dati e i relativi metadati viene affrontata in ambito informatico dalle **ontologie** e dalla **semantica** [Berners-Lee 1998; 2001; Shadbolt et al., 2006] pensate per modellare i collegamenti, concetti e definizioni tra loro. La duttilità dei sistemi che integrano dati associati ad ontologie forniscono le basi per fare ricerche e stabilire connessioni ben più evolute e complesse di quelle possibili con semplici strumenti basati ad esempio sulle chiavi di ricerca. L'ostacolo principale all'adozione delle ontologie è la complessità di gestione che richiede grandi energie in termini di tempo e umane per la creazione, il "riempimento" e l'aggiornamento di questi modelli.

1.7 Riconoscimento dei risultati della Ricerca (*Identification, citation*)

Negli ultimi due decenni si sta assistendo alla crescita esponenziale di nuovi contenuti raggiungibili tramite Internet. Di pari passo con l'aumento della quantità, stanno aumentando anche le tipologie di contenuti e, in particolare, di dati recuperabili, basti solo pensare al fenomeno dei cosiddetti *"Big Data"* e alla varietà derivante dal lento, ma inesorabile processo di digitalizzazione del patrimonio documentale storico. È facile intuire l'esigenza sottesa alla identificazione sicura e certificata del livello di attendibilità dei dati, e l'identificazione della fonte dei dati permetterà di dividere **fonti autoritative** da **fonti non certificate**.

L'**identificativo** è un **codice univoco** che viene associato a un elemento, rendendone recuperabile il contenuto. Un esempio è il codice fiscale, un identificativo che, per fini fiscali, associa in maniera univoca soggetti fisici o giuridici a un codice alfanumerico sin dal 1973. L'accezione informatica si è storicamente sviluppata nell'ambito delle banche dati, in cui è fondamentale riuscire a trovare un'informazione in modo veloce e sicuro tra milioni di altre, per poi disporre l'utilizzo nei modi più disparati tracciandone il percorso nel flusso delle operazioni. Un fondamentale aspetto nell'uso degli identificativi è di essere svincolati dalla posizione del dato a cui fanno riferimento. È sufficiente mantenere aggiornata l'associazione tra dato e la sua posizione per mantenere funzionante il sistema.

Assegnare un identificativo risolvibile sul Web che sia persistente e univoco ai dati, permette a ricercatori di **comunicare in modo non ambiguo i dati utilizzati** (si pensi, ad esempio, alla redazione di un articolo scientifico), contribuendo così alla trasparenza, riproducibilità e verificabilità delle procedure di ricerca. Identificativi persistenti, univoci e risolvibili sono componenti fondamentali per il meccanismo delle citazioni scientifiche, in quanto eliminano l'ambiguità relativa ai riferimenti utilizzati, rendendo più robusto il conteggio delle citazioni collegate e le relative metriche che valutano l'impatto della ricerca. Tra la corposa letteratura dedicata sull'argomento, una buona panoramica e guida all'uso disciplinare degli identificativi è fornita da Richards et al. [2011], ANDS [2011] e Archer et al. [2013] e da McMurry et al. [2017] da cui è tratta la Fig. 2 che fornisce una intuitiva indicazione sulle regole di assegnazione degli identificativi in relazione svolto nella creazione del dato.

YOUR CONTRIBUTION TO CONTENT	YOUR ROLE	CREATE NEW ID vs REUSE EXISTING
ORIGINAL CONTENT YOU GENERATED or AUTHORED	THE AUTHOR	CREATE
CONTENT DEPOSITED TO YOUR CARE by the GENERATOR	THE GUARDIAN	
CONTENT YOU CURATED	THE CURATOR	* *
MEANINGFULLY DIFFERENT, or EXPANDED CONTENT	THE ANNOTATOR	
INTEGRATIONS of EXISTING DATA, COMBINED in NOVEL WAYS, or at DIFFERENT LEVELS of GRANULARITY	THE INTEGRATOR	
FACTUAL CORRECTIONS, IMPROVEMENTS	THE CONTRIBUTOR	* *
IDENTICAL CONTENT or TRIVIALY CHANGED, AGGREGATED for INDEXING or SEARCH	THE INDEXER	
CONTENT COPIED to CIRCUMVENT DEPENDENCY or to UNIFY USER EXPERIENCE	THE APPLICATION PROVIDER	REUSE

Figura 2 Relazione tra il livello di contributo alla generazione del dato, riferita alle modalità di assegnazione di identificativi persistenti associati a quel dato [McMurry et al., 2017].

Figure 2 Contributions and roles related to content as they correspond to identifier creation versus identifier reuse [McMurry et al., 2017].

Un sistema solido di citazioni e collegamenti nel ciclo della ricerca scientifica assicura anche la dovuta attribuzione dei meriti e della paternità dei singoli elementi che compongono un sistema complesso, supportando anche sistemi di riconoscimento e finanziamento.

Tra gli innumerevoli sistemi di identificativi diffusi in ambito scientifico ne citiamo tre, l'*Handle System*, un sistema generico di identificativi usato per molte diverse tipologie di oggetti digitali, il *Digital Object Identifier (DOI)* [Paskin, 2010], basato su *Handle System* e che viene associato a pubblicazioni e set di dati o banche dati, l'*Open Researcher and Contributor ID (ORCID)* che viene associato alle persone fisiche ed è adottato da ANVUR per la VQR [ANVUR, 2015]. Gli identificativi DOI sono rilasciati da agenzie riconosciute dalla *DOI Foundation*. Ciascuna è specializzata in determinate aree di competenza e ha sviluppato degli schemi di metadati per descrivere il proprio ambito. In questa sede, si segnalano l'agenzia privata *Crossref* gestita dalla *Publishers International Linking Association Inc.*, che si occupa della gestione di pubblicazioni e di enti finanziatori (*Funder Registry*), e *DataCite* [Starr & Gastl, 2011; Starr et al., 2014] che si occupa di assegnare DOI e relativi metadati a set di dati.

Force11, una comunità autocostituita di insegnanti, bibliotecari, archivisti, editori e finanziatori, ha stilato una dichiarazione congiunta dei **principi della citazione dei dati** [Force11, 2014b], identificando 8 punti:

1. **Importanza.** I dati devono essere considerati a tutti gli effetti prodotti della ricerca e come

tali devono essere citabili. I dati devono essere considerati alla stessa stregua di altri prodotti come le pubblicazioni;

2. **Riconoscimento e attribuzione.** La citazione dei dati deve facilitare il riconoscimento e la legittima attribuzione della paternità a tutti coloro che hanno contribuito ai dati, riconoscendo che non esiste un unico metodo di attribuzione può essere adatto a tutte le tipologie di dati;
3. **Evidenza.** Nelle pubblicazioni scientifiche, ovunque e comunque un'affermazione si basa su dati, i dati corrispondenti devono essere chiaramente citati;
4. **Identificazione univoca.** La citazione dei dati deve includere un sistema persistente per l'identificazione univoca dei dati, un sistema che sia utilizzabile dalle macchine, univoco a livello internazionale, e largamente diffuso nella comunità;
5. **Accesso.** Le citazioni di dati devono semplificare l'accesso ai dati stessi, ai metadati relativi, alla documentazione, codice, o altri materiali che permettano un uso consapevole e appropriato, sia per le macchine, sia per le persone;
6. **Persistenti.** Gli identificativi univoci e i metadati che descrivono i dati e le relative disposizioni, devono essere disponibili a tempo indefinito, un tempo che superi persino il ciclo di vita dei dati stessi;
7. **Specificità e verificabilità.** La citazione dei dati deve poter semplificare l'identificazione, l'accesso, e la verifica di attendibilità di affermazioni basate sui dati utilizzati. Le citazioni, e i relativi metadati, dovrebbero includere informazioni sulla provenienza ed essere così stabili da permettere processi di verifica che garantiscano di trattare gli stessi identici dati usati per la costruzione delle affermazioni su di essi basati, indipendentemente dalla granularità, versione o periodo temporale di riferimento;
8. **Interoperabilità e flessibilità.** I metodi usati per la citazione di dati dovrebbero essere sufficientemente flessibili per adattarsi alle esigenze delle diverse comunità scientifiche, senza per questo differenziarsi a tal punto da compromettere l'interoperabilità tra le diverse pratiche di citazione adottate da ciascuna comunità.

1.8 Tracciabilità (*Traceability, Provenance, Lineage*)

Il processo scientifico si basa sulla riproducibilità e sull'affidabilità delle fonti, pertanto è fondamentale che venga tenuta traccia di tutte le procedure seguite nell'elaborazione dei dati. Anche il grande pubblico si sta accorgendo dell'importanza di riuscire a tracciare la provenienza delle informazioni e le elezioni presidenziali americane del 2016 hanno contribuito alla sensibilizzazione al problema [Allcott & Gentzkow, 2016]. Grande enfasi sulla provenienza dei dati e servizi dovrebbe essere parte integrante delle procedure di ricerca, e codificate sotto forma di metadati, e la loro compilazione, se opportunamente considerata in fase di progettazione del lavoro, può essere gestita automaticamente da strumenti predisposti allo scopo che aggiungono via via i metadati descrittivi delle operazioni svolte [Simmhan et al., 2005]. Esempi della struttura di sistemi che gestiscono questo tipo di dati si può trovare in Mayernik et al. [2013] e le linee guida da seguire specificatamente per i dati delle Scienze della Terra sono disponibili da GEO [2015]. Altri elementi che descrivano aspetti diversi dalle procedure condotte quali persone e ruoli ricoperti, sono solitamente gestiti manualmente, o nei casi più complessi, anch'essi gestiti in modo automatico da strumenti appositi. Al fine di fornire agli utenti finali una descrizione completa sulla provenienza, i metadati dovrebbero ereditare (includendoli o collegandoli) i metadati associati a tutti i dati o servizi utilizzati come input. **Il tracciamento completo delle procedure, la reputazione delle persone coinvolte e una descrizione esaustiva delle incertezze associate a ogni aspetto rilevante, sono tutti elementi che contribuiscono a definire il livello di qualità, e chiariscono agli utenti finali il livello di affidabilità raggiungibile**

utilizzando quanto fornito. Quanto descritto nei metadati deve poter essere accessibile dagli utenti, e diventare parte integrante del processo di ricerca. Poniamo ad esempio che venga descritto l'utilizzo di una procedura basata su uno strumento che non è pubblicamente accessibile: in questo caso è impossibile stabilire la bontà della procedura e in quanto salta la condizione di riproducibilità della procedura. **Un vantaggio nel codificare le procedure in maniera standardizzata, permetterebbe ad altri ricercatori di poter sfruttare l'esperienza già maturata da altri e riadattandola alle proprie esigenze** [Moreau, 2010]. Inoltre, il tracciamento è fondamentale per l'individuazione di eventuali problemi che generano errori nell'output, e se il tracciamento comprende anche gli input utilizzati, si aumentano le chance di sistemare problemi nell'intera filiera della ricerca permettendo poi di rigenerare un output senza errori.

1.9 Standard per l'interoperabilità di dati e servizi (*Interoperability*)

L'interoperabilità è da sempre un tema importante in ambito scientifico vista la necessità di poter scambiare tra diversi gruppi i risultati delle proprie ricerche ed è per questo strettamente collegato al tema dei metadati (vedi par. 1.6). Con il crescere della complessità dei sistemi digitali, è aumentata la complessità delle tipologie di soluzioni possibili per risolvere lo stesso problema, **creando un numero di "standard" per la codifica dei dati e dei servizi enorme, spesso confinati ad ambiti disciplinari molto ristretti**. L'esigenza di poter condurre ricerche interdisciplinari fa sì che la jungla di standard tra loro incompatibili diventi una barriera invalicabile. Esistono diverse organizzazioni che si occupano di affrontare il problema della interoperabilità, a livello internazionale come la *Research Data Alliance* (RDA) e, in ambito europeo, il progetto OpenAIRE (FP7 e H2020). Sia RDA che OpenAIRE promuovono delle linee guida che sono diventate un riferimento in ambito scientifico, per l'interoperabilità a **livello tecnologico** e a **livello semantico**, cioè metodi che permettono di associare un senso ai dati a prescindere dalla disciplina permettendo così lo scambio. Lo standard di metadati di riferimento per la ricerca supportato dalla Commissione Europea è il *Common European Research Information Format* (CERIF), mentre il *World Wide Web Consortium* (W3C) supporta il *Data Catalog Vocabulary* (DCAT). L'Unione Europea ha intrapreso un processo di armonizzazione a livello normativo, semantico e tecnologico dalla fine degli anni '90 all'interno del programma *Interoperability solutions for European Public Administrations* (ISA) dal 2010 al 2015 e nel suo prosieguo denominato ISA² [Decision (EU) 2015/2240]. ISA ha istituito il cosiddetto **European Interoperability Framework** [EIF; EC DG-Informatics 2017], che è il tramite tra le implementazioni del *National Interoperability Framework* (NIF) a livello di ciascuno stato membro e dei vari *Domain Interoperability Frameworks* (DIF).

1.10 Conservazione e sostenibilità (*Preservation, Sustainability*)

L'accesso ai dati della ricerca è strettamente legato alle modalità con cui questi dati vengono archiviati e gestiti. Uno dei principali problemi che impedisce l'accesso ai dati scientifici è dovuto al fatto che i ricercatori tendono a non condividerli tramite archivi istituzionali o disciplinari, poiché spesso preferiscono gestirli personalmente, spesso mantenendoli su hard disk, o altri sistemi di archiviazione personali. Tale comportamento, dovuto a una serie di fattori quali la maggiore flessibilità, l'esigenza di controllo o una più banale pigrizia, rischia di sfociare in una potenziale e irrimediabile perdita di dati che può acuirsi col passare del tempo. Molti sono i fattori che entrano in gioco: i dispositivi potrebbero degradarsi risultando illeggibili, si potrebbero perdere informazioni sui sistemi di codifica, i formati potrebbero diventare obsoleti e illeggibili con i moderni strumenti, potrebbero subentrare fattori personali quali il pensionamento o un

cambio di lavoro. Per ridurre la perdita dei dati, le PA devono organizzarsi per creare archivi istituzionali e obbligando i dipendenti a utilizzarli, archivi che devono essere “*predisposti in modo tale da garantire l'accesso aperto, libero e gratuito, dal luogo e nel momento scelti individualmente, l'interoperabilità all'interno e all'esterno dell'Unione Europea e la conservazione a lungo termine in formato elettronico*” (Decreto Legge 8/8/2013, n.91, Art.4, Comma 2), tutte indicazioni presenti anche nella raccomandazione della Commissione Europea del 17 luglio 2012 sull'accesso all'informazione scientifica e sulla sua conservazione (2012/417/UE). La **conservazione** di dati digitali richiede quindi di intraprendere diverse azioni, partendo da una attenta pianificazione calibrata alla propria realtà, fino a prevedere le azioni da intraprendere nell'eventualità che l'archivio possa chiudere [DPC, 2015]. L'esigenza della conservazione ha portato alla redazione delle linee guida certificate ISO 14721:2012 e definite nel “*Open Archival Information System (OAIS) reference model*” [CCSDS, 2012; Lavoie, 2014].

Gli archivi che seguono questo modello sono “*Trusted Digital Repositories*” (TDRs):

- hanno come missione quella di fornire accesso e preservare i dati rispondendo a una esigenza istituzionale o conformemente alle esigenze dell'ambito scientifico di riferimento;
- garantiscono un'attività pianificata che assicuri la conservazione del proprio contenuto;
- permettono il riuso dei dati nel tempo, assicurandosi che i metadati associati siano aggiornati e soddisfino l'esigenza di un futuro utilizzatore.

La costruzione di un piano per la conservazione dei dati si basa sulla programmazione di aspetti organizzativi, tecnologici e di sostenibilità e in questo senso si fa riferimento al bilanciamento delle tre gambe di uno sgabello [*“three-legged stool”* - Kenney & McGovern, 2003].

Se sul fronte tecnologico molte soluzioni valide sono disponibili ed esistono standard maturi sia in ambito internazionale (ISO), europeo (vedi le norme sulla conservazione dei dati pubblici della C.E.) e nazionale (linee guida AgID), ma, spesso, vengono trascurati gli aspetti legati alla **sostenibilità** di lungo periodo dell'infrastruttura che rende accessibili i dati. Trascurare l'aspetto della sostenibilità ha generato in passato problemi, basti pensare alle banche dati create grazie a progetti europei che, a causa dell'assenza di una pianificazione adeguata, sono stati abbandonati, diventando dei veri e propri relitti, nonostante avessero un potenziale utilizzo di lungo periodo. La Commissione Europea [EC DG-Research, 2017c] e ESFRI [2017], coscienti dell'importanza di una corretta pianificazione della sostenibilità delle infrastrutture ritenute utili nel lungo periodo, hanno pubblicato delle linee guida e delle raccomandazioni per supportare la gestione di questo aspetto. **ESFRI in particolare ha raggruppato le raccomandazioni in 7 punti fondamentali.**

1. Stabilire e supportare l'eccellenza per tutto il ciclo di vita di una infrastruttura di ricerca mediante l'adozione di strumenti adeguati, assicurando le condizioni più adatte al contorno, e aprendo le infrastrutture al mondo.
2. Assicurarsi che le persone giuste al posto e nel momento giusto contribuiscano alle infrastrutture di ricerca, armonizzando la Ricerca e il sistema educativo nazionale affinché le professionalità siano disponibili.
3. Armonizzare e integrare una visione convergente tra le infrastrutture di ricerca e le e-Infrastructure europee al fine di ottimizzare gli investimenti e i servizi offerti alla comunità.
4. Sfruttare appieno le potenzialità delle infrastrutture di ricerca nel loro ruolo di poli di innovazione incorporando strategie del loro sviluppo nei piani nazionali ed Europei.
5. Dotarsi di strumenti capaci di determinare l'effettivo valore economico e il più ampio valore sociale delle infrastrutture di ricerca, utilizzando e inserendo i risultati di queste analisi nei dialoghi tra decisori e la società.
6. Stabilire le condizioni al contorno per una gestione efficiente e sostenibile dei finanziamenti alle infrastrutture di ricerca nel lungo periodo, una gestione che va coordinata in modo altrettanto efficace.
7. Promuovere e incentivare il coordinamento a livello nazionale ed Europeo sin dalla

progettazione e pianificazione delle infrastrutture di ricerca nazionali ed Europee al fine di incrementarne il valore strategico.

Lo strumento di riferimento a supporto della pianificazione della conservazione e della sostenibilità del dato sono i **Data Management Plan** (DMP), ovvero dei piani di gestione dei dati di lungo periodo che includono anche la descrizione di come i dati vengono creati, processati o generati, quali metodologie e standard vengono adottati e i piani di accessibilità ai dati. Da tempo obbligatori negli ambienti scientifici anglosassoni, i DMP sono diventati obbligatori anche nei progetti H2020, e per la loro compilazione esistono linee guida consolidate e strumenti online, come ad esempio quelli del *Digital Curation Centre* [DCC, 2013].

1.11 Infrastrutture tecnologiche

L'*Open Science* necessita dell'aggregazione delle esistenti infrastrutture tecnologiche, attualmente divise tra varie discipline e Stati membri, nell'ottica di rendere l'accesso più semplice, la contaminazione tra le discipline più feconda, con l'intento di creare nuove opportunità di mercato e nuove soluzioni in settori come la sanità, l'ambiente o i trasporti. Per aggregare le infrastrutture di dati scientifici e superare la frammentazione, l'assetto istituzionale dovrà promuovere i finanziamenti a lungo termine, la sostenibilità, e la conservazione dei dati. *Science Europe*, un'associazione creata nel 2011 che raccoglie le maggiori istituzioni finanziatrici e di ricerca e che è, in parte, coinvolta nelle attività di *European Research Area*, ha stilato i quattro principi (Tab. 3) che dovrebbero caratterizzare i sistemi che gestiscono i dati della ricerca per supportare l'*Open Science* [Science Europe, 2016]. Le istituzioni europee hanno deciso di costruire l'*European Data Infrastructure*, basandosi sul potenziamento di infrastrutture tecnologiche paneuropee esistenti e coinvolgendo la comunità scientifica, i finanziatori e i responsabili dell'*European Strategy Forum on Research Infrastructures* [ESFRI, 2017], l'*Infrastructure for Spatial Information in Europe* [INSPIRE, Direttiva CE 2007/2/EC], l'*e-Infrastructures Reflection Group* [e-IRG, 2017], GÉANT [GÉANT, 2017], PRACE [PRACE, 2017], ELIXIR [ELIXIR, 2017], il Forum di Belmont [Forum di Belmont, 2017] e analoghe iniziative di aggregazione disciplinare come gli *European Research Infrastructure Consortium* (ERIC) esistenti e in via di costruzione.

Flexibility	I sistemi che gestiscono i dati della ricerca dovrebbero essere sufficientemente flessibili da permettere estensioni in termini di tipologie di elementi trattati, la loro definizione, metadattazione e permettere l'utilizzo di fonti di dati esterne.
Openness	I sistemi che gestiscono i dati della ricerca dovrebbero poter essere utilizzati dall'esterno, in linea con il principio "il più possibile aperti, chiusi quanto necessario" e, in ottemperanza alla direttiva CE 2013/37/EU, il processamento di dati non dovrebbe causare la perdita dell'informazione sulla paternità sui dati utilizzati in ingresso.
FAIRness	I sistemi che gestiscono i dati della ricerca dovrebbero promuovere la trovabilità, accessibilità, interoperabilità e riutilizzabilità dei dati gestiti adottando i "FAIR data Principles" dedicati alle attività di ricerca basate sui dati.
Data entry minimisation	I sistemi che gestiscono i dati della ricerca dovrebbero ridurre al minimo la necessità di immettere dati, facilitando il riutilizzo di dati già inseriti manualmente, adottando il motto "immetti una volta, riutilizza molte volte".

Tabella 3 Principi caratterizzanti i sistemi che gestiscono i dati per rendere possibile l'*Open Science* (Science Europe, 2016).

Table 3 Core principles of data infrastructures that makes Open Science possible (Science Europe, 2016).

In questo contesto, è bene distinguere due tipologie di infrastrutture: le *Research Infrastructure* e le *e-Infrastructure*. La definizione di **Research Infrastructure** è condivisa, sia a livello europeo (ESFRI, programma H2020) sia a livello del MIUR, per cui si intendono strutture, risorse e servizi usati dalle comunità di ricerca per promuovere l'innovazione nei rispettivi settori. Queste infrastrutture possono essere utilizzate oltre che per fini di Ricerca, ad esempio per scopi educativi o di servizio pubblico. Una *Research Infrastructure* può essere composta da attrezzature scientifiche di primaria importanza o serie di strumenti, risorse basate sulla conoscenza quali collezioni, archivi o dati scientifici, infrastrutture in rete quali sistemi di dati e calcolo e reti di comunicazione e qualsiasi altra infrastruttura di natura unica ed essenziale per raggiungere l'eccellenza nella Ricerca e nell'innovazione. Secondo questa definizione, gli ERIC ricadono in questa categoria.

Viceversa non c'è una definizione chiara e condivisa di **e-Infrastructure**, anche se generalmente è diffuso come termine per indicare infrastrutture basate su aspetti informatici e non sui contenuti. ESFRI adotta la definizione coniata da e-IRG [e-IRG, 2013], secondo cui si tratta di ambienti dove è possibile condividere ricerca e risorse per l'educazione (es.: networks, computer, spazi per l'archiviazione, software, dati). Nel Programma Nazionale per le Infrastrutture di Ricerca (PNIR) 2014-2020 del MIUR, le *e-Infrastructure* sono chiamate "Infrastrutture di Ricerca Virtuali" e definite come quelle infrastrutture informatiche che realizzano super-calcolo o servizi per la ricerca interfacciandosi alle *Research Infrastructure* produttrici di dati scientifici. In letteratura [Barbera et al., 2009] le *e-Infrastructure* sono definite come quelle infrastrutture informatiche, come ad esempio le risorse basate su soluzioni ICT (*Information and Communications Technology*) come reti distribuite, computer, spazi di archiviazione e software che semplificano la collaborazione all'interno delle comunità di ricerca tramite la condivisione delle risorse, degli strumenti di analisi e dei dati. Secondo le definizioni riportate, GÉANT, PRACE o ELIXIR ricadono in questa categoria.

L'attuazione della rivoluzione *Open Science* permetterà l'analisi e lo sfruttamento delle potenzialità dei *Big Data* con un impatto fondamentale sul cambiamento di approccio della ricerca scientifica e, a cascata, sull'economia europea e la società, offrendo nuove opportunità di grandi innovazioni in campo industriale e sociale. Fare *Open Science* sarà reso possibile dalle possibilità offerte dalle soluzioni informatiche che generalmente vengono identificate come "**Cloud**", cioè soluzioni frutto della combinazione di tre elementi interdipendenti: a) le infrastrutture che gestiscono e archiviano dati, b) le reti a banda larga che li trasmettono e c) i centri di supercalcolo che li elaborano. In questo contesto, la Commissione Europea risponde alla sfida di fare *Open Science* tramite soluzioni *Cloud* con un progetto H2020 sperimentale denominato *European Open Science Cloud pilot* (vedi par. 4.5) che le consenta di mantenere la sua posizione all'interno dell'economia basata sui *Big Data*.

2. Infrastrutture Tecnologiche di supporto alla Ricerca

Dato che l'Europa è tra i maggiori produttori di dati scientifici al mondo e si prefigge di raggiungere una posizione di leadership nell'utilizzo di *Big Data* in campo scientifico, la Commissione Europea sta realizzando un piano per promuovere lo sviluppo di servizi *cloud* e infrastrutture di dati. Il piano prevede l'integrazione, il consolidamento e la federazione di infrastrutture tecnologiche europee per lo sviluppo di servizi *Cloud* per l'*Open Science* [Comunicazione CE COM(2016) 178 final]. Tutto questo è attuabile usufruendo di strutture con eccezionale capacità di calcolo, rapidità di connessione e soluzioni *cloud* ad alta capacità. In questo contesto, analizzeremo quali sono le infrastrutture tecnologiche nazionali di eccellenza che partecipano alle infrastrutture paneuropee (*e-Infrastructure*) che più si prestano alla realizzazione di servizi *Cloud* per l'*Open Science*.

2.1 Analisi delle principali infrastrutture tecnologiche nazionali

Le infrastrutture tecnologiche nazionali sono strutture, risorse e servizi utilizzati da ricercatori per accrescere le conoscenze e il grado di innovazione nel proprio settore. Comprendono attrezzature scientifiche, archivi di dati, strutture elettroniche localizzate, distribuite e virtuali, che costituiscono un'eccellenza per il settore della ricerca e dell'innovazione. In Italia, le infrastrutture tecnologiche nazionali di eccellenza sono GARR, CINECA e CNAF-INFN. Queste eccellenze contribuiscono al progresso tecnologico ed industriale del sistema Paese e partecipano alle infrastrutture paneuropee.

2.1.1 GARR - la rete nazionale della ricerca

Il consorzio GARR (Gruppo per l'Armonizzazione delle Reti della Ricerca) è l'associazione senza fini di lucro fondata da CNR, ENEA, Fondazione CRUI e INFN con il patrocinio del MIUR che progetta, realizza e gestisce l'infrastruttura di rete telematica della comunità scientifica e accademica italiana per le attività di didattica e ricerca in ambito nazionale ed internazionale [GARR, 2017]. Il GARR assume il ruolo di *National Research & Education Network* (NREN) e partecipa alla Rete paneuropea **GÉANT**. La rete GARR-X (2012) unisce oltre 1000 sedi su tutto il territorio nazionale, in modo capillare, utilizzando 8.850 km di fibra ottica di dorsale, altri 6.400 km di accesso e una banda aggregata di oltre 1900 Gbps, per connettere oltre 2 milioni di utenti fra studenti, ricercatori e docenti. In questa ottica, si articola sul territorio con oltre 70 *Point of Presence* (PoP) e un'infrastruttura in fibra ottica con capacità massima di 100 Gbit/s. È collegata alle reti di ricerca europee e mondiali in modo affidabile attraverso un doppio collegamento a 100 Gbit/s verso la rete GÉANT e con collegamenti multipli a 10 Gbit/s verso le reti di operatori commerciali come Google, Akamai, Cogent Communications e Level 3.

Oltre all'infrastruttura [Wikipedia inglese, 2017d], GARR offre una serie di servizi per la rete che vanno dalla configurazione alla gestione dei suoi apparati (*Network Operations Center*), dalla raccolta alla pubblicazione delle statistiche di traffico (*GARR Integrated Network Suite*), dalla gestione dei guasti agli incidenti di sicurezza informatica (SCAnzioni Ripetute a Richiesta), dalla registrazione di nomi di dominio (*Local Internet Registry*) all'assegnazione di indirizzi IP (*Network Information Center*), dalle *certification authority* ai servizi applicativi di news, dalla multivideoconferenza al *mirroring*. Inoltre, fornisce un servizio per l'autenticazione e l'autorizzazione in un approccio federato, denominato IDEM (*IDentity Management Authentication and Authorization Infrastructure*) in accordo con eduGAIN e il servizio per il supporto alla mobilità di Eduroam.

2.1.2 CINECA - il centro di supercalcolo nazionale

Il CINECA (Consorzio Interuniversitario del Nord-Est per il Calcolo Automatico) è il centro di supercalcolo nazionale per la ricerca scientifica [CINECA, 2017] e, grazie al *Tier-0* Marconi, è alla dodicesima posizione nel mondo [TOP500, 2016]. Al consorzio aderiscono 70 Università italiane, CNR, Crea, OGS, INDIRE, INVALSI, la Stazione Zoologica "Anton Dohrn" e il MIUR. Nell'ambito della ricerca pubblica, il supercalcolo permette ai ricercatori di affrontare le sfide scientifiche con strumenti efficaci e confrontabili con gli altri Paesi avanzati, mentre in quello industriale, permette di ottimizzare i tempi e contenere i costi necessari per lo sviluppo dei prodotti, aumentandone la qualità e la competitività nella sfida del mercato globale. Il supercalcolo non si appoggia solo su piattaforme hardware, ma anche su ambienti operativi, su librerie per la parallelizzazione dei programmi, per l'analisi delle prestazioni e per l'ottimizzazione dei codici. In questo senso, CINECA

offre alla propria utenza un ambiente integrato di servizi che ne permette una fruizione semplice ed efficace. L'infrastruttura per il supercalcolo è composta da sistemi:

- *High Performance Computing* - computer ad alte prestazioni riservati al calcolo scientifico
- Memorizzazione Dati - un insieme di apparati di memorizzazione di massa organizzati in una *Storage Area Network* e connessi in fibra ottica ai sistemi di calcolo
- Grafica e Realtà virtuale

A livello internazionale è partner di **PRACE** (*Partnership for Advanced Computing in Europe*) con l'obiettivo di realizzare un ecosistema europeo HPC. Inoltre, è membro di *European Technology Platform for HPC* che unisce imprese e enti di ricerca, per definire le priorità per lo sviluppo di un ecosistema europeo di supercalcolo competitivo a livello mondiale.

2.1.3 CNAF - la grid nazionale

Il CNAF (Centro Nazionale Analisi Fotogrammi) è il centro nazionale dell'INFN per la ricerca e lo sviluppo nelle tecnologie informatiche e telematiche che gestisce e sviluppa il *middleware* della *Grid* (il sistema di calcolo distribuito su scala geografica [CNAF, 2017]). Partecipa al consorzio *Worldwide LHC Computing Grid (WLCG)*, una collaborazione di più di 150 centri di elaborazione in 40 paesi collegati a *Grid* nazionali ed internazionali, che ospita gli esperimenti del *Large Hadron Collider (LHC)* di Ginevra, fornendo servizi alle attività di *storage*, distribuzione e analisi dei dati [WLCG, 2017]. Il CNAF ospita un centro di calcolo *Tier-1* e uno dei principali PoP del GARR. La quantità di dati archiviati e processati è circa 30 PB (PetaByte), ed è previsto un aumento di un ordine di grandezza nei prossimi tre anni. Dal 2009 contribuisce a *European Grid Initiative (EGI)*, un'organizzazione per la sostenibilità della *Grid* in Europa con l'obiettivo di fornire un'*e-Infrastructure* di calcolo e *storage* per gli esperimenti, sempre più performante, basata su tecnologie di calcolo distribuito e *Cloud*. In particolare, attraverso il progetto *European Middleware Infrastructure* garantisce la "manutenibilità", il supporto e lo sviluppo del *middleware* di EGI. Il CNAF dispone di una linea dedicata a 40 Gbps con il CERN e una linea a 10 Gbps con gli altri Tier di WLCG.

2.2 Analisi delle principali e-Infrastructure

Nell'ottica di rafforzare la posizione europea nell'innovazione *data-driven*, migliorare la sua coesione e contribuire a creare un unico mercato digitale [Comunicazione CE COM(2015) 192 final], la Commissione Europea ha adottato un'iniziativa strategica per realizzare un'*European Data Infrastructure* attraverso l'integrazione, il consolidamento e la federazione delle *e-Infrastructure* già esistenti. Le *e-Infrastructure* consentono ai ricercatori un accesso sicuro a strutture e risorse per riprodurre ricerche e per ri-utilizzare i risultati dell'innovazione attraverso reti di comunicazione, calcolo ad alte prestazioni e ad alto throughput, gestione dei dati multidisciplinare e collaborazione su software scientifico in uno scenario tutto europeo. In questo scenario, la partecipazione nazionale alle *e-Infrastructure* è dovuta non solo a Università e a Enti Pubblici di Ricerca, ma soprattutto a centri di eccellenza come:

- GARR che contribuisce a GÉANT,
- CINECA a PRACE e EUDAT,
- e CNAF a EGI.

2.2.1 GÉANT - la rete della ricerca europea

GÉANT è l'infrastruttura paneuropea per la rete telematica della ricerca e dell'istruzione che interconnette le reti nazionali di ricerca (NREN) di tutta Europa, consentendo la collaborazione su progetti che vanno dalla scienza biologica ai fenomeni geologici, dalla tecnologia alle arti [GÉANT, 2017]. Il progetto GÉANT combina un'elevata ampiezza di banda, ad alta capacità, a 50.000 Km di cavi e una gamma crescente di servizi. GÉANT è strategica perché supporta tutte le altre *e-Infrastructure* con una rete telematica *terabit-ready* e servizi online per l'accesso *trusted* sul territorio (eduGAIN). In questo senso, il ruolo di GÉANT è unico in Europa, in quanto, connette 50 milioni di utenti in oltre 10000 istituzioni. Attraverso GÉANT è possibile collegarsi alle reti della ricerca di altri continenti come Internet2 ed ESnet in the USA, AfricaConnect in Africa, TEIN in Asia e RedCLARA in America Latina [Wikipedia inglese, 2017e]. Fondato dalla collaborazione tra la Commissione Europea e le reti telematiche della Ricerca, il progetto GÉANT è supportato dalla collaborazione di 41 partners: 38 NREN, TERENA [TERENA - *Trans-European Research and Education Networking Association*], di DANTE (*Delivery of Advanced Network Technology to Europe*) e di RIPE [RIPE - *Réseaux IP Européens*]. Il progetto GÉANT è fondamentale per *European Research Area* e *Digital Agenda Europe* in *Horizon 2020*. Per connettere tutti gli utenti di GÉANT è stato realizzato eduGAIN [Wikipedia inglese, 2017b] una interfederazione di infrastrutture di Autenticazione e Autorizzazione. In questo modo, un utente può accedere, tramite le sue credenziali di federazione, a servizi dislocati in altri paesi [eduGAIN, 2017].

2.2.2 PRACE ed EuroHPC per il supercalcolo europeo

PRACE (*Partnership for Advanced Computing in Europe*) è un'associazione internazionale senza fini di lucro fondata nel 2010 per creare un'infrastruttura paneuropea di supercomputer per la ricerca [PRACE, 2017]. L'infrastruttura è composta da centri di supercalcolo come **BSC** in Spagna, **CINECA** in Italia, **CSCS** in Svizzera, **GCS** in Germania e **Genci** in Francia e integrata da centri HPC nazionali di tutta Europa [Wikipedia inglese, 2017f]. Aperta alla ricerca sia accademica che industriale in un contesto di *Open Innovation*, PRACE è un catalizzatore fondamentale per affrontare le nuove sfide della società rafforzando la competitività europea. Per stare al passo con lo sviluppo tecnico delle diverse comunità scientifiche, i sistemi di PRACE sono aggiornati continuamente per rendere le tecnologie HPC più avanzate e accessibili. PRACE promuove scuole stagionali e *workshop* in tutta Europa per un uso efficace della propria infrastruttura. La distribuzione dell'accesso tempo di calcolo è supervisionato da un processo di *peer-review* realizzato tramite un comitato scientifico composto da importanti ricercatori europei. PRACE riceve *feedback* attraverso il *Forum PRACE User*, dove gli utenti discutono e condividono la loro esperienza. PRACE è stato finanziato nella sua fase implementativa dal Settimo Programma Quadro dell'UE (2007-2013) e nella sua quarta implementazione, PRACE-4IP, da *Horizon 2020*. Gli obiettivi di PRACE-4IP sono:

- garantire la sostenibilità a lungo termine delle infrastrutture,
- promuovere la leadership europea nelle applicazioni HPC,
- aumentare le risorse umane europee qualificate in HPC e nelle sue applicazioni,
- sostenere un ecosistema equilibrato delle risorse HPC per i ricercatori in Europa,
- valutare le nuove tecnologie e sostenere il percorso europeo per l'utilizzo di risorse dell'exaflop/s.

A marzo 2017, diversi governi europei hanno sottoscritto una dichiarazione d'intenti² che ha portato alla costituzione di EuroHPC, un framework di cooperazione per sviluppare un supercomputer hexascale europeo per la comunità scientifica europea basandosi sull'esperienza di PRACE. La dichiarazione proposta da sette stati europei (tra cui l'Italia) durante il "Digital Day", è stata successivamente sottoscritta da molti altri paesi europei. A gennaio 2018, la Commissione Europea, ha fatto una proposta [Comunicazione CE COM(2018) 8 final] che mira ad istituire la "impresa comune europea per il calcolo ad alte prestazioni" finanziando l'iniziativa promossa dagli stati membri con oltre un miliardo di euro. Successivamente, a marzo 2018, la Commissione Europea ha lanciato la creazione del Consorzio *European Processor Initiative* (EPI) per identificare le migliori soluzioni tecnologiche per i processori da implementare nella realizzazione dell'EuroHPC.

2.2.3 EGI - la Grid europea

La *European Grid Infrastructure* [EGI, 2017] è un'infrastruttura distribuita e multidisciplinare che integra più di 300 *Service Provider* e almeno 20000 utenti raggruppati in 200 *Virtual Organization*. EGI fornisce ai ricercatori europei l'accesso al calcolo, allo storage e a servizi *Cloud*. Realizza e fornisce soluzioni *open* per la scienza e le infrastrutture di ricerca federando risorse digitali e competenze tra differenti comunità oltre i confini nazionali. EGI offre l'accesso a 730000 *core* per lo *High Throughput Computing* (HTC), 6600 per il *Cloud Computing*, 285 PB di *storage online* e 280 PB di *storage library*. EGI è coordinato e gestito dalla fondazione EGI.eu istituita nel 2010 [Wikipedia inglese, 2017c]. Tra i suoi partecipanti ci sono le *National Grid Initiatives* (NGI), l'*European Intergovernmental Research Organisation* (EIRO) e l'*European Research Infrastructure Consortiums* (ERICs). Attualmente, il principale progetto è EGI-Engage [EGI-Engage, 2016], che promuove l'adozione delle *Open Science Commons*, per permettere al ricercatore di usare un servizio digitale *open* per accedere ai dati e alla conoscenza di cui ha bisogno. L'*Open Science Commons* è basata su tre pilastri:

- *e-Infrastructure Commons* - l'ecosistema di servizi chiave,
- *Open Data Commons* - a cui l'utente può accedere per (ri)usare i dati e
- *Knowledge Commons* - in cui le comunità hanno condiviso la paternità della loro conoscenza e partecipano al co-sviluppo di *software* e servizi digitali.

2.2.4 EUDAT - la gestione delle banche dati europee

EUDAT è l'iniziativa europea nata nel 2011 per realizzare una *Collaborative Data Infrastructure* (CDI) come soluzione paneuropea alla sfida della proliferazione dei dati scientifici [EUDAT, 2017]. La CDI consente ai ricercatori di condividere i dati tra le comunità scientifiche e promuovere ricerche interdisciplinari. EUDAT offre la gestione di un servizio dati geograficamente distribuito attraverso una rete di 35 organizzazioni europee, supportando sia i singoli ricercatori sia le comunità scientifiche. I servizi condivisi e le risorse di *storage* sono distribuiti su 15 nazioni europee, mentre i dati sono memorizzati in alcune dei centri di supercalcolo europeo. I servizi offerti da EUDAT si occupano dell'intero ciclo di vita dei dati, assicurando sia l'accesso che il deposito, sia la condivisione che l'archiviazione a lungo termine, supportando l'identificazione, *discoverability* e *computability*. In molte comunità di ricerca, c'è la consapevolezza che la crescente quantità di dati necessita di nuovi approcci per gestirli, per conservarli e per condividerli. La gestione degli archivi

² Declaration on the "Cooperation framework on High Performance Computing", March 2017.
<http://eurohpc.eu/declaration>

dati e dei *Big Data* è una questione che permea tutte le infrastrutture di ricerca. Sin dal suo inizio, il progetto EUDAT ha lavorato con varie comunità scientifiche europee per identificare i servizi di cui i ricercatori hanno bisogno per la gestione dei dati delle ricerche. EUDAT ha creato l'infrastruttura per la gestione dei dati, supportando comunità di ricerca a cui sarebbe risultato difficile realizzarla autonomamente, contribuendo allo scenario digitale europeo (i.e. the *European Data Initiative*, the *European Open Science Cloud*, *Research Data Management* e *Open Access*). EUDAT, inizialmente nato da due progetti finanziati dalla Commissione Europea, prima su fondi FP7 e poi H2020, nel febbraio 2018 è diventata una società privata (EUDAT Ltd.) non a scopo di lucro e si candida come la e-infrastruttura di riferimento a livello europeo per la gestione dei dati scientifici e supportata in questo ruolo dai nodi della rete della *Collaborative Data Infrastructure*.

3. Analisi dei casi studio

Durante la realizzazione di questo studio, sono stati presi in considerazione quattro differenti casi studio per capire come il paradigma dell'*Open Science* è stato recepito e declinato a livello nazionale. I quattro casi presi in esame appartengono a piani differenti, rispettivamente:

1. un caso legato alla realizzazione di politiche (*governance*) e iniziative tese a migliorare l'accesso alle informazioni scientifiche e circolazione della conoscenza a livello europeo - ***Standing Working Group (SWG) on Open Science and Innovation***;
2. un caso legato all'applicazione dell'*Open Science* in un ambito disciplinare più specifico, in particolare sulle Scienze della Terra, ovvero l'***European Plate Observing System (EPOS)***, un consorzio di ricerca europeo di lungo termine finalizzato alla progressiva integrazione delle infrastrutture di ricerca nazionali nell'ambito della Scienze della Terra;
3. un caso simile al precedente, ma in cui il consorzio europeo in questione è legato alle attività di ricerca in ambito marino, ovvero l'***European Multidisciplinary Seafloor and water-column Observatory (EMSO)***;
4. un ultimo caso legato alla realizzazione di una piattaforma per l'innovazione, lo sviluppo e la competitività regionale basata sulla condivisione e integrazione di infrastrutture e dati, definita ***Emilia-Romagna Big Data Community***.

3.1 Standing Working Group on Open Science & Open Innovation

Lo *Standing Working Group on Open Science & Open Innovation* dell'*European Research Area (ERA)* e del comitato per l'innovazione *European Research Area and Innovation Committee (ERAC)* è un gruppo di supporto istituito dalla Commissione Europea per lo sviluppo e l'attuazione di politiche e iniziative legate al contesto dell'*Open Access* e dell'*Open Innovation*, per migliorare l'accesso alle informazioni scientifiche e all'uso delle conoscenze per la ricerca e l'innovazione. Il gruppo di lavoro si concentra sulla priorità 5 (*Optimal Circulation and Transfer of Knowledge*) dell'*ERA Roadmap 2015-2020* che prevede l'attuazione delle politiche di *Open Access* e di trasferimento di conoscenze a livello nazionale per massimizzare la diffusione, l'acquisizione e lo sfruttamento dei risultati scientifici.

L'obiettivo della priorità 5 è, inoltre, quello di facilitare l'individuazione e l'accesso ai risultati dei progetti di ricerca finanziati da fondi pubblici. L'azione specifica che si prevede di implementare è legata alla realizzazione di una piattaforma per l'offerta di servizi informativi per la ricerca che, oltre a facilitare la circolazione della produzione scientifica all'interno del sistema pubblico nazionale, favorisca l'accesso ai risultati della ricerca pubblica da parte delle imprese. Il compito del gruppo è quello di:

- condividere le "*best practices*" tra gli Stati membri e nei paesi associati su come assicurare

e promuovere una circolazione ottimale della conoscenza scientifica e suggerire iniziative pertinenti a livello UE;

- superare ostacoli associati alla libera circolazione della conoscenza per elaborare raccomandazioni, fornire consulenza all'ERAC (entro la metà del 2017) e su qualsiasi altro argomento relativo alla scienza digitale e aperta;
- promuovere la priorità 5 dell'ERA *Roadmap 2015-2020* attraverso l'attuazione delle politiche di accesso e di trasferimento di conoscenze a livello nazionale al fine di massimizzare il programma di diffusione, acquisizione e valorizzazione dei risultati scientifici;
- promuovere l'attuazione delle azioni concordate nell'ERA [Comunicazione COM(2012) 392 final];
- fornire una consulenza politica all'ERAC sulla circolazione delle conoscenze scientifiche in Europa;
- nell'espletamento delle sue attività il Gruppo è vincolato a cinque priorità tematiche:
 - *Open Data e-Infrastructure*;
 - *Open Access* alle pubblicazioni: modelli, costi e metriche;
 - *Open Innovation*;
 - valutazione della ricerca, incentivi e valutazione impatto dei risultati scientifici;
 - formazione e competenze nel contesto di *Open Science* e *Open Innovation*.

Il gruppo, al termine del suo mandato, dovrà redigere e presentare una relazione annuale all'ERAC, fornendo tempestivamente una panoramica strategica e operativa delle questioni relative alla ricerca e all'innovazione coerentemente con quanto stabilito dalla priorità 5 della ERA *Roadmap* per la quale è responsabile. A sua volta, tale documento, servirà all'ERAC per preparare la propria relazione annuale da sottoporre all'approvazione del Consiglio.

3.2 European Plate Observing System (EPOS)

European Plate Observing System (EPOS - <https://www.epos-ip.org>) è un piano di lungo termine finalizzato alla progressiva integrazione delle infrastrutture di ricerca europee nell'ambito della Scienze della Terra e realizzato nella forma di un *European Research Infrastructure Consortium* (ERIC), un consorzio approvato dall'*European Strategy Forum on Research Infrastructures* (ESFRI) e supportato da 47 membri provenienti da 25 paesi europei.

EPOS è finalizzato a dotare la comunità scientifica di riferimento di una serie di servizi infrastrutturali sostenibili e condivisi realizzati con moderne soluzioni tecnologiche, e appoggiandosi dove possibile, ad infrastrutture europee già esistenti o in costruzione. Gli ambiti disciplinari coperti finora sono: sismologia, dati satellitari, osservatori vulcanologici e geomagnetici, pericolosità legata ad attività antropogeniche, dati geologici, laboratori multi-scala ed esperimenti legati alle geo-risorse a bassa emissione di anidride carbonica.

La progettazione di EPOS è iniziata nel 2002 ed è stata ammessa nel piano d'azione ESFRI nel 2008. La realizzazione dell'ERIC, un'entità legale a tutti gli effetti e di lungo mandato con una sede e un'amministrazione propria, è stata portata avanti in diverse fasi; si è partiti con la "*Conception Phase*" (2002-2008) basata su una serie di progetti nazionali ed europei, una "*Preparatory Phase*" (2010-2014) basata su un progetto FP7, e una "*Implementation Phase*" (2014-2019) basata su un progetto H2020. La struttura legale dell'EPOS ERIC è diventata operativa il 30 settembre del 2018. L'obiettivo finale di EPOS è diventare il "CERN delle Scienze della Terra", puntando alla creazione di nuove *facilities* su grande scala che permetteranno anche il *data mining*, ossia l'analisi e l'elaborazione di grandi quantità di dati per fare correlazioni ed aumentare le potenzialità del loro utilizzo. Una panoramica più dettagliata dell'infrastruttura di EPOS è pubblicata in Bailo et al. [2018].

Un'infrastruttura che collega la dimensione nazionale a quella europea

EPOS mira a valorizzare le infrastrutture di ricerca esistenti favorendo l'adozione di soluzioni standardizzate su diversi fronti, da una politica dei dati (*data policy*) condivisa, a servizi per l'accesso ai dati interoperabili. L'adozione di soluzioni standardizzate favorisce l'accesso trasparente degli utenti a dati interdisciplinari e servizi, indipendentemente dalla propria posizione, creando un'infrastruttura distribuita. Grazie a EPOS si tende a favorire l'accesso, l'uso di dati, servizi e laboratori che, oltre a innovare il processo della ricerca, mirano a rendere la ricerca scientifica più efficiente riducendo la ridondanza dei servizi.

La struttura organizzativa di EPOS (Fig. 3) parte da un piano nazionale in cui sono presenti le **National Research Infrastructure (NRI)**. A livello europeo i servizi di accesso ai dati di ciascuna NRI si consorziano su base tematica in **Thematic Core Services (TCS)**, ciascuno governato da un comitato di coordinamento. In Italia, al fine di raccogliere in un unico soggetto legale tutti gli enti coinvolti (INGV, CNR, ISPRA, INOGS, Università di Trieste, di Genova, Federico II, Roma Tre, AMRA, CINECA, EUCENTRE) è stata creata una **Joint Research Unit (JRU)** denominata "EPOS-Italia". Il piano successivo vede gli **Integrated Core Services (ICS-C)**, ovvero i servizi che permettono l'accesso interdisciplinare ai dati e sono di due tipi, a seconda del modo in cui accedono ai dati di base: ICS-C, dove "C" sta per *Central Hub*, e ICS-d, dove "d" sta per *distributed*. I servizi ICS sono il vero valore aggiunto e l'aspetto più innovativo dell'intera infrastruttura, in quanto i servizi TCS più avanzati sono il frutto di precedenti progetti FP6 e FP7 e altri servizi TCS sono attualmente in corso di sviluppo tramite alcuni progetti H2020 ancora in corso. I servizi ICS-C includono, oltre a infrastrutture tecnologiche e sistemi di metadato sperimentali, anche i piani per l'integrazione legale, amministrativa e finanziaria.

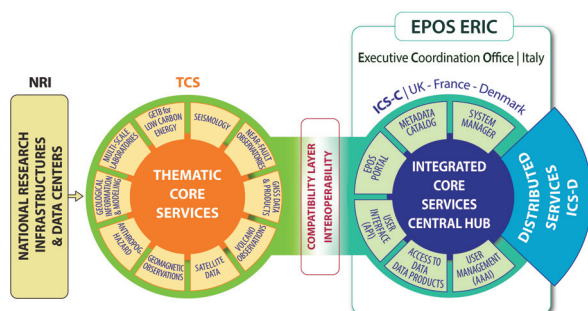


Figura 3 Schema semplificato dell'infrastruttura che compone l'European Research Infrastructure Consortium "EPOS".

Figure 3 Simplified scheme of the infrastructure of the European Research Infrastructure Consortium named "EPOS".

L'infrastruttura che gestisce il flusso dei dati e i meccanismi che regolano la comunicazione tra *National Research Infrastructure (NRI)*, *Thematic Core Service (TCS)*, e *Integrated Core Service (ICS)* è entrata in una prima fase di test operativo durante il 2017, e a partire da ottobre 2017 è oggetto di un complesso sistema di validazione da parte di referenti scientifici e tecnologici della Commissione Europea, che seguiranno linee guida comuni a quelle per altre infrastrutture [e-IRG, 2017a;b].

La gestione amministrativa e il coordinamento dell'ERIC EPOS [Atakan et al., 2017], definito come l'*Executive and Coordination Office (ECO)*, è gestito da INGV, mentre ICS-C è in capo al *British Geological Survey (BGS)*, al *Bureau De Recherches Géologiques et Minières (BRGM)*, con il supporto tecnologico da parte del *Geological Survey of Denmark and Greenland (GEUS)*. Il governo dell'ERIC EPOS è in capo alla *General Assembly (GA)* composta dai rappresentanti di tutti i membri. Poiché l'ERIC è ancora in corso di costruzione, le decisioni sono temporaneamente in capo al *Board of Governmental Representatives (BGR)*.

Intervista al coordinatore dell'ERIC EPOS e del progetto H2020 EPOS-IP

A gennaio 2017 abbiamo intervistato Massimo Cocco presso il MIUR, il responsabile dell'attività di costruzione dell'ERIC in coordinamento con il MIUR, e del progetto H2020 EPOS *Implementation Phase*. Gli argomenti affrontati hanno spaziato su diverse tematiche, tra cui

problemi di *governance* e modalità di finanziamento delle infrastrutture di ricerca, che comprendono tra le altre cose i costi per sostenere l'*Open Access* e piani di conservazione dei dati di lungo periodo con un orizzonte temporale di circa 50 anni.

Alcune delle risposte le forniranno altri progetti europei, come ad esempio EUDAT (vedi par. 2.2.4), ma su alcune problematiche non è ancora chiaro chi potrà fornire le risposte. Un esempio fra tutti è il costo della sostenibilità dell'*Open Access* relativo ai dati, per cui, contrariamente all'ambito delle pubblicazioni scientifiche, non è ancora chiaro chi dovrà sostenerlo. Se per quanto riguarda le infrastrutture di ricerca *single-site* il soggetto coinvolto è uno, e quindi con un livello di complessità ridotto, le infrastrutture distribuite che gestiscono i dati su più nodi spesso gestiti da soggetti diversi, portano con sé un livello di complessità elevato che travalica quasi sempre i confini nazionali.

Contrariamente ad altre infrastrutture che gestiscono dati, EPOS intende mantenere direttamente il controllo sulla gestione dei dati interfacciandosi direttamente con chi quei dati li produce, ovvero le infrastrutture di ricerca, e quindi in maniera del tutto indipendentemente da dove i dati sono fisicamente archiviati. EPOS si avvantaggerà della collaborazione con soggetti nazionali che hanno esperienza nella gestione di *Big Data* e conoscenze tecnologiche adatte ad affrontare la sfida, in Italia si parla del CINECA (vedi par. 2.1.2), INFN-CNAF (vedi par. 2.1.3) e dell'associazione Big Data. La buona riuscita di questo approccio è fortemente legata all'interoperabilità, che non solo coinvolge i dati, bensì, cosa anche più importante, i metadati. È, infatti, fondamentale che le procedure di gestione che abilitano servizi centralizzati, come ad esempio la ricerca dei dati, possano avere accesso ad un set di metadati standardizzato e indipendente dall'ambito scientifico specifico di riferimento. Anche sul piano della *governance* dovrà essere affrontato il tema dell'interoperabilità, in quanto dovrà essere garantita la proprietà intellettuale e le eventuali limitazioni d'uso delle diverse licenze originali legate ai dati provenienti da diverse infrastrutture di ricerca, e si dovrà regolamentare la produzione di nuovi dati generati attraverso i servizi a livello di *e-Infrastructure*.

Altro tema molto importante è quello legato alla terminologia. In ambito scientifico, alcuni termini variano di significato al variare dell'ambito, causando incomprensioni e fraintendimenti. È fondamentale quindi che in un progetto complesso come EPOS si stabilisca una comune tassonomia per promuoverne il relativo uso nelle diverse comunità scientifiche coinvolte.

Si è affrontata l'interazione tra ERIC diversi sul piano sistema di coordinamento e collaborazione, anche al fine di ottimizzare le risorse disponibili. Fino al 2010 non esisteva nessun meccanismo di coordinamento, poi nacquero i "*cluster*", un tentativo nato da una prima discussione tra il *Directorate-General for Research and Innovation* (DG-Research) e il *Communications Networks, Content and Technology* (DG-Connect) della Commissione Europea che fino ad allora finanziavano le rispettive infrastrutture di ricerca e di calcolo in modo totalmente separato. I cluster funzionarono parzialmente per la parte riguardante la ricerca scientifica, molto meno efficacemente per la parte di calcolo. A complicare la situazione esistono in Europa altre due entità che hanno margini di sovrapposizione e che non hanno un sistema di coordinamento con gli altri DG: il *Directorate-General Joint Research Centre* e l'*European Space Agency*. La nuova risposta al problema del coordinamento nasce con il progetto pilota H2020 *European Open Science Cloud* (EOSC) lanciato a gennaio 2017. EOSC ha però un problema alla base che potrebbe inficiare lo spirito di coordinamento tra i due DG, poiché da un lato sarà finanziato direttamente da DG-Research che si occuperà di *governance*, sostenibilità e architettura, e dall'altro DG-Connect che finanzierà i singoli progetti che porteranno alla costruzione delle infrastrutture tecnologiche, dando per scontato che nasca spontaneamente il coordinamento tra le due azioni. Un altro punto importante che deve essere ancora risolto è la gestione delle licenze associate ai dati, un tema è già stato affrontato da tempo in ambiti come quello del *software open source* e delle pubblicazioni scientifiche. Una delle questioni più spinose per un ERIC come EPOS è la gestione di enormi quantità di dati provenienti da fonti diverse a cui ogni istituzione ha associato

licenze diverse che spesso rendono l'uso dei dati tra loro incompatibili, soprattutto nell'ottica della generazione di dati derivati dalla combinazione di dati diversi. Questo aspetto, combinato con aspetti riguardanti la conservazione a lungo termine degli archivi di dati, viene affrontato con il *Data Management Plan*, uno strumento che permette di chiarire moltissimi aspetti, a volte affrontati solo in maniera superficiale e poco strutturata.

La gestione delle nuove infrastrutture che saranno i motori dell'*Open Science* prevedono conoscenze e capacità operative poco diffuse, per cui è fondamentale l'attivazione di nuovi percorsi formativi che creino *data managers*, che coprano sia la parte di *governance*, legale e tecnologica.

EPOS ha tentato nel 2013 un ingaggio con l'industria petrolifera e dell'aviazione privata in Norvegia. Dalle discussioni è emerso che non c'era volontà di collaborazione, ma piuttosto la richiesta di fornire prodotti e servizi finiti utilizzabili per esempio durante le emergenze. Un altro esempio di tentativo di ingaggio è stato fatto con le compagnie minerarie in ambito "*anthropogenic hazard*", in cui è stata espressa l'esigenza di ottenere da EPOS degli strumenti di monitoraggio delle attività, apponendo una serie di vincoli senza l'intenzione di collaborare scientificamente alle attività.

Gli ultimi argomenti affrontati riguardano il coinvolgimento dei ricercatori sia di Enti di Ricerca, sia di Università, che spesso non comprendono le potenzialità dei futuri servizi offerti da EPOS o altri ERIC, anche perché non vedono una oggettiva valorizzazione delle attività legate alla produzione, gestione e sfruttamento dei dati. Il problema qui è molto ampio, e spazia da nuovi sistemi di valutazione dell'ANVUR, a sistemi tecnologici per il tracciamento efficace dell'uso dei dati, fino alla valutazione dell'impatto sulla società e alla regolamentazione dello sfruttamento con ritorno economico da parte dei privati.

3.3 European Multidisciplinary Seafloor and water-column Observatory (EMSO)

A febbraio 2018 è stato possibile intervistare Laura Beranzoli, consigliere tecnico-scientifico del delegato italiano in EMSO ERIC (www.emso-eu.org).

EMSO ERIC è stato fondato nell'ottobre 2016 ed è quindi un soggetto legale autonomo riconosciuto dalla Commissione Europea e dagli stati membri dell'Unione Europea. Il consorzio coordina la realizzazione, lo sviluppo e il mantenimento di osservatori multidisciplinari che acquisiscono dati geofisici, oceanografici e ambientali nei mari europei. EMSO ERIC, essendo una istituzione internazionale, può sottoscrivere accordi di collaborazione con altre istituzioni e può aderire alla partnership di progetti. Lo statuto dell'ERIC, la cui pubblicazione sulla Gazzetta Europea³ ha sancito la sua costituzione, è stato sottoscritto da otto stati europei: Italia, Portogallo, Spagna, Francia, Grecia, Romania, Gran Bretagna, e Irlanda. Ciascuno governo firmatario dello Statuto, per mezzo del proprio Ministero della Ricerca, individua tra i propri istituti di ricerca una o più Representing Entities, ovvero uno o più istituti a cui viene demandata l'attuazione dei piani delle attività dell'ERIC. Le Representing Entities sono riconosciute tali perché proprietarie di parti dell'infrastruttura EMSO che produce i dati (serie temporali) e fornisce servizi di accesso ai dati e agli osservatori.

EMSO ERIC al momento coordina una infrastruttura distribuita costituita da otto osservatori multiparametrici ovvero piattaforme equipaggiate di strumenti geofisici e oceanografici. Accanto ai nodi per le osservazioni multiparametriche, esistono tre siti in acque basse con piattaforme cablate per scopi sperimentali, utili per lo sviluppo e calibrazione dei prototipi di strumenti che andranno posizionati successivamente in acque profonde e dei relativi software di acquisizione e gestione dei dati.

³ [http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016Y1001\(01\)](http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016Y1001(01))

Grazie all'azione di coordinamento di EMSO ERIC sulle *Representing Entities* è stato possibile presentare un progetto europeo H2020 denominato *European multidisciplinary Seafloor and water-column observatory development*⁴ (EMSOdev) approvato successivamente dalla Commissione Europea, che nel periodo 2015-2018 deve realizzare due principali obiettivi:

1. raggiungere la standardizzazione di un sottoinsieme delle misure acquisite dai nodi dell'infrastruttura per mezzo della progettazione di un modulo prototipo per il monitoraggio marino denominato *EMSO Generic Instrument Module* (EGIM), in grado di acquisire in modo continuo la misura di 7 parametri fisici (temperatura, pressione, conducibilità, ossigeno disciolto, torbidità, rumore ambientale) con sensori selezionati in funzione di precise caratteristiche tecniche;
2. omogeneizzare all'interno dell'infrastruttura gli aspetti relativi alla codifica, trasferimento e condivisione dei dati, attraverso l'adozione di una *Data Management Platform*, che faccia un esteso uso di *web services* e renda accessibili i dati dei nodi di EMSO.

Nell'ambito di EMSOdev si compie anche uno sforzo mirato a gettare le basi per la definizione di una politica per la gestione dei dati (*Data Policy*) anche in modo coerente e coordinato con le altre infrastrutture di ricerca ambientali. Al momento la documentazione pubblicata sulle questioni relative alla gestione dei dati in EMSO ERIC [Pappalardo et al., 2015] riporta un'analisi tecnologica degli standard di codifica e trasferimento dei dati, opera uno sforzo di armonizzazione della terminologia in uso e traccia i principi politici generali per la condivisione e diffusione dei dati, recependo i principi dell'*Open Access* suggeriti in ambito dei progetti europei.

L'implementazione delle componenti principali del data management di EMSO ERIC è tuttavia uno degli obiettivi del programma delle attività del progetto H2020 EMSO-link⁵ (2017-2020) coordinato da EMSO ERIC. EMSO-link mira al pieno sviluppo e operatività dell'organizzazione dell'ERIC e all'elaborazione di un piano di sostenibilità di lungo periodo di EMSO, anche nell'ottica della realizzazione dell'*European Ocean Observing System* (EOOS). Questo progetto dà supporto alla progettazione e all'avvio dei servizi che EMSO fornirà agli *stakeholders* con particolare attenzione a:

1. *Data Services*, per la visualizzazione, utilizzo, controllo di qualità e validazione dei dati attraverso EMSO Data Portal;
2. *Logistics and Engineering Services*, per il coordinamento interno delle scelte tecniche e tecnologiche e il supporto agli utenti che richiedano un accesso fisico agli osservatori;
3. *Communication Services*, per la diffusione e la divulgazione delle attività e delle opportunità offerte da EMSO agli utenti scientifici e agli *stakeholders* e quindi anche delle 'regole' che gli utenti devono osservare nell'uso dei dati e dei prodotti;
4. *Science Services*, per la realizzazione di prodotti scientifici, derivanti dall'analisi e dalle elaborazioni dei dati di EMSO, che incontrino le esigenze dell'utenza scientifica, formativa, istituzionale, industriale.

Per la natura distribuita dell'infrastruttura e per l'attuale organizzazione (gli osservatori sono di proprietà delle *Representing Entities* che spesso non hanno adottato *Data Policies*), le questioni legate alla gestione dei dati di EMSO riguardano l'intera catena del *workflow*. Per fare alcuni esempi certamente non esaustivi tali questioni comprendono la definizione della proprietà dei dati che fluiscono dagli osservatori verso la *Data Management Platform* e che sottende i ruoli dell'ERIC e delle *Representing Entities*; l'adozione di un adeguato sistema di identificazione e tracciabilità dei dati stessi per un corretto ri-uso nel lungo termine fino alla definizione delle

⁴ https://cordis.europa.eu/project/rcn/197997_en.html

⁵ Implementation of the Strategy to Ensure the EMSO ERIC's Long-term Sustainability
https://cordis.europa.eu/project/rcn/207911_en.html

procedure per la *Data Curation* (ad es. necessità di archiviazione di grandi capacità, ridondanza degli archivi, mantenimento dell'infrastruttura dati allo stato dell'arte); la definizione di una adeguata *Data Policy* verso le industrie quando esse accedono all'infrastruttura per testare e validare nuovi prodotti; la definizione di una *Data Policy* che tuteli l'ERIC quando questa concordi con altri soggetti la fornitura di dati e di prodotti che alimentano servizi gestiti da terzi.

In questo contesto, EMSO ERIC diviene quindi un "laboratorio" cui guardare per ulteriori casi studio.

La dimensione italiana di EMSO

La *Representing Entity* di EMSO ERIC per l'Italia è INGV, scelta che riconosce sia il ruolo promotore nella creazione dell'ERIC, sia il ruolo guida sul piano tecnico-scientifico. Come avviene per EPOS, anche per EMSO è stata costituita ancor prima della costituzione dell'ERIC, una **Joint Research Unit (JRU)** nazionale composta da otto istituti (INGV, CNR, INFN, SZN, OGS, ISPRA, ENEA, CoNISMa) che, tramite un accordo formale di collaborazione, mira a costituire un organismo super-partes che svolga un ruolo di coordinamento delle attività connesse con quelle di EMSO e condotte da ciascun ente. Tra gli altri scopi dichiarati dalla JRU c'è il supporto ad attività di formazione, al trasferimento tecnologico, alla progressiva integrazione di altri osservatori marini regionali e a possibili accordi con altre JRU. Nell'accordo si auspica inoltre che l'attività sinergica porti ad accrescere le possibilità di partecipare a bandi per progetti nazionali ed europei, che sarebbe più difficile per ogni istituto preso singolarmente.

La JRU ha il compito di partecipare alle attività di EMSO ERIC con le competenze e le peculiarità dei membri della JRU stessa (enti di ricerca) e anche con le proprie *facilities*. Questa modalità di coordinamento per una efficace partecipazione nazionale ad una infrastruttura europea è adottata anche da altri Paesi Membri. È Evidente che la JRU aggiunge necessariamente un ulteriore grado di complessità, quello di livello nazionale, alle questioni relative alla condivisione e gestione dei dati di una infrastruttura a scala europea. Possiamo menzionare a titolo di esempio la necessità di allineare e armonizzare le procedure di documentazione dei dati (metadati) ambientali in ambito nazionale, oppure la necessità di conciliare *data policies* differenti anche all'interno dello stesso ente quando questo partecipa a più infrastrutture europee come nel caso di CNR, OGS, SZN, INGV, INFN.

3.4 Emilia-Romagna Big Data Community

Per attuare "un'unica strategia digitale per il mercato digitale europeo", abbiamo bisogno di "azioni con ripercussioni a lungo termine sulla competitività industriale europea, sugli investimenti per Cloud Computing e Big Data, sulla ricerca e l'innovazione nonché sulle competenze". In questo contesto di *Open Science*, l'Italia deve offrire:

- alle Piccole e Medie Imprese (PMI) - un facile accesso alle infrastrutture ICT per aumentare la loro competitività, stimolare la creazione di posti di lavoro e la crescita economica,
- alle Pubbliche Amministrazioni (PA) - un'infrastruttura ICT per i servizi al cittadino, nel settore dell'istruzione, sanità, trasporti e favorire l'e-Government.

In Emilia-Romagna, sono ubicati, storicamente, cluster di supercalcolo, grandi archivi di dati e reti ad alta velocità grazie agli sforzi a livello europeo di INFN, INGV, CNR, GARR e CINECA. Queste infrastrutture ICT sono alla base di *Cloud Computing* e analisi di *Big Data* e potrebbero promuovere la crescita economica regionale permettendo alle PMI e alle PA non in grado di realizzare tali infrastrutture da sole, di usufruire di tali servizi. In questo contesto, nasce la piattaforma *Emilia-Romagna Big Data Community* che ritiene che la crescita economica e il benessere saranno promosse dal *Cloud Computing* e dal *Big Data* in uno scenario *Open Science* e che saranno PMI e PA a caldeggiare lo studio e lo sviluppo delle future infrastrutture

tecnologiche e degli scenari scientifici. Come ultimo atto, il 6 giugno 2018 a Bologna, è stata costituita l'associazione *Big Data* con lo scopo di promuovere le attività della piattaforma fondata in Emilia-Romagna di centri ed infrastrutture di ricerca di eccellenza nel campo del supercalcolo e del trattamento dei Big Data.

3.4.1 Il Mandato della Regione Emilia-Romagna per la creazione della piattaforma

L'Emilia-Romagna ha completato la ricognizione su *Cloud Computing* e *Big Data* [ER, 2016a]. Bonaccini, il presidente della Regione, ha manifestato l'interesse nella creazione di una *piattaforma per l'innovazione, lo sviluppo e la competitività regionale* definita **Emilia-Romagna Big Data Community**: *“La Regione si è data come obiettivo di mandato di fare di Bologna e dell'Emilia-Romagna un grande Hub europeo della ricerca. Le università, i centri di ricerca, gli enti pubblici di ricerca, il Cineca, il Rizzoli, che con CNR e Università sono la frontiera più avanzata dell'innovazione, la nostra rete Alta Tecnologia, le molte imprese che operano come fornitori dei laboratori più avanzati, costituiscono un insieme che ha titolo ad essere riconosciuto come grande infrastruttura, e noi ci candidiamo ad essere l'istituzione che mette a sistema tutte queste eccellenze”*. Supportato dall'assessore alla Ricerca e all'Università (Patrizio Bianchi), [Sole24Ore, 2016]: *“Il 70% della capacità nazionale di super calcolo è in Emilia-Romagna. Adesso dobbiamo passare dalla quantità al valore, realizzare una politica industriale utile al riposizionamento del Paese che passi attraverso il sistema della ricerca. In questa regione, sono occupati in questo settore 1.800 ricercatori, 230 ricercatori stranieri, tra il 2013 e il 2015 sono stati realizzati 60 percorsi di alta formazione. Non c'è ambito della ricerca e dell'innovazione per cui non sia fondamentale la capacità di gestire grandi quantità di dati. La nostra intenzione adesso è valorizzare specializzazioni e complementarità maturate da tutti i centri di ricerca che lavorano nell'ambito del supercalcolo e del big data e creare un sistema aggregato più competitivo anche a livello europeo”*. Si inserisce l'assessore all'Attività Produttive (Palma Costi) [ER, 2016b]: *“L'attività ha previsto una ricognizione delle infrastrutture esistenti a livello regionale, che abbiamo realizzato con il supporto di Aster. Da tale ricognizione è stato possibile ottenere una situazione aggiornata sull'esistenza in regione di strutture di ricerca e innovazione che presentano le potenzialità scientifiche, tecnologiche ed organizzative adatte per incrementare le capacità competitive delle imprese, in coerenza con quanto previsto dalla Strategia di specializzazione intelligente regionale. Supercalcolo e big data, materiali avanzati e sistemi di produzione innovativi e genomica, medicina rigenerativa e biobanche sono state identificate come le tematiche di rilevanza strategica per la regione, e per sostenerle abbiamo destinato 7 milioni di euro”*. Da questa dichiarazione si intuisce che l'obiettivo della Regione è attrarre nuovi investimenti per realizzare un'infrastruttura in linea con *Horizon 2020* che vuole creare un'eccellenza scientifica per garantire una produzione di ricerca a livello mondiale che assicuri all'Europa competitività a lungo termine. Per meglio affrontare questi temi la *Big Data Community* è diventata l'associazione *Big Data* con lo scopo di:

- promuovere nei propri soci l'adozione di tecnologie, modelli, processi volti a massimizzare l'integrazione delle infrastrutture di calcolo disponibili;
- promuovere azioni per favorire il potenziamento delle infrastrutture disponibili;
- formulare alla Regione Emilia-Romagna proposte per la realizzazione di programmi ed iniziative volte al rafforzamento della capacità del sistema produttivo di sfruttare le potenzialità offerte dai Big Data;
- creare sinergie ed alleanze attraverso lo sviluppo e il potenziamento di reti e collegamenti coordinati e stabili con altre aggregazioni pubblico-private attive in ambiti analoghi a livello nazionale ed europeo.

Per realizzare i questi scopi l'associazione parteciperà a bandi o inviti a presentare proposte nell'ambito di programmi di finanziamento regionali, nazionali, europei ed internazionali. INGV assieme a ASTER Soc. Cons. p.a, CINECA, CMCC, CNR, ENEA, INAF, INFN, IOR, LEPIDA Spa, Università di Bologna, Università di Ferrara, Università di Modena e Reggio Emilia e Università di Parma è uno dei soci fondatori.

3.4.2 Applicazioni della piattaforma

Le possibili applicazioni della piattaforma spaziano da aree di ricerca come la fisica delle particelle, l'esplorazione spaziale, lo studio dei terremoti, a domini applicativi come l'analisi finanziaria, la salute, il monitoraggio ambientale, il monitoraggio di sismicità indotta, la gestione del patrimonio culturale, l'agricoltura di precisione, la multimedialità ecc. La ricognizione sul periodo 2013-2015 [Regione Emilia-Romagna, 2016] ha evidenziato che nella Regione intorno al tema *Big Data*: lavorino quasi 1800 ricercatori, sono stati ospitati 230 ricercatori stranieri, si sono svolti 94 eventi internazionali e ci sono state 60 iniziative didattiche (Corsi di *PhD*, Laurea Magistrale, Master e *Summer Schools*). Per capire le potenzialità di questa piattaforma è necessario capire quali sono le infrastrutture tecnologiche in gioco. Queste infrastrutture saranno agevolate attraverso la realizzazione di un Tecnopolo che ospiterà anche l'*European Centre for Medium-range Weather Forecasts* (ECMWF). La connettività della Regione è legata sia a GARR e sia a Lepida. GARR (vedi par. 2.1.1) fornisce una larghezza di banda fino a 100Gb/s e 4 *Point of Presence* che permettono l'accesso alla Rete Nazionale della Ricerca. Lepida offre una capacità di banda pari a 100Gb/s, più di 140000 Km di fibra ottica e 2500 nodi di accesso, 42 *Point of Presence*, per connettere tutta la Regione. L'infrastruttura *High Performance Computing* (HPC) ospitata dal CINECA (vedi par. 2.1.2) è attualmente basata su un sistema *Tier-0* per il super calcolo che aderisce a PRACE e EuroHPC (vedi par. 2.2.2) e un *Tier-1* che realizzerà elaborazione e analisi per *Big Data*, con l'introduzione di un servizio di archivio online per la conservazione e preservazione a lungo termine che aderisce a EUDAT (vedi par. 2.2.4). L'infrastruttura *High Throughput Computing* (HTC) ospitata dal CNAF (vedi par. 2.1.3) nasce col compito di mantenimento della *Grid* italiana e dello sviluppo del suo *middleware* e ora sviluppa un'infrastruttura Cloud tramite la partecipazione a EGI (vedi par. 2.2.3) che permette l'accesso a grandi banche dati con elevata larghezza di banda. Per aumentare la competitività della piattaforma, tramite il progetto tecnopolo è stato potenziato il collegamento tra il *Tier-0* HPC del Cineca e il *Tier-1* HTC del CNAF-INFN attraverso un link ultra veloce in fibra dedicato con capacità fino a 1,2 Terabit/s [CINECA, 2018].

4. Le sfide dell'Open Science

4.1 Nuovi approcci per il coinvolgimento, valutazione e incentivazione

L'avvento di smartphone e di social media ha permesso la circolazione sempre più veloce di informazioni più o meno attendibili. Con l'attuazione della cosiddetta "*Amministrazione Trasparente*" (D.Lgs. 14/03/2013, n.33, G.U. n.80 del 05/04/2013), e delle politiche di *Open Access* alla produzione scientifica, tutte le persone possono avere accesso a informazioni dettagliate su molti aspetti amministrativi e scientifici prodotti da Università ed EPR. La Società si trova per la prima volta ad avere accesso ad a informazioni riservate, avvicinando mondi fino a qualche anno fa distantissimi. Tutto questo comporta un livello di coinvolgimento che il mondo della Ricerca sta faticosamente imparando ad affrontare.

Tradizionalmente il mondo della ricerca basa la misurazione del valore della propria produzione scientifica misurando l'impatto dei propri articoli pubblicati su riviste con diversi metodi. Per garantire un buon livello di qualità dei contenuti ha sviluppato sistemi di *peer-review*, cioè sistemi di valutazione esterna degli articoli da parte di esperti del settore. Purtroppo la valutazione dei risultati, specialmente nei sistemi accademici, si è evoluta in un sistema autoreferenziale di incentivazione all'attività vincolante per l'avanzamento di carriera dei ricercatori. **Oggi l'attività scientifica è caratterizzata da una eccessiva competizione che si basa più sulla quantità piuttosto che sulla qualità della produzione scientifica e al suo impatto ("publish or perish")**, generando risultati basati su processi difficilmente riproducibili e dando eccessivo peso a pubblicazioni su riviste scientifiche ad alto impatto, piuttosto che su modi di condivisione moderni più efficaci ed efficienti [Leonelli, 2017]. In questo senso, si ricorda la "San Francisco Declaration On Research Assessment" [DORA; ASCB, 2012], che, attraverso un elenco di raccomandazioni, cerca di limitare l'adozione dell'*Impact Factor* delle riviste come unico metro di misurazione dei risultati.

Nell'attuale discussione su quali siano le opportunità derivanti dall'adozione dell'*Open Science*, e in particolar modo nell'ambito del movimento *Open Access* legato alle pubblicazioni scientifiche, si stanno sperimentando nuovi sistemi di revisione dei risultati chiamati *open peer-review* [Ross-Hellauer, 2017] che provano a sfruttare maggiormente le possibilità offerte dai moderni sistemi di condivisione. Nonostante già da alcuni decenni si siano sviluppati sistemi più o meno efficaci per la misurazione dell'impatto di un articolo pubblicato, è ancora in corso una complessa discussione [EASE, 2007; Dolenc et al., 2016; Leonelli, 2017] che coinvolge il movimento d'opinione legato all'*Open Access* [UNESCO, 2015]. Viceversa, **la discussione scientifica su quali siano le possibili soluzioni per la valutazione dell'impatto scientifico che può avere una banca dati è solo agli albori** [Hodson et al., 2015], e solo recentemente si iniziano a vedere proposte concrete [Open Data Barometer, 2017] grazie alla disponibilità di soluzioni tecnologiche basate sul tracciamento della disponibilità e uso dei dati.

In Italia, la valutazione della produzione scientifica è affidata dal 2006 all'Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca (ANVUR) che periodicamente conduce i bandi di Valutazione della Qualità della Ricerca (VQR). Al momento sono stati condotti due bandi, uno per il periodo 2004-2010 e l'altro per il periodo 2011-2014. La valutazione del primo bando era incentrata su criteri bibliometrici, dando peso ai cosiddetti indici di impatto delle riviste scientifiche che molto dibattito hanno generato [ANVUR, 2011; Galimberti, 2012]. Nel secondo bando è stata ricalibrata la valutazione di prodotti scientifici alternativi alle pubblicazioni, dando più peso a prodotti alternativi come le banche dati.

Il successo dell'*Open Science* sarà direttamente proporzionale al grado di adozione delle nuove pratiche di condivisione dei risultati dei ricercatori, e qui i sistemi da valutazione sia a livello centrale (ANVUR), sia a livello istituzionale, giocheranno un ruolo fondamentale poiché saranno gli unici sistemi di incentivazione tangibili che vedranno applicati i ricercatori. La Commissione Europea suggerisce una serie di indicatori da tenere presente per la gestione dell'evoluzione della carriera dei ricercatori, definendo la cosiddetta "**Open Science Career Assessment Matrix**" [EU Working Group on Open Science Rewards, 2017], che tiene conto dei risultati della ricerca, dei processi adottati per conseguirla, dei servizi a supporto alla ricerca offerti, dell'impatto, della trasmissione della conoscenza e all'esperienza maturata. Nell'approntare nuovi, più efficaci sistemi di valutazione, è importante proteggere i giovani ricercatori [Bratislava Declaration of Young Researchers, 2016], che se da un lato sono spinti a condividere liberamente i risultati, dall'altro sono più facilmente soggetti ad azioni di prevaricazione, sfruttamento e più in generale, hanno minori capacità di far riconoscere il valore dei propri risultati. Nuovi sistemi dovrebbero inoltre tenere conto delle peculiarità delle diverse discipline, ognuna caratterizzata da diversi parametri che possano dimostrare l'impatto dei risultati, diversità particolarmente accentuate tra l'ambito delle scienze umanistiche più legate ad aspetti qualitativi, rispetto ad ambiti delle

scienze naturali, più legate ad aspetti quantitativi.

L'auspicio per il futuro è quindi quello di evitare la via di uscita semplicistica basata sull'adozione di sistemi automatizzati di valutazione dell'impatto, che rischiano di causare seri problemi nelle decisioni relative ai finanziamenti [*"inappropriate indicators create perverse incentives"*, in Wilsdon et al., 2015; Kenna et al., 2017].

Al fine di tentare di misurare l'impatto dell'*Open Science*, la Commissione Europea ha istituito l'*Open Science Monitor* [Elta et al., 2016; RAND Europe, 2017], una piattaforma che si basa sull'analisi di una serie di indicatori (Fig. 4) come il livello di adozione dell'*Open access* relativo alle pubblicazioni e i dati scientifici, alle attività di comunicazione relative all'*Open science*, e indagini su attività di *Citizen science* e sul coinvolgimento della popolazione nelle attività della ricerca scientifica.

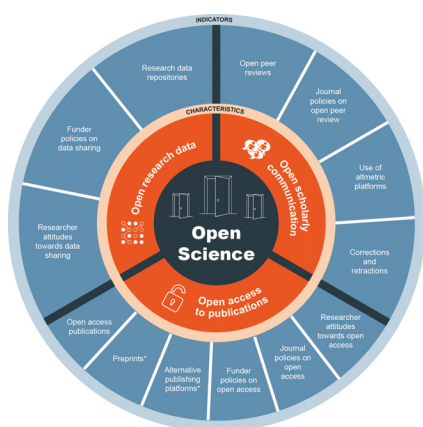


Figura 4 Schema che rappresenta gli indicatori su cui si basa l'Open Science Monitor [Elta et al., 2016; RAND Europe, 2017].

Figure 4 Indicators selected for the "Open Science Monitor" [Elta et al., 2016; RAND Europe, 2017].

A maggio 2018 è stato pubblicato un primo studio della Commissione Europea [EC DG-Research, 2018a] sui risultati e l'impatto delle iniziative a supporto dell'*Open Science* nei diversi stati membri. Da questa prima indagine pare che l'impatto stia generando ricadute positive e un certo grado di miglioramento è rilevato soprattutto nell'ambito della politica di accesso alle pubblicazioni scientifiche, anche se si rileva come esista un enorme divario da stato a stato, e pure da istituzione a istituzione. Nell'ambito di sistemi innovativi per la misurazione dell'impatto della Ricerca Scientifica, viene segnalata in questo studio l'iniziativa positiva del Regno Unito denominata "*Researchfish*" (<https://www.researchfish.net/>) che, nonostante sia uno strumento che non fornisca liberamente i risultati delle proprie statistiche, tenta di fornire strumenti operativi per identificare e quantificare l'impatto dell'*Open Access* e più in generale dell'*Open Science*. Sul fronte degli avanzamenti negli aspetti di valutazione dell'impatto di politiche per i Dati Aperti (*Open Data*), si rileva nello studio come ci sia ancora poca attenzione sull'argomento rilevando che addirittura un terzo degli stati non ne sta discutendo affatto a livello istituzionale.

4.2 Nuove professionalità

Gli sviluppi in ambito digitale degli ultimi due decenni hanno incrementato drasticamente la complessità del sistema che ruota intorno ai dati digitali. La Commissione Europea ha redatto un interessante documento in cui sono state analizzati i nuovi approcci operativi al fare *Open Science* [EC DG-Research, 2017b] rilevando come siano sempre più necessarie nuove professionalità capaci di utilizzare e sfruttare appieno le potenzialità sia di infrastrutture di ricerca in corso di realizzazione, ma anche di quelle già esistenti. L'incremento della complessità porta quindi alla richiesta di nuove professionalità molto specifiche quali ad esempio *Data Manager*, *Digital Media Specialist*, *ICT Security Specialist*, *E-Learning Specialist*. Questi profili non sono facilmente reperibili nel mercato del lavoro, e le poche persone disponibili sono spesso esclusivo appannaggio delle

aziende private, anche solo a causa dell'impossibilità della PA di poter contrattare il profilo retributivo. È diffusa l'esperienza di concorsi per profili ICT andati a vuoto sia in ambito universitario e degli EPR, ed è pertanto raro imbattersi in dipendenti della PA che siano veri esperti professionisti ICT (*Information and Communications Technology*). È chiaro che le istituzioni che riescono ad assicurarsi uno di questi profili nel proprio organico fanno di tutto per tenerceli. Queste professionalità vengono inquadrare negli EPR come tecnologi o CTER (collaboratore tecnico degli enti di ricerca), e nelle università come PTA (personale tecnico amministrativo).

La scarsità dei profili necessari obbliga la PA, e quindi anche Università ed EPR, ad affidarsi a consulenze professionali demandando di fatto all'esterno il controllo di alcuni aspetti delle proprie infrastrutture. Al fine di permettere un maggiore controllo e consapevolezza di quali siano i profili professionali esistenti, quali siano le rispettive competenze e quali siano i migliori modi per stipulare contratti (di qualunque natura essi siano) nel mondo ICT che garantiscano buoni livelli di qualità e sicurezza, AgID rende pubbliche le "Linee guida per la qualità delle competenze digitali nelle professionalità ICT" [AgID, 2017a]. Grazie a questa guida, la PA è facilitata nel compito di individuare quali siano le professionalità di cui ha bisogno, e soprattutto è messa nelle condizioni di utilizzare un gergo adatto a parlare con il mondo delle agenzie di consulenza private. Un altro strumento utile per monitorare le professionalità ICT è l'osservatorio delle competenze digitali [AICA et al., 2017], un progetto comune tra Aica, Assinform, Assintel, Assinter, varie associazioni attive nel settore dell'ICT, che nel corso del 2017 ha analizzato oltre 175.000 annunci di lavoro pubblicati sul Web, con particolare attenzione alla domanda proveniente dalla Pubblica Amministrazione. Questo osservatorio è al quarto anno di attività e dai risultati emerge chiaramente come ormai tutte le professionalità hanno un grado più o meno elevato di pervasività delle competenze digitali, chiamato "Skill Digital Rate", per cui si auspica che l'importanza della verifica di tali competenze nelle sedi concorsuali tenga conto di queste necessità.

A livello europeo la situazione è resa ancora più complessa per la difformità con cui vengono definiti, concepiti, formati e selezionati i profili professionali nei diversi stati membri dell'Unione Europea. La difformità esistente è minore rispetto al passato, anche grazie all'attuazione del "Bologna Process", una serie di azioni iniziate nel 2010 definite nella "Bologna Declaration" siglata dai paesi europei nel 1999 e che ha portato alla creazione dell'European Higher Education Area (EHEA). Per cercare di armonizzare il panorama complesso ed eterogeneo delle professionalità digitali, la Commissione Europea ha lanciato nel giugno 2016 l'iniziativa **New Skills Agenda for Europe**. Tra le azioni proposte da questa agenda si segnala la *Digital Skills and Jobs Coalition*, che mira a coinvolgere gli stati membri dell'Unione, facendo cooperare le istituzioni educative e società private e delle associazioni no-profit attive nel settore della formazione professionale. L'attività di cooperazione intrapresa tra questi soggetti si potrà avvantaggiare del costituendo Quadro europeo delle qualifiche per l'apprendimento permanente [European Qualifications Framework, EQF, COM/2016/0383 final]. Nella definizione dell'EQF, si terrà conto delle definizioni approntate dall'*European Skills, Competences, Qualifications and Occupations* (ESCO) che traduce in tutte le lingue europee le diverse tipologie di profili professionali. Sempre nel quadro dell'armonizzazione tra gli stati membri, anche la modulistica pubblicata da Europass a cura di Cedefop si uniformerà alle linee guida che emergeranno dalle azioni menzionate, fornendo le basi per una più semplice circolazione delle professionalità a prescindere dallo stato in cui le professionalità sono state formate.

Vista la diffusa necessità di creare nuovi profili professionali adatti a gestire con competenza i vari aspetti legati al fare *Open Science*, stanno nascendo strumenti formativi ottimizzati per questo ambito. Tra questi si segnala il portale del progetto FP7 FOSTER "**Facilitating Open Science Training for European Research**" [Orth et al., 2017], una piattaforma che rende disponibile il materiale didattico adatto alla crescita delle competenze per perseguire l'*Open Science*. La formazione mira all'organizzazione dei dati prodotti dalla Ricerca, il loro ingresso nel ciclo della Ricerca, la loro disseminazione e loro archiviazione, tutte attività che ricadono

nell'ambito del **Research Data Management** (RDM). Grazie a queste attività è possibile assicurare l'attendibilità e la verificabilità dei risultati ottenuti, permettendo nuove e innovative ricerche basate sulle informazioni già esistenti [Whyte & Tedds, 2011], potenziando l'uso di strumenti come i **Data Management Plans** (DMP), lo sviluppo di politiche aperte per pubblicazioni (*Open Access*) e dati (*Open Data*), lo sviluppo di servizi web per la condivisione dei dati e sistemi alternativi per la misurazione dei risultati della Ricerca. Sempre nell'ambito dell'incremento delle conoscenze per permettere ai ricercatori di essere messi nelle condizioni di affrontare una migliore gestione dei propri dati, si segnala anche l'iniziativa "*Support Your Data*" promossa dalla California Digital Library che ha pubblicato un semplice guida che illustra azioni concrete adottabili direttamente dai ricercatori [Borghi et al., 2018].

4.3 Nuove soluzioni per la sostenibilità

Per quanto riguarda i finanziamenti, è fondamentale anzitutto considerare la necessità di riorientare Horizon 2020 e il prossimo programma quadro per estendere i dati attuali alle già esistenti Infrastrutture di ricerca (distribuite e virtuali) oltre il ciclo di vita di un progetto, per integrarle e federarle in modo sostenibile.

Tra gli Stati membri esiste già un elevato livello di consapevolezza circa la necessità di ridurre la frammentazione e i costi di funzionamento. Una combinazione di finanziamenti potrebbe essere immaginato come la combinazione tra fondi Strutturali, FFO (Fondo di Finanziamento Ordinario) e FOE (Fondo Ordinario per gli Enti e le istituzioni di ricerca finanziati dal Ministero).

Horizon 2020 offre già un finanziamento significativo di 2 miliardi Euro per costruire EOSC. Sarà necessario coprire altri 4,7 miliardi di euro di risorse combinate tra gli Stati membri che svolgeranno un ruolo di traino nell'iniziativa in parola.

I modelli di riferimento sono attualmente in corso di definizione in base a risorse di *governance* esistenti sia a livello europeo (per esempio GÉANT) e Livello globale (ad esempio il Forum di Belmont). Per stabilire sostenibili ipotesi di finanziamento è necessario individuare una struttura di *governance* chiara per prendersi cura delle "regole del gioco" e controllare i flussi di decisione, rispettando il principio di sussidiarietà.

La *governance* di EOSC, così come ipotizzata, contribuirà a prevenire duplicazioni di sforzi, frammentazione e soluzioni isolate. Inoltre, avrà la capacità di semplificare i processi decisionali in relazione al *science data sharing*.

Per migliorare le modalità di valutazione della qualità e dell'impatto dei risultati scientifici da parte delle agenzie di finanziamento e delle istituzioni accademiche vale la pena ricordare in questa sede la **San Francisco Declaration on Research Assessment** [DORA; ASCB, 2012] nata durante un dibattito all'incontro annuale di *American Society for Cell Biology* a San Francisco. La dichiarazione, sottoscritta da circa 900 istituzioni dedite alla ricerca e da circa 13000 ricercatori, ha sviluppato una serie di raccomandazioni che riassumiamo come segue:

- **Raccomandazioni generali.** Non usare le metriche (es.: *impact factor* delle riviste) come surrogato della qualità degli articoli, per valutare il livello di contributo dei ricercatori, per decidere assunzioni, promozioni, o decisioni di finanziamento.
- **Per gli enti finanziatori.** Usare criteri trasparenti di valutazione della produttività scientifica, esplicitando chiaramente che il contenuto scientifico hanno più valore delle metriche.
- **Per le istituzioni.** Esplicitare chiaramente i criteri usati per le assunzioni, gli incarichi, le promozioni includendo nelle valutazioni l'impatto di tutte le tipologie di prodotti scientifici, non solo degli articoli pubblicati.
- **Per gli editori.** Ridurre o cessare l'enfasi posta sulle metriche della propria rivista come elemento di promozione, o al limite, presentare le metriche contestualizzandole, come

presentando l'andamento degli ultimi 5 anni, o presentando diverse tipologie di metriche. Sarebbe auspicabile sviluppare l'uso di metriche basate sui contenuti scientifici, piuttosto che su sistemi che misurano il numero di citazioni. Promuovere la buona pratica degli autori di descrivere il loro livello di contributo nei lavori pubblicati. Tendere a rimuovere le limitazioni nel riutilizzo delle informazioni pubblicate, incentivando l'adozione di licenze *Creative Commons* che permettano tali pratiche. Rimuovere le eventuali limitazioni al numero di riferimenti bibliografici utilizzati, favorendo la citazione di lavori originali, non recensioni o lavori derivati di scarso valore.

- **Per le organizzazioni che forniscono metriche.** Adottare sistemi chiari, fornendo i dati di base utilizzati e i metodi di calcolo applicati (che dovrebbero adattarsi alle varie tipologie di contributi), e permettendo il riuso dei risultati ottenuti. Comunicare chiaramente che la manipolazione delle metriche non è tollerata, esplicitando cosa si intende per manipolazione e quali intenzioni si adotteranno per eliminarle.
- **Per i ricercatori.** Utilizzare criteri di valutazione basati sui contenuti scientifici quando si tratta di decidere dell'utilizzo di fondi, per assunzioni, cattedre, promozioni. Preferire la citazione di fonti originali e non recensioni o rielaborazioni, al fine di promuovere gli autori che per primi hanno dato un nuovo contributo. Usare molte metriche come indicatori per supportare le tesi sostenute. Contestare dove possibile le cattive pratiche che utilizzano in maniera errata le metriche delle riviste per prendere decisioni importanti.

4.4 Ricerca, Università e Pubblica Amministrazione

Nel Sistema Paese, Università e Enti Pubblici di Ricerca sono considerati Pubblica Amministrazione [ISTAT, 2017; Wikipedia italiano, 2017c], come i Ministeri, gli Istituti e le Scuole di ogni ordine e grado, le Aziende autonome, le Regioni, le Province, i Comuni, le Comunità Montane, gli istituti autonomi Case Popolari, le Camere di Commercio, Industria, Artigianato e Agricoltura, gli enti pubblici non economici nazionali (e.g. ACI), regionali e locali (e.g. Agenzie Regionali per la Protezione Ambientale), le amministrazioni, le Aziende Sanitarie Locali e gli enti del Servizio Sanitario Nazionale, l'Agenzia per la rappresentanza negoziale (e.g. ARAN) e le agenzie fiscali (e.g. Agenzia delle dogane e dei monopoli, Agenzia del demanio e Agenzia delle Entrate). *Fare Ricerca*, l'attività principale di Enti Pubblici di Ricerca e dell'Università, è un'attività **sperimentale, estremamente dinamica, spesso di nicchia, che rimette continuamente in discussione i suoi processi e che poco si sposa con procedure complesse e laboriose che troppo spesso cercano di ridurre le spese semplicemente spingendo all'adozione di economie di scala** o di favorire la rotazione tra i fornitori presenti sul mercato. Questo diventa ancora più problematico quando si parla di scelte tecnologiche. Nel nostro Paese, l'Agenzia per l'Italia Digitale [AgID, 2017b] ha il compito di elaborare una strategia nazionale, individuando l'insieme di azioni e norme per lo sviluppo delle tecnologie, dell'innovazione e dell'economia digitale, in linea con l'Agenda Digitale Europea. L'AgID guida le scelte tecnologiche della PA che, spesso, entrano in sovrapposizione o, addirittura, in contrapposizione con quelle fondamentali per la Ricerca, creando potenzialmente un problema di partecipazione a progetti europei e internazionali come nel caso della scelta della rete di connettività ad Internet e dell'identità digitale. In questi due casi, sono state fatte eccezioni o trovate soluzioni specifiche che sono artifici per risolvere l'incompatibilità che è stata riconosciuta. Questo rimane, però, un problema potenziale, in quanto l'AgID potrebbe non comprendere tempestivamente l'insorgere di una nuova incompatibilità, né trovare una soluzione in tempo. Il legislatore ha preso atto delle peculiarità del settore degli enti pubblici di ricerca (EPR), e ha recentemente introdotto delle riforme con la cosiddetta legge "Madia" (Legge 7 agosto 2015, n. 124, Art.13) e relativo decreto attuativo (D.Lgs. 25 novembre 2016, n. 218), in cui gli EPR vengono inquadrati in un "sistema di

regole più snello e appropriato”, ribadendone “l’autonomia decisionale, anche attraverso la riduzione dei controlli preventivi ed il rafforzamento di quelli successivi” e spingendo per un maggiore “coordinamento con le migliori pratiche internazionali”.

In questa realtà scientificamente e amministrativamente molto complessa non è facile orientarsi, e INGV ha istituito nel 2015 un gruppo chiamato “PoliDat” [INGV, 2015] che si occupa di definire una politica dei dati istituzionale ispirata ai principi dell’*Open Access*, di definire il quadro normativo a livello nazionale, europeo e internazionale, e di redigere regolamenti e indicazioni per i propri ricercatori compatibili col fare *Open Science* [INGV, 2016; 2017a; 2018a], anche tramite l’istituzione di un “Ufficio Gestione Dati” [INGV, 2018b]. Dal punto di vista infrastrutturale, ha istituito un gruppo di lavoro “Backup servizi e dati” [INGV, 2017b] col compito di progettare lo studio di fattibilità per le procedure di “*backup and restore*” per le sale operative del Centro Nazionale Terremoti, della Sezione di Napoli e di Catania e uno “Rete dati” [INGV, 2017c] con lo scopo di progettare, realizzare e coordinare le reti informatiche dell’Istituto per ottimizzare la loro gestione e per armonizzare la condivisione della base dati e dei dati registrati dalle reti di monitoraggio tra le diverse Sezioni dell’ente, nel modo più efficiente, sicuro e affidabile possibile. Anche le Università italiane hanno istituito nel 2016 un gruppo di lavoro transuniversitario che tenta di promuovere la diffusione dell’*Open Science*, un gruppo chiamato *Italian Open Science Support Group* (IOSSG) che fornisce utili linee guida in diversi ambiti [IOSSG, 2016].

4.5 European Open Science Cloud (EOSC)

Ad aprile 2016 [Comunicazione CE COM(2016) 0178 final] la Commissione Europea ha riconosciuto l’esigenza di realizzare un piano per lo sviluppo del *Cloud Computing* al fine di evitare di trattare altrove i dati della ricerca europea per l’assenza di un’adeguata rete infrastrutturale di calcolo distribuito.

A tal fine, a gennaio 2017 è stato presentato il primo progetto pilota di *European Open Science Cloud* (EOSC), chiamato appunto EOSC Pilot [EOSC, 2017], finanziato per 10 milioni di euro e coordinato da *Science and Technology Facilities Council* (STFC). Il progetto, la cui struttura è ancora in fase di preparazione, vede la partecipazione di numerose agenzie di finanziamento, di *e-Infrastructure*, come EGI, GÉANT, PRACE, EUDAT, e di primari enti di ricerca in tutta Europa, tra cui gli italiani CNR, INAF, INFN e INGV, e di consorzi italiani, come CINECA e GARR. Nello specifico, l’azione dell’EOSC si sostanzierà in uno «spazio sicuro ed aperto» in cui la comunità scientifica potrà «archiviare, condividere e riutilizzare dati e risultati scientifici».

L’EOSC Pilot progetterà, stabilirà e sosterrà un quadro di *governance* orientato agli *stakeholder* con il coinvolgimento diretto dei soggetti coinvolti e per raggiungere tale fine è stato istituito il *Governance Development Forum* [EGDF, 2017].

In particolare, con tale iniziativa, si intende proporre un’**infrastruttura federata di dati e servizi di accesso, analisi, interoperabilità e comunicazione scientifica, gestita attraverso un modello di governance “stakeholder driven”**, cioè in grado di coinvolgere efficacemente tutti i portatori di interesse nella gestione dell’iniziativa, dagli utenti, ai *service provider*, agli enti di ricerca, fino ai finanziatori.

La call EOSC Pilot ha messo insieme le diverse realtà a livello europeo e nazionale, ma al tempo stesso ha evidenziato le differenti visioni esistenti nella comunità. Pertanto è necessario controllare che il lavoro svolto nel Pilot di EOSC sia coerente con le vigenti legislazioni dei singoli Stati membri, incoraggiando l’adozione di nuovi modi di lavoro e di pratiche scientifiche condivise a livello sovranazionale. La Commissione sta attualmente intraprendendo un intenso lavoro preparatorio per l’attuazione dell’iniziativa in parola. Gli Stati membri hanno un ruolo fondamentale nell’attuazione dell’iniziativa *Cloud*, sia come finanziatori di ricerca che come responsabili politici con un diretto mandato sulle infrastrutture di dati scientifici esistenti.

Da un punto di vista squisitamente politico occorre mappare i contributi italiani nelle diverse iniziative Europee e quantificarne l'impegno tecnologico, le risorse umane e finanziarie messe a disposizione (*cash, in-kind e secondment*), inserendo nel piano strategico nazionale la quantificazione reale del contributo italiano per EGI, EUDAT e PRACE. A tal fine, il MIUR dovrebbe predisporre un *framework* di coordinamento nazionale tra *Research Infrastructures* ed *e-Infrastructures*. Questa operazione è fondamentale per capire e valutare il modello di *governance* e sostenibilità che sarà discusso e adottato da EOSC.

Sul piano della *governance* occorre chiarire, infine, cosa significhi adottare un approccio federato per l'EOSC: una soluzione potrebbe essere quella di una federazione di contributi nazionali a livello europeo, identificando i *service providers* a livello nazionale e definirne i costi totali.

I benefici per il sistema nazionale sarebbero evidenti: dall'ottimizzazione dei costi, al controllo sul ritorno per il paese in termini di visibilità e riconoscimento in ambito internazionale. Solo allineando le pratiche in Europa, lavorando verso una versione sostenibile e federata dell'EOSC che mira ad accelerare e supportare la transizione corrente a scienza più aperta verso un unico mercato digitale, si porranno le basi solide per un accesso a servizi e sistemi condivisi, promuovendo al contempo la semplificazione nel riutilizzo di dati scientifici. Questo dovrebbe essere fatto basandosi sul successo di esistenti sistemi informatici, riducendo la frammentazione attraverso la creazione di un ecosistema di infrastrutture. Secondo la Commissione, l'EOSC, vuole conferire all'Europa un ruolo di leadership globale nelle infrastrutture dei dati scientifici, affinché gli scienziati europei raccolgano tutti i vantaggi della scienza dei dati. Queste azioni si baseranno sulle risorse e sulle capacità già messe a disposizione dalle infrastrutture di ricerca e dalle infrastrutture virtuali dei singoli Stati, massimizzando l'utilizzo in tutta la comunità scientifica di riferimento. Ciò ridurrà la frammentazione tra le infrastrutture lavorando attraverso settori scientifici ed economici, paesi e modelli di *governance*, migliorando l'interoperabilità tra le infrastrutture, dimostrando come i dati e le risorse possano essere condivisi anche quando sono grandi e complessi e in vari formati.

EOSC Pilot ambisce a migliorare le capacità di preservare e riutilizzare i dati, un passo importante verso la creazione di un ambiente di innovazione aperto e affidabile dove i dati provenienti da ricerche pubbliche finanziate sono aperti con chiari incentivi per la condivisione di dati e risorse. Tra gli obiettivi dell'EOSC:

- a. Obiettivo della *Governance* - Progettare e sperimentare un quadro di *governance* orientato agli *stakeholder* con il coinvolgimento di comunità di ricerca, istituti di ricerca, infrastrutture di ricerca, incluse le infrastrutture informatiche e gli organismi di finanziamento, a formare e sorvegliare lo sviluppo futuro della *European Open Science Cloud*. I modelli di riferimento sono attualmente in corso di definizione in base a risorse di *governance* esistenti sia a livello europeo (per esempio GÉANT) e livello globale (ad esempio il Forum di Belmont). È necessaria una struttura di *governance* per prendersi cura delle "regole del gioco" e controllare le decisioni, rispettando il principio di sussidiarietà. La *governance* dell'EOSC contribuirà a prevenire duplicazioni di sforzi, frammentazione e soluzioni isolate, abbassando le barriere all'interazione tra gli stati. Inoltre, avrà la capacità di semplificare i processi decisionali in relazione al *science data sharing*.
- b. Obiettivo Scienza - Sviluppare un certo numero di casi scientifici di applicazione in alcuni settori di riferimento che mostreranno la rilevanza e l'utilità dei Servizi EOSC e la loro abilitazione al riutilizzo dei dati. I casi indicati serviranno a far comprendere le potenzialità dell'EOSC supportandone lo sviluppo.
- c. Obiettivo Servizi - Creare una serie di servizi pilota di EOSC che federino dati, infrastrutture e servizi che promuovano la ricerca multidisciplinare attraverso le frontiere geografiche e nel lungo periodo, attraverso la conservazione dei dati. Inizialmente la base di utenti sarà circoscritta alla comunità scientifica, ma, successivamente, la Commissione intende estendere i servizi anche alla PA e all'industria, con la creazione di «soluzioni e

tecnologie che apporteranno vantaggi a tutti i settori dell'economia e della società», ai fini dello sviluppo del mercato unico digitale.

- d. Obiettivo di interoperabilità - L'interoperabilità dei dati richiede norme tecniche specifiche, la certezza del diritto per quanto concerne l'accesso e l'utilizzo, e la condivisione dei dati della ricerca. La piena condivisione dei risultati delle attività scientifiche, infatti, è «ostacolata anche dalla loro dimensione, dai diversi formati, dalla complessità dei software necessari per analizzarli» e da una profonda «settorialità» delle diverse discipline. Detta problematica, può essere risolta adottando «metadati comuni» per lo scambio e utilizzo di dati, per poi, successivamente, renderli ampiamente accessibili da parte di coloro che intendono elaborarli mediante strumenti di analisi di dati comuni. A tale fine occorre definire e implementare specifiche interfacce, standard e processi che consentano e supportino l'interoperabilità e la condivisione dei dati e delle infrastrutture EOSC in tutte le discipline e tra i fornitori.
- e. Obiettivo di impegno a livello comunitario - Sviluppare *standard* e criteri comuni di valutazione per assicurare che le organizzazioni e gli individui siano motivati a sviluppare le capacità e le competenze che si basano sull'EOSC e implementare una strategia di formazione in materia di EOSC e coordinarne la fornitura di servizi. Individuare e riunire, attraverso un efficace piano di impegni e di strategia di comunicazione, i principali gruppi di *stakeholder* del settore della ricerca, del settore privato e pubblico, accoppiati a sostenere il progetto attraverso una strategia efficace di comunicazione e approfondimento basata su contenuti orientati ai risultati. Queste azioni strategiche soddisferanno le principali attività e i principali obiettivi delle infrastrutture di ricerca ESFRI, tra cui la più volte citata EPOS.

Il lancio dell'EOSC Pilot, rappresenta un'opportunità senza precedenti per EPOS, per sfruttare appieno il potenziale dei dati, in particolare i dati inerenti le discipline delle Scienze della Terra. EPOS ha già sviluppato alcuni servizi finalizzati all'integrazione dei dati e metadati, coinvolgendo utenti e *stakeholders*. Sebbene EOSC possa rappresentare una soluzione per la federazione di infrastrutture e fornitori tecnologici esistenti e nuovi, EPOS non sarà in grado di delegare a EOSC la manutenzione dei suoi servizi integrati e tematici. Al pari di EOSC, la visione di EPOS, relativa all'infrastruttura europea dei dati, è costituita dall'assegnazione di un ruolo fondamentale alle infrastrutture di ricerca paneuropea con sforzi, risorse e capacità dedicate per federare i fornitori di dati e le infrastrutture all'interno di singoli settori scientifici.

L'*European Data Infrastructure* non dovrebbe essere considerata solo come un'opportunità per la realizzazione di supercomputer; piuttosto potrebbe rappresentare un'opportunità per adottare un modello coerente per la federazione delle infrastrutture di ricerca in sinergia con gli sforzi di EOSC per la **federazione delle infrastrutture informatiche**. Per migliorare l'interoperabilità dei sistemi, il progetto EOSC Pilot, punta a creare un'appropriata struttura di governo comune in tutta l'UE e a fornire le specifiche per la condivisione dei dati tra diverse discipline e infrastrutture. L'iniziativa afferma che la *privacy* e la protezione dei dati saranno basati su standard riconosciuti e garantiti dal design dell'EOSC. I metodi suggeriti per garantire la conformità della proposta con la legge sulla protezione dei dati comprendono l'anonimato irreversibile dei dati sensibili prima della loro integrazione con altre fonti e la creazione di spazi di *personal data* all'interno del *Cloud*. Al fine di dimostrare l'effettiva perseguibilità del piano, la Commissione Europea si propone di arrivare a collegare le prioritarie infrastrutture europee di ricerca all'*European Open Science Cloud* entro il 2017.

Il 26 ottobre 2017 la Commissione Europea ha pubblicato una dichiarazione a supporto dell'EOSC che mira a coinvolgere tutti i soggetti scientifici potenzialmente interessati al fine di coinvolgerli [Science|Business, 2017] e di arrivare alla realizzazione dell'EOSC entro il 2020. Dopo un mese oltre 70 soggetti, tra cui università, centri di ricerca, associazioni e fondazioni, hanno

sottoscritto la dichiarazione d'intenti. Oltre alle consultazioni intraprese dalla Commissione Europea in ambito scientifico, alcuni soggetti potenzialmente interessati alla costituzione e all'uso dell'EOSC hanno posto l'attenzione sugli aspetti economici relativi alla sua costituzione. L'associazione Science|Business [2018], composta da soggetti privati, della ricerca e istituzionali, riconosce il potenziale di innovazione e aumento della produttività della ricerca europea pubblica e privata, ma ha anche posto l'accento sui costi connessi per garantire la sostenibilità a lungo termine, di supporto alla pubblicazione di FAIR data, dei sistemi di interconnessione tra i diversi data provider, dello sviluppo di sistemi di sicurezza informatica e del sistema di governo dello stesso EOSC.

Il suggerimento di Science|Business per rendere più efficiente la spesa per la costruzione dell'EOSC, è quella di adottare strategie guidate dalla domanda concreta di servizi e dati da parte dei ricercatori, invece che costruire qualcosa basandosi su teorie ipotizzando che poi l'interesse della comunità scientifica e i fondi connessi alle attività verranno alla fine, quando EOSC sarà implementato.

A gennaio 2018 la Commissione Europea ha iniziato un progetto H2020 triennale da oltre 30 milioni di euro complementare a EOSC Pilot chiamato EOSC-hub con oltre 150 partner provenienti da oltre 50 paesi. Mentre EOSC Pilot si concentrerà sulla progettazione della governance, EOSC-hub si occuperà della progettazione dell'infrastruttura tecnologica a partire dalle risorse della federazione EGI come EUDAT CDI, INDIGO DataCloud, GÉANT ed altre realtà europee che si occupano della gestione di servizi collegati ai dati della ricerca.

A marzo 2018, a seguito della dichiarazione di supporto all'EOSC e alle molteplici consultazioni intercorse tra la Commissione Europea, gli ambienti scientifici europei e i maggiori soggetti istituzionali coinvolti, è stata approvata la tabella di marcia per la costituzione dell'EOSC [EC Staff Working Group SWD (2018), 83 final] in cui si dettaglia il piano d'intervento, il cui principale auspicio è quello di ridurre l'attuale frammentazione caratterizzata da una bassa efficienza e scarso supporto ai FAIR data. Il piano presentato cerca di inquadrare tutte le attività proposte nell'ambito della costruzione del *Digital Single Market* e la costruzione di un'economia dei dati europea [COM(2017) 09 final]. Infine, il piano analizza e chiarisce le relazioni con le iniziative condotte dalle infrastrutture selezionate e supportate da ESFRI, ponendo molta attenzione alla loro sostenibilità [SWD (2017), 323 final]. Il piano presentato identifica sei linee d'azione e una precisa tempistica, azioni che cercano di mettere a frutto i risultati ottenuti nell'ambito dei progetti H2020 che tanto spingono per la pubblicazione di FAIR data e passando per un maggiore coinvolgimento delle iniziative condotte dai singoli stati membri dell'Unione. Il piano prevede una versione preliminare ma funzionante dell'EOSC entro la seconda metà del 2020 in cui saranno incluse le infrastrutture di dati che saranno giudicate idonee entro la fine del 2019. Dopo il 2020, quando andrà a regime, gli studi della Commissione Europea sostengono che EOSC potrà autosostenersi finanziariamente, ipotesi che pare validata dalle valutazioni di Science|Business [2018]. Recentemente è stato reso disponibile un primo rapporto sullo stato di avanzamento dei lavori per la realizzazione dell'EOSC che cerca di focalizzare le principali problematiche emerse cercando al contempo di individuare alcune soluzioni pratiche [EC DG-Research, 2018b].

4.6 Modelli di business delle e-Infrastructure (intervista a Sanzio Bassini)

Nello svolgimento di questo lavoro, abbiamo avuto la possibilità di affrontare tematiche inerenti a *Big Data* e *Open Science* con esperti del settore. In un'intervista a Sanzio Bassini, direttore del dipartimento *SuperComputing Applications and Innovation* del CINECA e Presidente del Concilio di PRACE nel periodo 2014-2016, è emerso uno dei temi fondamentali: i modelli di *business* delle *e-Infrastructure*. Il Dott. Bassini spiega come l'integrazione, il consolidamento e la federazione di infrastrutture tecnologiche europee partono da modelli tanto diversi. La ragione

principale è storica. L'Università e EPR hanno costruito queste infrastrutture comprendendone l'enorme potenziale ai fini della Ricerca: ad esempio, hanno permesso ai ricercatori di connettersi attraverso una rete pubblica (GÉANT) che non fosse legata a scelte di *business*, e hanno promosso l'accesso gratuito a supercalcolo performante (PRACE). Il Dott. Bassini evidenzia come le *e-Infrastructure* basilari prese in considerazione, adottino quattro modelli di *business* diversi: universale, a contributo, a brokeraggio e di buona volontà.

GÉANT è considerata l'infrastruttura tecnologica di base per eccellenza, in quanto su di essa transitano tutte le comunicazioni telematiche. Senza GÉANT le altre infrastrutture non esisterebbero. Nel modello di GÉANT, che potrebbe essere definito "*universale*", le singole nazioni mantengono le proprie Reti Nazionali della Ricerca (NREN) mentre la Comunità Europea sovvenziona le infrastrutture di interscambio per connetterle. Il modello è basato sulla reciprocità, in quanto ogni Stato mette a disposizione degli altri la propria rete della ricerca.

La missione di PRACE è promuovere l'impatto delle scoperte scientifiche e lo sviluppo delle tecnologie attraverso tutte le discipline per migliorare la competitività europea a beneficio della società. PRACE cerca di realizzare questa missione offrendo al ricercatore la possibilità di accedere a servizi di supercalcolo e di gestione dei dati attraverso una *peer review*. Tutti i ricercatori degli Stati membri di PRACE possono partecipare a bandi per vincere ore di calcolo. Il modello di *business* di PRACE è basato sulla competitività con lo scopo di promuovere la "buona ricerca". Il modello di PRACE potrebbe essere definito "*a contributo*" (dal motto *to compute or to contribute*). In questo modello, è stato definito un numero minimo di risorse computazionali da condividere (i.e. *to compute*). Gli Stati che non possono mettere a disposizione tali risorse, contribuiscono alle spese del team di supporto specialistico (i.e. *to contribute*) per le infrastrutture di calcolo di quelli che le condividono.

EGI realizza servizi *High Throughput Computing* per ricercatori e progetti internazionali sia del mondo della Ricerca che del *business*, federando *cloud providers* e *data center* pubblici e privati. Nel modello di *business* "*a brokeraggio*" di EGI viene assegnato un valore alle risorse. Le risorse non usate dai proprietari sono a disposizione di terzi. In momenti diversi, è possibile accedere a risorse di terzi con un modello *pay for use a somma zero*. Questo significa che, nel computo totale, il numero delle risorse sfruttate (i.e. debito) deve essere pari a quelle date in condivisione (i.e. credito). Questo modello è stato pensato per promuovere lo sfruttamento a saturazione delle risorse in ogni istante. Il tema più caldo e, ancora, non risolto rimane l'immagazzinamento, la conservazione e la preservazione dei dati.

EUDAT è la *e-Infrastructure* per i servizi sulla gestione, l'analisi e il riuso dei dati della ricerca e, per questo motivo, in linea con il piano *European Open Science Cloud*. Le infrastrutture e i suoi servizi sono stati sviluppati in collaborazione con comunità di ricerca di differenti discipline scientifiche e coinvolte in tutte le parti del processo. Purtroppo il modello di *business* di EUDAT è di "*buona volontà*", in quanto non esiste un modello adeguato di sostenibilità. EUDAT è un progetto nato dal FP7 e finanziato da H2020. La parte più calda rimane la conservazione e la preservazione dei dati che implicano un mantenimento a lungo termine e, di conseguenza, la necessità di un modello di sostenibilità chiaro, anche se nel febbraio 2018 è diventata una società privata (EUDAT Ltd.) non a scopo di lucro.

Una possibile infrastruttura EOSC potrebbe essere realizzata tramite la combinazione di EGI, (per la gestione del *Cloud Computing*) e EUDAT, per la conservazione dei dati. Inoltre, risolverebbe il problema della sostenibilità di EUDAT. GÉANT essendo ritenuta l'infrastruttura telematica di base viene finanziata singolarmente. Rimane da capire come sfruttare la potenza del supercalcolo all'interno di EOSC. Una soluzione interessante è legata alla Emilia Romagna Big Data Community che ha collegato direttamente il Tier-1 HTC del CNAF, che lavora sulla infrastruttura EGI, con il Tier-0 del supercalcolo del CINECA, che lavora sia sull'infrastruttura PRACE che EUDAT. Questa soluzione innovativa potrebbe portare la Regione Emilia Romagna ad essere un Tecnopolo di eccellenza in tutta Europa.

Conclusioni

Secondo la Commissione Europea e il Consiglio Europeo, l'accesso alle grandi moli di dati prodotti dalla ricerca scientifica, la loro libera circolazione e il loro efficace utilizzo sono una condizione necessaria per un adeguato sviluppo di una società innovativa e della conoscenza. Affinché i dati possano avere un impatto innovativo e generare ricadute economiche sul sistema Paese, questi devono risultare facilmente accessibili e utilizzabili per differenti scopi commerciali, in particolar modo per le Piccole e Medie Imprese.

La ricerca condotta ha permesso di analizzare il complesso ecosistema dell'*Open Science*, delle principali infrastrutture di ricerca e tecnologiche coinvolte, degli ambiti di *governance* e di sostenibilità, fornendo una panoramica documentata che permette di evidenziare i nodi ancora da sciogliere, e, di conseguenza, dove è necessario concentrare i maggiori sforzi.

Attuare l'*Open Science* significherà risolvere le attuali problematiche connesse alla gestione della proprietà intellettuale, chiarendo quali siano i soggetti demandati alla scelta e all'apposizione di specifiche licenze, trovando soluzioni efficaci per la gestione di eventuali incompatibilità in termini di utilizzo. Sarà necessario modificare le attuali metriche e i sistemi di valutazione e incentivazione della produzione scientifica, monitorando gli utilizzi dei prodotti della ricerca per dimostrarne le ricadute sulla società. Questi cambiamenti saranno possibili solo migliorando la sinergia tra chi fa ricerca e chi svolge un ruolo di impulso politico, poiché entrambi saranno chiamati a rispondere a una società sempre più connessa e quindi sempre più presente nelle attività scientifiche.

L'EOSC rappresenta un ambito privilegiato per creare un modello volto ad assicurare la sostenibilità a lungo termine per la condivisione di *Big Data* e permetterà di sperimentare un sistema per la compensazione dei diritti di proprietà intellettuale, in relazione all'accesso e all'utilizzo di set di dati specifici. Questo livello di ambizione richiederà un forte coinvolgimento degli Stati membri dell'Unione, in particolare, mediante l'utilizzo dei fondi strutturali e delle garanzie del Fondo Europeo per gli Investimenti Strategici (EFSI).

Gli elementi considerati in questo testo costituiscono un solido riferimento per la comunità scientifica, per i finanziatori della ricerca e per i decisori coinvolti nell'elaborazione delle iniziative relative all'*Open Science*.

Per quanto riguarda l'analisi relativa ai modelli di *business* adottati per garantire la sostenibilità alle infrastrutture tecnologiche descritte, si è rilevata una forte eterogeneità delle realtà esistenti a livello europeo: in quest'ottica diventa difficile pensare ad un modello unico di riferimento. Una possibile infrastruttura alla base dell'EOSC potrebbe essere realizzata tramite la combinazione di EGI, per la gestione del *Cloud Computing* ed EUDAT, per la gestione e conservazione dei dati. GÉANT, essendo ritenuta l'infrastruttura di base per eccellenza, è finanziata direttamente dagli Stati membri e dalla Commissione Europea. Rimane da risolvere il nodo legato a come sfruttare la potenza del supercalcolo all'interno di EOSC: una soluzione interessante emersa durante la nostra analisi pare essere quella dell'Emilia-Romagna Big Data Community che ha collegato direttamente il Tier-1 di calcolo distribuito del CNAF su infrastruttura EGI, con il Tier-0 del supercalcolo del CINECA, su infrastruttura PRACE. Questa soluzione innovativa potrebbe portare la Regione Emilia Romagna ad essere un Tecnopolo di eccellenza a livello europeo.

Nella nostra ricerca emerge che uno dei più grandi vincoli al cambiamento, necessario all'implementazione dell'*Open Science*, è legato alla lentezza, pervasività, scarsa elasticità e contraddittorietà della burocrazia imperante nella Pubblica Amministrazione italiana. Fare Ricerca è un'attività sperimentale per definizione, dinamica ed estremamente settoriale che poco si sposa con procedure complesse e laboriose che cercano di contrarre la spesa pubblica spesso basandosi su tentativi di implementare economie di scala laddove queste non potranno mai dare i risultati attesi. Negli ultimi anni si è assistito ad un livello crescente di invasività capillare degli enti regolatori che, seppure con lo scopo di ottimizzare le già scarse risorse economiche

disponibili e di armonizzare gli strumenti tecnologici, impongono regolamenti e linee guida che impattano drasticamente sulla capacità d'innovazione auspicata nel settore della Ricerca. Il personale dedicato alla realizzazione delle infrastrutture tecnologiche, già di per sé sottodimensionato, anziché concentrarsi sulle sole attività di ricerca e sperimentazione, viene frequentemente impiegato in processi macchinosi legati all'espletamento delle pratiche burocratiche. Le linee guida indicate dall'AgID impongono indiscriminatamente a tutta la Pubblica Amministrazione scelte tecnologiche che, spesso, entrano in sovrapposizione o, addirittura, in contrapposizione con quelle richieste per la partecipazione a progetti europei e internazionali. Per risolvere questi problemi tipicamente italiani, sono state compiute spesso scelte di compromesso che altre istituzioni europee non hanno dovuto affrontare.

Un miglioramento a questa situazione può essere effettuato con l'entrata in vigore della cosiddetta Legge "Madia" (Legge 7 agosto 2015, n. 124, Art.13) e del relativo decreto attuativo "Semplificazione delle attività degli enti pubblici di ricerca" (D.Lgs. 25 novembre 2016, n. 218) che dovrebbe portare un netto miglioramento nella dinamicità e snellezza delle procedure, grazie ai "principi di responsabilità ed autonomia decisionale" e alla "razionalizzazione e semplificazione dei vincoli amministrativi, contabili e legislativi" concessi agli Enti Pubblici di Ricerca come l'INGV. Alla luce di questa riforma, il miglioramento della gestione è ora in capo alle singole amministrazioni che, mediante strumenti di autogestione come lo Statuto e i Piani Triennali di attività, hanno un maggiore margine di libertà, pur nel rispetto delle linee guida presenti nel Programma Nazionale della Ricerca pubblicato dal MIUR.

Dalle nostre analisi risulta chiaro come il percorso che porterà all'attuazione di buone pratiche basate sull'*Open Science* è ancora lungo e complesso, le sfide in gioco molte ed eterogenee, attuabili esclusivamente con il contributo attivo e coordinato dei diversi soggetti coinvolti, a livello governativo, a livello degli istituti che fanno Ricerca e a livello della società. Il nostro lavoro vuole dare degna rappresentazione ad alcune tra le novità più significative emerse negli ultimi anni partendo da un approccio tradizionale per arrivare all'*Open Science*. Secondo le intenzioni della Commissione Europea questa rivoluzione permetterà di passare da un sistema estremamente competitivo come quello attuale che ha ampi margini di inefficienza assolutamente non competitivi a livello globale, a pratiche più collaborative, che, grazie ai nuovi strumenti del *Cloud Computing*, dovrebbero permettere all'Europa di sfruttare appieno il suo primato in termini di produzione di dati scientifici a livello mondiale, rendendola più attrattiva per i ricercatori di tutto il mondo.

Bibliografia

- AgID, Agenzia per l'Italia Digitale, (2017a). *Linee guida per la qualità delle competenze digitali nelle professionalità ICT*. Aggiornamento del manuale operativo "Dizionario dei profili di competenza per le professioni ICT". <http://open.gov.it/wp-content/uploads/2017/05/professioni-ICT.pdf>
- Aica, Assinform, Assintel, Assinter, (2017). *Osservatorio delle competenze digitali 2017. Scenari, gap, nuovi profili professionali e percorsi formativi*. Report del progetto condotto in collaborazione con AgID e MIUR. http://www.agid.gov.it/sites/default/files/osservatorio_competenze_digitali_2017.pdf
- Allcott H., Gentzkow M. (2017). *Social Media and Fake News in the 2016 Election*. National Bureau of Economic Research (NBER), Working Paper No. 23089. DOI: <https://doi.org/10.3386/w23089>
- ALLEA, All European Academies, (2017). *The European Code of Conduct for Research Integrity, Revised Edition*. Berlin, 11 pp. http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020-ethics_code-of-conduct_en.pdf
- ANDS, Australian National Data Service, (2011). *Persistent Identifiers*. Australian National Data Service Guides, Expert level. <http://ands.org.au/guides/persistent-identifiers-expert.pdf>
- ANVUR, Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca, (2011). *Sul documento*

- ANVUR relativo ai criteri di abilitazione scientifica nazionale: commenti, osservazioni critiche e proposte di soluzione. http://www.anvur.org/images/Riferimenti_normativi/Documento_02_11.pdf
- ANVUR, Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca, (2015). ORCID ID: il nuovo identificativo dei ricercatori italiani. Progetto IRIDE, ultimo accesso ottobre 2015. http://www.anvur.org/index.php?option=com_content&view=article&id=829
- Archer P., Goedertier S., Loutas N., (2013). *Study on persistent URIs: with identification of best practices and recommendations on the topic for the Member States and the European Commission*. <http://philarcher.org/diary/2013/uripersistence>
- ASCB, American Society for Cell Biology, (2012). *San Francisco Declaration on Research Assessment (DORA)*. <http://www.ascb.org/files/SFDeclarationFINAL.pdf>
- Atakan K., Bazin PL, Bozzoli S., Freda C., Giardini D., Hoffmann T., Kohler T., Kontkanen P., Lauterjung J., Pedersen H., Saleh K., Sangianantoni A., (2017). *Setting the stage for the EPOS ERIC: Integration of the legal, governance and financial framework*. EGU General Assembly 2017, Geophysical Research Abstracts, vol.19, EGU2017-12890. <http://meetingorganizer.copernicus.org/EGU2017/EGU2017-12890.pdf>
- Bailo D., Paciello R., Rabissoni R., Sbarra M., Vinciarelli V., (2018). *Integration of heterogeneous data, software and services in Solid Earth Sciences: the EPOS system design and roadmap for the building of Integrated Core Services*. Rapporti Tecnici INGV, vol.393, 22pp. <http://www.ingv.it/editoria/rapporti/2018/rapporto393/>
- Barbera R., Ardizzone V., Ciuffo L. (2009). *Grid INFN Virtual Laboratory for Dissemination Activities*. In: Udoh E. and Zhigang Wang F. (eds.), "Handbook of Research on Grid Technologies and Utility Computing: Concepts for Managing Large-Scale Applications", IGI Global. DOI: <http://doi.org/10.4018/978-1-60566-184-1.ch024>
- Berners-Lee T., (1998). *Semantic Web Road map*. <https://www.w3.org/DesignIssues/Semantic.html>
- Berners-Lee T., Hendler J. and Lassila O., (2001). *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American, 284(5). <https://www.scientificamerican.com/magazine/sa/2001/05-01/>
- Borghi J., Abrams S., Lowenberg D., Simms S., Chodacki J., (2018). *Support Your Data: A Research Data Management Guide for Researchers*. Research Ideas and Outcomes, 4:e26439. DOI: <https://doi.org/10.3897/rio.4.e26439>
- CCSDS, Consultative Committee for Space Data Systems, (2012). *Reference Model for an Open Archival Information System (OAIS)*. Washington, DC, CCSDS Secretariat. <https://www.iso.org/obp/ui/#iso:std:iso:14721>
- CRUI, (2013a). *Position statement sull'accesso aperto sui risultati della ricerca scientifica in Italia*. http://www.cnr.it/sitocnr/lservizi/Biblioteche/Position_statement_OA_IT.pdf
- Cox A.M. and Ting Tam W.W., (2018). *A critical analysis of lifecycle models of the research process and research data management*. Aslib Journal of Information Management, 70(2):142-157. DOI: <https://doi.org/10.1108/AJIM-11-2017-0251>
- DCC, Digital Curation Centre (2013). *Checklist for a Data Management Plan, v4.0*. http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf
- Dolenc J., Hünenberger P., Renn O., (ed.) (2016). *Metrics in Research – For better or worse?* ETH Zurich, Infozone Special Issue S1. http://www.infozentrum.ethz.ch/uploads/user_upload/pdf/PDFs_Infozone/Infozone_Special_Issue_1.pdf
- DPC, Digital Preservation Coalition, (2015). *Digital Preservation Handbook, 2nd Edition*. <http://handbook.dpconline.org/>
- EASE, European Association of Science Editors (2007). *EASE statement on inappropriate use of impact factors*. https://www.ease.org.uk/wp-content/uploads/ease_statement_ifs_final.pdf
- EC DG-Informatics, Directorate General Informatics of the European Commission, Interoperability unit (2017). *New European Interoperability Framework, Promoting seamless services and data flows for European public administrations*. https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission (2016)(a). *Open Innovation Open Science Open to the World - a vision for Europe*.

- DOI: <http://doi.org/10.2777/061652>
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission (2016)(b). *Realising the European Open Science Cloud*. First report and recommendations of the Commission High Level Expert Group Research and on the European Open Science Cloud. https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission (2017)(a). *Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission (2017)(b). *Providing researchers with the skills and competencies they need to practise Open Science*. 36pp. DOI: <http://doi.org/10.2777/121253>
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission, (2017c). *Sustainable European Research Infrastructures - A call for action*. 60 pp. DOI: <http://doi.org/10.2777/76269>
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission, (2018a). *Access to and preservation of scientific information in Europe*. Report on the implementation of Commission Recommendation C(2012) 4890 final - Study, 96 pp. DOI: <http://doi.org/10.2777/642887>
- EC DG-Research, Directorate-General for Research and Innovation of the European Commission, (2018b). *Prompting an EOSC in practice*. Interim report and recommendations of the Commission 2nd High Level Expert Group [2017-2018] on the European Open Science Cloud (EOSC). 88 pp. https://ec.europa.eu/info/sites/info/files/conferences/eosc_summit_2018/prompting_an_eosc_in_practice_eosc_hleg_interim_report.pdf
- EDP, European Data Portal, (2015). *Creating Value through Open Data*. Capgemini Consulting for the European Data Portal Project. https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf
- EDP, European Data Portal, (2017). *Re-using Open Data - A study on companies transforming Open Data into economic & societal value*. Capgemini Consulting for the European Data Portal Project. https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf
- EGI, European Grid Infrastructure, (2015). *Open Science Commons*, v3. White paper, European Grid Infrastructure (EGI). <http://go.egi.eu/oscpw>
- e-IRG, e-Infrastructure Reflection Group, (2013). *White paper*. Report "eirg-1005", H2020 Project "e-Infrastructure Reflection Group Support Programme 5" (e-IRGSP5). <http://e-irg.eu/documents/10920/11274/e-irg-white-paper-2013-final.pdf>
- e-IRG, e-Infrastructure Reflection Group (2017a). *Evaluation of e-Infrastructures and the development of related Key Performance Indicators*. Report "eirg-1005", H2020 Project "e-Infrastructure Reflection Group Support Programme 5" (e-IRGSP5). <http://e-irg.eu/catalogue/eirg-1005>
- e-IRG, e-Infrastructure Reflection Group, (2017b). *Guide to e-Infrastructure Requirements for European Research Infrastructures*. Report "eirg-1004", H2020 Project "e-Infrastructure Reflection Group Support Programme 5" (e-IRGSP5). <http://e-irg.eu/catalogue/eirg-1004>
- Elta Smith E., Gunashekar S., Lichten C., Parks S., Chataway J. (2016). A framework to monitor open science trends in the EU. RAND Europe, Cambridge UK. https://www.oecd.org/sti/063%20%20OECD%20Blue%20Sky%202016_Open%20Science.pdf
- EPOS, European Plate Observing System, (2016). *Legal framework, Governing EPOS, Data Policy and Access Rules*. <https://www.epos-ip.org/node/167/pdf>
- ESFRI, European Strategy Forum on Research Infrastructures, (2016). *Strategy report on research infrastructures, Roadmap 2016*. https://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/esfri_roadmap_2016_full.pdf
- ESFRI, European Strategy Forum on Research Infrastructures, Long-Term Sustainability Working Group, (2017). *Long-Term Sustainability of Research Infrastructures*. ESFRI Scripta Volume II, 48 pp.

- http://www.esfri.eu/sites/default/files/u4/ESFRI_SCRIPTA_SINGLE_PAGE_19102017.pdf
- EU Working Group on Open Science Rewards, (2017). *Evaluation of Research Careers Fully Acknowledging Open Science Practices: Rewards, Incentives and/or Recognition for Researchers practicing Open Science*. European Commission, Directorate-General for Research and Innovation, Directorate B - OpenInnovation and Open Science, Unit B.2 – Open Science and ERA Policy, 34 pp.
https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf
- Floridi L., (2015). *Semantic Conceptions of Information*. The Stanford Encyclopedia of Philosophy.
<http://plato.stanford.edu/entries/information-semantic/>
- Force11, Data Citation Synthesis Group, (2014a). *The FAIR data principles*. Martone M. (ed.), San Diego.
<https://www.force11.org/group/fairgroup/fairprinciples>
- Force11, Data Citation Synthesis Group, (2014b). *Joint Declaration of Data Citation Principles*. Martone M. (ed.), San Diego. <https://www.force11.org/group/joint-declaration-data-citation-principles-final>
- Galimberti P., (2012). *Qualità e quantità: stato dell'arte della valutazione della ricerca nelle scienze umane in Italia*. Italian Journal of Library, Archives and Information Science (JLIS), vol.3, n.1. DOI:
<http://doi.org/10.4403/jlis.it-5617>
- GEO, Group on Earth Observations, (2015). *Data Management Principles Implementation Guidelines*.
<http://www.earthobservations.org/dswg.php>
- Hodson S., Uhlir P., Chu W., (2015). *The Value of Open Data Sharing*. Report for Group on Earth Observations, CODATA.
https://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf
- Hodson S., Jones S., Collins S., Genova F., Harrower N., Mietchen D., Petrauskaitė R. and Wittenburg P., (2018a). *FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data*. Interim report of the European Commission Expert Group on FAIR data. DOI:
<http://doi.org/10.5281/zenodo.1285290>
- Hodson S., Jones S., Collins S., Genova F., Harrower N., Laaksonen L., Mietchen D., Petrauskaitė R. and Wittenburg P., (2018b). *Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data*. 86 pp. Interim report of the European Commission Expert Group on FAIR data. DOI: <http://doi.org/10.5281/zenodo.1285272>
- INGV, Istituto Nazionale di Geofisica e Vulcanologia, (2015). *Costituzione del Gruppo di Lavoro sulla politica dei dati PoliDat*. Decreto del Presidente INGV n.223 del 11 giugno 2015.
<http://comunicazione.ingv.it/images/AmministrazioneTrasparente/decreto-Presidente-223.2015.pdf>
- INGV, Istituto Nazionale di Geofisica e Vulcanologia, (2016). *Principi della Politica dei Dati dell'INGV*. Decreto del Presidente INGV n.200 del 26 aprile 2016. <http://istituto.ingv.it/l-ingv/amministrazione-trasparente-1/contenuti-at-vari/decreti-e-delibere-anno-2016/decreto-dp-200.2016.pdf>
- INGV, Istituto Nazionale di Geofisica e Vulcanologia (2017a). *Politica di accesso aperto ai contributi della ricerca scientifica INGV*. Delibera del CdA n.408: Allegato N al Verbale n. 08/2017, del 17 luglio 2017.
http://istituto.ingv.it/images/altri_regolamenti/PolicyAccessoContributiRicercaINGV-CdA-def.pdf
- INGV, Istituto Nazionale di Geofisica e Vulcanologia (2017b). *Costituzione Gruppo di Lavoro denominato "Backup servizi e dati"*. Decreto del Direttore Generale INGV n.101 del 9 febbraio 2017.
<http://istituto.ingv.it/l-ingv/amministrazione-trasparente-1/contenuti-at-vari/2017-decreti-delibere/decreto-dg-101.2017.pdf>
- INGV, Istituto Nazionale di Geofisica e Vulcanologia (2017c). *Costituzione Gruppo di Lavoro denominato "Rete dati"*. Decreto del Direttore Generale INGV n.102 del 9 febbraio 2017.
<http://istituto.ingv.it/l-ingv/amministrazione-trasparente-1/contenuti-at-vari/2017-decreti-delibere/decreto-dg-102.2017.pdf>
- INGV, Istituto Nazionale di Geofisica e Vulcanologia (2018a). *Approvazione del Documento di attuazione della politica dei dati dell'INGV*. Delibera n.651/2018, Allegato U al Verbale n.05/2018, del 25 luglio 2018. <http://istituto.ingv.it/it/amministrazionetrasparente/disposizioni-general/atti-general/atti-amministrativi-general/decreti-e-delibere/anno-2018/delibere-del-consiglio-di-amministrazione/8740-n-651-allegato-u-verbale-n-5-2018-approvazione-del-documento-di-attuazione-della-politica-dei>

[dati-dell-ingv-pdf/file.html](#)

- INGV, Istituto Nazionale di Geofisica e Vulcanologia (2018b). Costituzione Ufficio Gestione Dati "UGD". Decreto del Presidente INGV n.349/2018, del 9 novembre 2018.
<http://istituto.ingv.it/it/amministrazionetrasparente/disposizioni-general/atti-general/atti-amministrativi-general/decreti-e-delibere/anno-2018/decreti-del-presidente-e-del-direttore-generale-4/9100-n-349-dp-costituzione-ufficio-gestione-dati-ugd-pdf/file.html>
- Kenna R., Mryglod O., Berche B., (2017). *A scientists' view of scientometrics: not everything that counts can be counted*. Condensed Matter Physics, vol.20, n.1. DOI: <http://10.5488/CMP.20.13803>
- Kenney R.A. and McGovern N.Y., (2003). *The Five Organizational Stages of Digital Preservation*. In: Digital Libraries: A Vision for the Twenty-first Century, a festschrift to honor Wendy Lougee. University of Michigan Scholarly Monograph Series. DOI: <http://doi.org/10.3998/spobooks.bbv9812.0001.001>
- Landry B.C., Mathis B.A., Meara N.M., Rush J.E., Young C.E., (1973). *Definition of some basic terms in computer and information science*. Journal of the American Society for Information Science, vol.24, n.5. DOI: <http://doi.org/10.1002/asi.4630240504>
- Lavoie B.F., (2014). *The Open Archival Information System Reference Model: Introductory Guide (2nd edition)*. Digital Preservation Coalition, Technology Watch Series Report 14-02 October 2014, 37 pp. DOI: <http://www.dpconline.org/docman/technology-watch-reports/1359-dpctw14-02/file>
- Leonelli S., (2017). *Mutual Learning Exercise Open Science - Altmetrics and Rewards Incentives and Rewards to engage in Open Science Activities*. European Commission, Directorate-General for Research and Innovation Directorate A, Policy Development and Coordination, Thematic Report No 3. https://ore.exeter.ac.uk/repository/bitstream/handle/10871/31348/ML-OS-Report-3_Final.pdf
- LERU, League of European Research Universities, Research Data Working Group, (2016). *LERU Roadmap for Research Data*.
http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf
- Mayernik M.S., Di Lauro T., Duerr R., Metsger E., Thessen A. E., Choudhury G.S. (2013). *Data conservancy provenance, context, and lineage services: key components for data preservation and curation*. Data Science Journal, vol.12. DOI: <http://doi.org/10.2481/dsj.12-039>
- McMurry J.A., Juty N., Blomberg N., Burdett T., Conlin T., Conte N. et al., (2017). *Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data*. PLoS Biology 15(6): e2001414. DOI: <https://doi.org/10.1371/journal.pbio.2001414>
- MIUR, gruppo di lavoro sui Big Data, (2016). *Big Data @MIUR*. Rapporto tecnico, Ministero dell'Istruzione dell'Università e della Ricerca. <http://www.istruzione.it/allegati/2016/bigdata.pdf>
- Moreau L., (2010). *The Foundations for Provenance on the Web*. Foundations and Trends in Web Science, vol. 2, Nos. 2-3 (2010) 99-241. DOI: <http://doi.org/10.1561/1800000010>
- Nielsen M., (2011). *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press, 280 pp. <http://press.princeton.edu/titles/9517.html>
- OECD, Organisation for Economic Co-operation and Development, (2015). *Making Open Science a reality*. DOI: <http://doi.org/10.1787/5jrs2f963zs1-en>
- Orth A., Pontika N., Ball D., (2017). *FOSTER's Open Science Training Tools and Best Practices*. IOS Press, pp.135-141. DOI: <http://doi.org/10.3233/978-1-61499-649-1-135>
- Pappalardo M., Beranzoli L., Favali P., Mairi Best, Piera J., Dañobeitia J.J., Huber R., Waldmann C., Rolin J.F., Petihakis G., (2015). *Data Management Plan*. H2020 EMSODEV, Work Package 6 "Data management platform", Deliverable 6.1, 36pp.
<https://emdesk.eu/shared/56a87775876a9-16f059520226aeb16659f00b93d0522a>
- Paskin N., (2010). *Digital Object Identifier (DOI) System*. Encyclopedia of Library and Information Sciences, Third Edition, pp.1586-1592. DOI: <http://doi.org/10.1081/E-ELIS3-120044418>
- Potočnik J., (2007). *The EU's Fifth Freedom: creating free movement of knowledge*. Informal Competitiveness Council, Wuerzburg (Germany), 26 April 2007. http://europa.eu/rapid/press-release_SPEECH-07-257_en.pdf
- RAND Europe (2017). *Open Science Monitoring. Methodological Note*. European Commission Directorate-

- General for Research and Innovation and received funding under specific contract 2015/RTD/A6/SC/PP-02741-2015.
- Regione Emilia-Romagna, (2016). *Emilia Romagna Big Data Community, From Volume to Value, 2nd edition*. Regione Emilia-Romagna, rapporto di programmazione 2014-2020, 48pp. http://formazioneelavoro.regione.emilia-romagna.it/alta-formazione-ricerca/allegati/From_volume_to_value.pdf/at_download/file/Big-data-second_edition.pdf
- RDA, (2015). *Data Foundation and Terminology guidelines*. Research Data Alliance, Data Foundation and Terminology Work Group Products. DOI: <http://doi.org/10.15497/06825049-8CA4-40BD-BCAF-DE9F0EA2FADF>
- RDA-CODATA, Legal Interoperability Interest Group, (2016). *Legal Interoperability of Research Data: Principles and Implementation Guidelines, v1.0*. Research Data Alliance. DOI: <http://doi.org/10.5281/zenodo.162241>
- Richards K., White R., Nicolson N., Pyle R., (2011). *A Beginner's Guide to Persistent Identifiers, version 1.0*. Global Biodiversity Information Facility (GBIF), Copenhagen, 33 pp. http://links.gbif.org/persistent_identifiers_guide_en_v1.pdf
- Ross-Hellauer T., (2017). *What is open peer review? A systematic review* [version 2; referees: 1 approved, 3 approved with reservations]. F1000Research 2017, 6:588. DOI: <http://doi.org/10.12688/f1000research.11369.2>
- Shadbolt N., Hall W. and Berners-Lee T., (2006). *The Semantic Web Revisited*. IEEE Intelligent Systems Journal, 21(3):96-101. DOI: <http://doi.org/10.1109/MIS.2006.62>
- Schöpfela J., Prosta H., Rebouillatb V., (2017). *Research Data in Current Research Information Systems*. Procedia Computer Science, vol.106. DOI: <http://doi.org/10.1016/j.procs.2017.03.030>
- Science|Business, (2017). *Governing the Open Science Cloud*. Science|Business Network's Cloud Consultation Group.
- Science|Business, (2018). *The European Open Science Cloud: Who pays for what?* Science|Business Network's Cloud Consultation Group.
- Science Europe, Working Group on Research Policy and Programme Evaluation, (2016). *Position Statement on Research Information Systems*. Science Europe, Brussels. http://www.scienceeurope.org/wp-content/uploads/2016/11/SE_PositionStatement_RIS_WEB.pdf
- Simghan Y.L., Plale B., Gannon D., (2005). *A Survey of Data Provenance Techniques*. Computer Science Department, Indiana University, Technical Report IUB-CS-TR618. <https://www.cs.indiana.edu/ftp/techreports/TR618.pdf>
- Starr J. et al., (2014). *DataCite metadata schema for the publication and citation of research data, metadata schema, version 4.0*. DataCite e.V., Hannover, Germany. DOI: <http://doi.org/10.5438/0012>
- Starr J. and Gastl A., (2011). *isCitedBy: a metadata scheme for DataCite*. D-Lib Magazine, 17(1/2). DOI: <http://doi.org/10.1045/january2011-starr>
- Stewart Lowndes J.S., Best B.D., Scarborough C., Afflerbach J.C., Frazier M.R., O'Hara C.C., Jiang N., Halpern B.S., (2017). *Our path to better science in less time using open data science tools*. Nature Ecology & Evolution, Vol.1, 7pp. DOI: <http://doi.org/10.1038/s41559-017-0160>
- STOA, Science and Technology Options Assessment, (2014). *Measuring scientific performance for improved policy making*. Science and Technology Options Assessment, European Parliamentary Research Service, 24pp. [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN_ET\(2014\)527383\(SUM01\)_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN_ET(2014)527383(SUM01)_EN.pdf)
- Thanos C., (2017). *Research Data Reusability: Conceptual Foundations, Barriers and Enabling Technologies*. Publications 2017, 5(1), 2, 19pp. DOI: <https://doi.org/10.3390/publications5010002>
- UNESCO, (2015). *Research Evaluation Metrics*. United Nations Educational, Scientific and Cultural Organization, UNESCO, Paris. <http://unesdoc.unesco.org/images/0023/002322/232210E.pdf>
- Vezzoso S., (2008). *Open Access: scelte istituzionali e ruolo del diritto d'autore*. In: Atti del Convegno "Pubblicazioni scientifiche, diritti d'autore e Open Access", 20 giugno 2008, Università degli Studi di Trento, Quaderni del Dipartimento di Scienze Giuridiche, vol.79, pp.81-95.

http://eprints.biblio.unitn.it/archive/00001589/02/unico_2_versione_12_5_2009.pdf

Whyte A. and Tedds J., (2011). *Making the Case for Research Data Management*. Digital Curation Centre (DCC) Briefing Papers, Edinburgh, 8pp. <http://www.dcc.ac.uk/resources/briefing-papers>

Wiley (2014). *Researcher data sharing insights*.

<https://hub.wiley.com/community/exchanges/discover/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont?referrer=exchanges>

Wilkinson M.D. et al., (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, vol.3, Macmillan Publishers Ltd. DOI: <http://doi.org/10.1038/sdata.2016.18>

Wilsdon J., Allen L., Belfiore E., Johnson B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Higher Education Funding Council for England. DOI: <http://doi.org/10.13140/RG.2.1.4929.1363>

Sitografia

Tutti i link presentati sono validi alla data di redazione di questa bibliografia, maggio 2018.

AgID, Agenzia per l'Italia Digitale, (2017b). <http://www.agid.gov.it/>

Bratislava Declaration of Young Researchers, (2016).

<https://web.archive.org/web/20170716120203/http://www.eu2016.sk/data/documents/bratislava-declaration-of-young-researchers-final.pdf>

CINECA, Consorzio Interuniversitario del Nord-Est per il Calcolo Automatico, (2017).

<https://www.cineca.it/>

CINECA, (2018). "Un link ultra veloce tra il centro LHC Tier1 del CNAF-INFN è Cineca"

<https://www.cineca.it/it/news/un-link-ultra-veloce-tra-il-centro-lhc-tier1-del-cnaf-infn-e-cineca>

CNAF-INFN, Centro Nazionale Analisi Fotogrammi dell'Istituto Nazionale di Fisica Nucleare, (2017).

<https://www.cnaf.infn.it/>

Comunicazione CE COM 6 (2000).

<https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:52000IP0131>

Comunicazione CE COM 392 (2012).

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012AE2075>

Comunicazione CE COM 192 final (2015).

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015AE3604>

Comunicazione CE COM 0178 final (2016). <https://publications.europa.eu/en/publication-detail/-/publication/47cc5480-0612-11e6-b713-01aa75ed71a1>

COM 09 final, (2017). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2017:9:FIN>

Comunicazione CE COM 8 final (2018).

<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018AE0354>

Decision (EU) 2015/2240 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015D2240>

Direttiva CE 2012/417/EU <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012H0417>

Direttiva CE 2013/37/EU <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32013L0037>

EC Staff Working Group SWD, 83 final, (2018).

https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

EGDF, EOSCpilot Governance Development Forum, (2017). <https://eoscpilot.eu/about/governance>

EPOS <https://www.epos-ip.org>

European Qualifications Framework, EQF, COM 0383 final (2016).

<https://publications.europa.eu/en/publication-detail/-/publication/e3a30a2e-434b-11e6-9c64-01aa75ed71a1>

ER, (2016a). "Ricerca e innovazione, l'Emilia-Romagna investe nella sua Big Data Community", notizia dal portale della Regione Emilia-Romagna, 8 febbraio 2016.

- <http://www.regione.emilia-romagna.it/notizie/2016/febbraio/ricerca-e-innovazione-lemilia-romagna-investe-nella-sua-big-data-community>
- ER, (2016b). “*Big Data, nasce la Community Emilia Romagna*”, notizia dal portale della Regione Emilia-Romagna, 9 febbraio 2016.
<http://www.regione.emilia-romagna.it/fesr/notizie/2016/febbraio/bigdata-nasce-la-community-emilia-romagna>
- EOSC, European Open Science Cloud, (2017). progetto H2020. <https://eoscpilot.eu/>
http://cordis.europa.eu/project/rcn/207500_en.html
- eduGAIN, EDUcation Global Authentication INfrastructure, (2017). <https://technical.edugain.org/>
- EGI, European Grid Infrastructure, (2017). <https://www.egi.eu>
- EGI-Engage, (2016). “Engaging the EGI Community towards an Open Science Commons, pagina ufficiale”, sezione delle pagine Wiki di EGI. <https://wiki.egi.eu/wiki/EGI-Engage>
- e-IRG, (2017). e-IRG, e-Infrastructures Reflection Group, progetto H2020 “e-Infrastructure Reflection Group Support Programme 5” (e-IRGSP5). <http://e-irg.eu/>
- ELIXIR, (2017). ELIXIR, distributed infrastructure for biological data. <https://www.elixir-europe.org/>
- EUDAT, (2017). EUDAT, pagina ufficiale, ultimo accesso 29 Maggio 2017. <https://eudat.eu/>
- Forum di Belmont, (2017). Forum di Belmont, pagina ufficiale. <https://www.belmontforum.org/>
- GARR News 14, (2016). “*Identità digitale, cosa cambia con SPID?*”, notizia da GARR News N°14, Giugno 2016. <http://www.garrnews.it/index.php/ricerche/427>
- GARR, (2017). GARR, Gruppo per l’Armonizzazione delle Reti della Ricerca. <http://www.garr.it/it/>
- GÉANT, (2017). Associazione GÉANT. <https://www.geant.org/>
- ISTAT, (2017). Elenco delle unità istituzionali appartenenti al settore delle Amministrazioni Pubbliche. <https://www.istat.it/it/archivio/190748>
- IOSSG, (2016). Italian Open Science Support Group. <https://sites.google.com/view/iossg/>
- Open Data Barometer, (2017). The Open Data Barometer, <http://opendatabarometer.org/>
- Open Definition <https://opendefinition.org/>
- Open Definition Open Definition principles that define “openness” in relation to data and content. <http://opendefinition.org/>
- PRACE, (2017). PRACE, Partnership for Advanced Computing in Europe. <http://www.prace-ri.eu>
- Raccomandazione CE 2005/251/EC <https://eur-lex.europa.eu/eli/reco/2005/251/oj>
- RIPE, (2017). RIPE, Réseaux IP Européens. <https://www.ripe.net/>
- Sole24Ore, (2016). “*Bologna crea il polo nazionale dei big data: il 70% della capacità di supercalcolo corre lungo la via Emilia*”, notizia scritta da I. Vesentini, Sole24Ore, sezione Impresa & Territorio, 8 Febbraio 2016. <http://www.ilsole24ore.com/art/impresa-e-territori/2016-02-08/-bologna-crea-polo-nazionale-big-data-70percento-capacita-supercalcolo-corre-la-via-emilia-161253.shtml>
- SWD (2017) 323 final <https://ec.europa.eu/transparency/regdoc/rep/10102/2017/EN/SWD-2017-323-F1-EN-MAIN-PART-1.PDF>
- TERENA, (2017). TERENA, pagina ufficiale, ultimo accesso 29 Maggio 2017. <https://www.terena.org/>
- TOP500, (2016). TOP500 Supercomputer Sites. <https://www.top500.org/list/2016/11/>
- Wikipedia italiano, (2017b). GARR, da Wikipedia, l’enciclopedia libera, versione italiana, ultima modifica il 9 feb 2017. <https://it.wikipedia.org/wiki/GARR>
- Wikipedia italiano, (2017c). Pubblica Amministrazione italiana, da Wikipedia, l’enciclopedia libera, versione italiana, ultima modifica 26 Aprile 2017. https://it.wikipedia.org/wiki/Pubblica_amministrazione_italiana
- Wikipedia inglese, (2017a). DANTE, da Wikipedia, l’enciclopedia libera, versione italiana, ultima modifica il 21 Ottobre 2016. <https://en.wikipedia.org/wiki/DANTE>
- Wikipedia inglese, (2017c). EGI, da Wikipedia, l’enciclopedia libera, versione inglese, ultima modifica il 20 Marzo 2017. https://en.wikipedia.org/wiki/European_Grid_Infrastructure
- Wikipedia inglese, (2017d). GARR, da Wikipedia, l’enciclopedia libera, versione inglese, ultima modifica il 14 April 2017. <https://en.wikipedia.org/wiki/GARR>

Wikipedia inglese, (2017e). GÉANT, da Wikipedia, l'enciclopedia libera, versione inglese, ultima modifica il 29 Maggio 2017. <https://en.wikipedia.org/wiki/GÉANT>

Wikipedia inglese, (2017f). PRACE, da Wikipedia, l'enciclopedia libera, versione inglese, ultima modifica il 4 Gennaio 2017. https://en.wikipedia.org/wiki/Partnership_for_Advanced_Computing_in_Europe

WLHC, (2017). WLHC, Worldwide LHC Computing Grid. <http://wlcg.web.cern.ch/>

QUADERNI di GEOFISICA

ISSN 1590-2595

<http://istituto.ingv.it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/>

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941

<http://istituto.ingv.it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/>

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651

http://istituto.ingv.it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

Coordinamento editoriale e impaginazione

Francesca DI STEFANO, Rossella CELI
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico e impaginazione

Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

©2018
Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

