

Esperienze di gestione degli Open Data negli Enti pubblici di Ricerca

Mario Locati (INGV), Alessandra Giorgetti (OGS), Angela Saraò (OGS) Roberta Vigni (ISPRA)

Con il contributo di

Elisabetta Poltronieri (ISS), Paola Carrabba (ENEA), Monica Sala (ENEA), Daniela Palma (ENEA),
Dario Pileri (INRIM), Emanuela Secinaro (INRIM), Carlo Cipolloni (ISPRA), Simona Olivadese (ISPRA),
Riccardo Scano (CREA), Vincenzo Patruno (ISTAT), Marco Biagetti (INAPP), Cristian Mezzetti (CNR),
Stefano Cozzini (AREA SCIENCE PARK), Stefano Bianco (INFN), Riccardo Smareglia (INAF).



CoPER Gruppo Open Science

CONSULTA DEI PRESIDENTI
DEGLI ENTI DI RICERCA

1. Open Science e Open Data
2. Governare la condivisione di Open Data
3. Panoramica sulle attuali politiche Open Data
4. Sfide nell'attuazione delle politiche Open Data
5. Conclusioni

Il paradigma dell'Open Science come guida alla gestione dei dati

Il paradigma dell'Open Science mira a una condivisione piena, libera, efficiente, efficace, affidabile e duratura dei prodotti della ricerca scientifica in un'ottica multidisciplinare secondo il principio *as Open as possible, as Closed as necessary*

I **principi FAIR** caratterizzano gli Open (Research) Data
Findable, Accessible, Interoperable, Reusable

I principi FAIR vanno intesi nella loro **accezione informatica**:
chi compie un'azione potrebbe essere un software in modo autonomo, senza che l'azione sia necessariamente innescata da una persona.

Principi FAIR, condizione per la condivisione dei dati

Findable

Esistono molteplici standard di metadati, solo parzialmente sovrapponibili
Più i metadati sono ricchi di informazioni, più i motori di ricerca sono precisi
Necessario usare identificativi univoci, gli stessi dati sono accessibili da più canali

Accessible

I dati devono poter essere recuperabili direttamente dagli utilizzatori (umani e non)
La privacy degli utilizzatori va rispettata, ridurre al minimo il tracciamento

Interoperable

Esistono molteplici standard per codificare i dati e vari standard di servizi web
Oltre a condividere file è utile adottare servizi web per i dati dinamici

Reusable

L'utente va informato sui termini di utilizzo dei dati tramite una licenza standard
La licenza CC BY 4.0 è consigliata e vincola l'utente a citare la fonte

Principi FAIR, condizione per la condivisione dei dati

Findable

Esistono molteplici standard di metadati, solo parzialmente sovrapponibili
Più i metadati sono ricchi di informazioni, più i motori di ricerca sono precisi
Necessario usare identificativi univoci, gli stessi dati sono accessibili da più canali

Accessible

I dati devono poter essere recuperabili direttamente dagli utilizzatori (umani e non)
La privacy degli utilizzatori va rispettata, ridurre al minimo il tracciamento

Interoperable

Esistono molteplici standard per codificare i dati e vari standard di servizi web
Oltre a condividere file è utile adottare servizi web per i dati dinamici

Reusable

L'utente va informato sui termini di utilizzo dei dati tramite una licenza standard
La licenza CC BY 4.0 è consigliata e vincola l'utente a citare la fonte

I moderni software che gestiscono gli **archivi di dati** svolgono tutte le funzioni in un'unica soluzione.
In caso di archivi distinti può essere utile dotarsi di un **catalogo di metadati**.

Il legislatore detta modi e tempi per la condivisione dei dati

Il 2019 è uno spartiacque nello scenario legislativo riguardante i dati della ricerca.

- Prima** Direttiva europea per il riutilizzo dell'informazione del settore pubblico
Public Sector Information (PSI) Directive
Ampio margine di libertà di gestione per gli enti titolari dei dati
- Dopo** Direttiva 2019/1024 recepita in Italia nel novembre del 2021
Ridotto margine di libertà di gestione per gli enti titolari dei dati

Il legislatore detta modi e tempi per la condivisione dei dati

Il 2019 è uno spartiacque nello scenario legislativo riguardante i dati della ricerca.

Prima Direttiva europea per il riutilizzo dell'informazione del settore pubblico
Public Sector Information (PSI) Directive
Ampio margine di libertà di gestione per gli enti titolari dei dati

Dopo Direttiva 2019/1024 recepita in Italia nel novembre del 2021
Ridotto margine di libertà di gestione per gli enti titolari dei dati

Forte spinta europea (e nazionale) a creare il **Digital Single Market** che avrà ricadute sulle attività economiche, deve essere possibile sfruttare liberamente i dati della ricerca.

L'**European Open Science Cloud (EOSC) Market Place** è un primo e concreto esempio della direzione intrapresa esplicitamente dalla **European Research Area (ERA)**.

Dotarsi di una politica dedicata agli Open Data è utile e necessario

1. Il legislatore fornisce una serie di **indicazioni normative** di carattere generale.

Dotarsi di una politica dedicata agli Open Data è utile e necessario

1. Il legislatore fornisce una serie di **indicazioni normative** di carattere generale.
2. Gli enti adattano le indicazioni alla propria struttura dotandosi di una **Politica dei Dati**.

Dotarsi di una politica dedicata agli Open Data è utile e necessario

1. Il legislatore fornisce una serie di **indicazioni normative** di carattere generale.
2. Gli enti adattano le indicazioni alla propria struttura dotandosi di una **Politica dei Dati**.
3. Si individua un **soggetto istituzionale preposto** all'attuazione della Politica dei Dati.

Dotarsi di una politica dedicata agli Open Data è utile e necessario

1. Il legislatore fornisce una serie di **indicazioni normative** di carattere generale.
2. Gli enti adattano le indicazioni alla propria struttura dotandosi di una **Politica dei Dati**.
3. Si individua un **soggetto istituzionale preposto** all'attuazione della Politica dei Dati.
4. Si creano (o adattano) **strumenti gestionali** funzionali alla gestione dei dati
 - per **utenti** interni ed esterni (es. archivi di dati);
 - per **finalità** di ricerca e amministrative (es. cataloghi di metadati, DMP);
 - conformi a standard tecnologici internazionali.



Dotarsi di una politica dedicata agli Open Data è utile e necessario

1. Il legislatore fornisce una serie di **indicazioni normative** di carattere generale.
2. Gli enti adattano le indicazioni alla propria struttura dotandosi di una **Politica dei Dati**.
3. Si individua un **soggetto istituzionale preposto** all'attuazione della Politica dei Dati.
4. Si creano (o adattano) **strumenti gestionali** funzionali alla gestione dei dati
 - per **utenti** interni ed esterni (es. archivi di dati);
 - per **finalità** di ricerca e amministrative (es. cataloghi di metadati, DMP);
 - conformi a standard tecnologici internazionali.
5. Si adattano i **processi interni** per adeguarli alla Politica dei Dati, ad esempio:
 - specifici articoli dedicati ai dati negli **accordi di collaborazione**;
 - **pianificazione triennale** delle attività;
 - **monitoraggio** e auto-valutazione delle attività e degli obiettivi.



Alcune tipologie di dati richiedono una gestione specifica

Oltre alla condivisione per finalità di ricerca, alcuni dati devono soddisfare **vincoli giuridici e tecnici** e devono adottare strumenti, codifiche e **canali istituzionali già formalizzati a livello governativo**.

Il sottoinsieme di **“dati di elevato valore”** è definito nella Direttiva Europea 2019/1024 e suddiviso nelle seguenti categorie tematiche:

- Dati geospaziali
- Dati relativi all'osservazione della terra e all'ambiente
- Dati meteorologici
- Dati statistici
- Dati relativi alle imprese e alla proprietà delle imprese
- Dati relativi alla mobilità



Strumenti utili per la definizione di una politica dei dati

Le linee guida dell'Agenzia per l'Italia Digitale (AgID) sono vincolanti per la PA:

- aiutano a districarsi nel complesso ambito normativo e gestionale
- forniscono utili suggerimenti sia su aspetti organizzativi sia tecnologici
- regole valide per tutta la Pubblica Amministrazione, richiedono un adattamento

Strumenti utili per la definizione di una politica dei dati

Le linee guida dell'Agenzia per l'Italia Digitale (AgID) sono vincolanti per la PA:

- aiutano a districarsi nel complesso ambito normativo e gestionale
- forniscono utili suggerimenti sia su aspetti organizzativi sia tecnologici
- regole valide per tutta la Pubblica Amministrazione, richiedono un adattamento

E' necessario un **percorso formativo per dirigenti, ricercatori e tecnici**

perché essere conformi ai principi FAIR ha ricadute sulla pianificazione (es. DMP), sul modo di fare ricerca e su aspetti amministrativi e legali.

Strumenti utili per la definizione di una politica dei dati

Le linee guida dell'Agenzia per l'Italia Digitale (AgID) sono vincolanti per la PA:

- aiutano a districarsi nel complesso ambito normativo e gestionale
- forniscono utili suggerimenti sia su aspetti organizzativi sia tecnologici
- regole valide per tutta la Pubblica Amministrazione, richiedono un adattamento

E' necessario un **percorso formativo per dirigenti, ricercatori e tecnici**

perché essere conformi ai principi FAIR ha ricadute sulla pianificazione (es. DMP), sul modo di fare ricerca e su aspetti amministrativi e legali.

Per governare l'Open Science, è utile **collegare tra loro tutti i tipi di prodotti della ricerca**, le persone e le organizzazioni coinvolte e i progetti finanziatori.

Strumenti utili per la definizione di una politica dei dati

Le linee guida dell'Agenzia per l'Italia Digitale (AgID) sono vincolanti per la PA:

- aiutano a districarsi nel complesso ambito normativo e gestionale
- forniscono utili suggerimenti sia su aspetti organizzativi sia tecnologici
- regole valide per tutta la Pubblica Amministrazione, richiedono un adattamento

E' necessario un **percorso formativo per dirigenti, ricercatori e tecnici**

perché essere conformi ai principi FAIR ha ricadute sulla pianificazione (es. DMP), sul modo di fare ricerca e su aspetti amministrativi e legali.

Per governare l'Open Science, è utile **collegare tra loro tutti i tipi di prodotti della ricerca**, le persone e le organizzazioni coinvolte e i progetti finanziatori.

Attenzione: **realizzare un archivio di dati non è solo una questione tecnologica**, richiede piano di gestione e sostenibilità (personale e fondi) su base almeno decennale.

Differenze tra enti vigilati e non vigilati dal MUR

La conoscenza reciproca nel Gruppo Open Science della CoPER è all'inizio.

Esistono sicuramente delle differenze tra enti vigilati e non vigilati dal MUR.

Alcuni caratteri specifici:

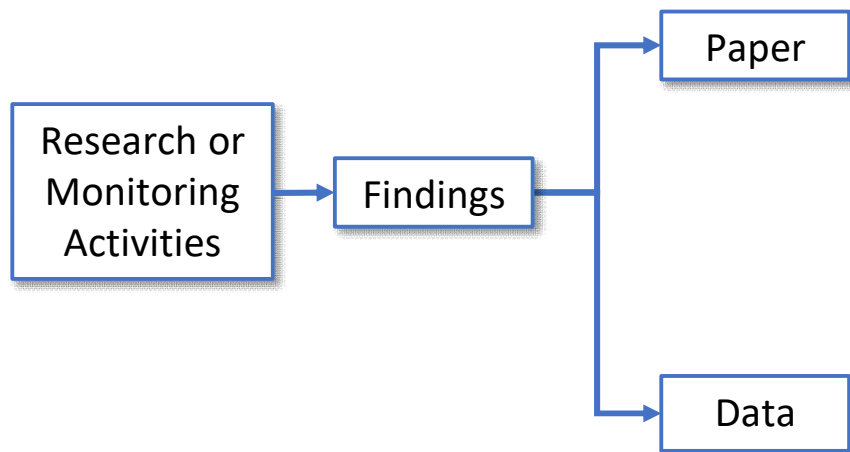
- partecipazione al **sistema di valutazione ANVUR**
(non tutti devono sottostare alla regola “publish or perish”)
- **priorità** e meccanismi di funzionamento della Ricerca
- sistemi e interlocutori legati ai **finanziamenti**

La finalità di condividere Open Data è un punto in comune.

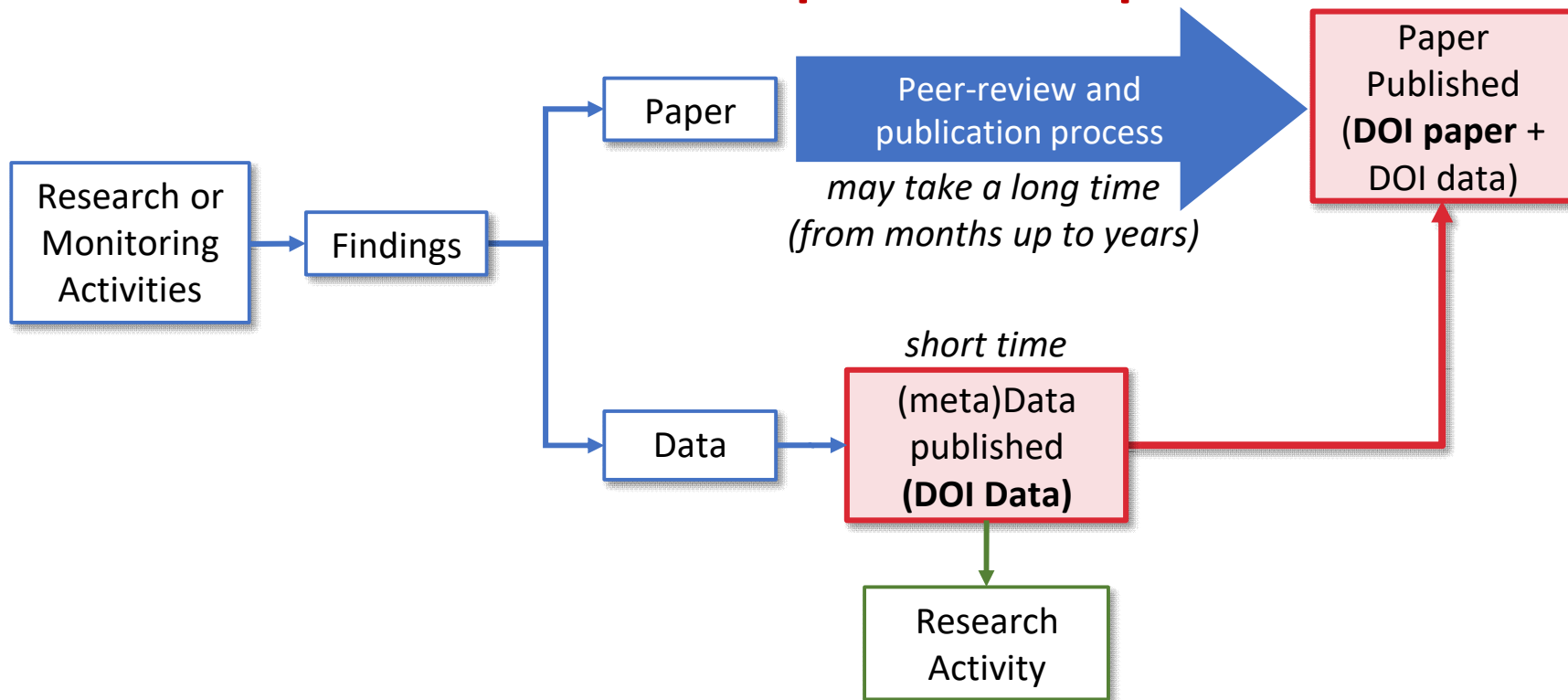
Sondaggio sulla gestione istituzionale degli Open Data *(preliminare)*

	Politica dati	Catalogo (meta)dati	Archivio dati	Soggetto istituzionale preposto
CREA Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria	SI	-	parziale (archivi distinti)	Ufficio sistemi informativi
ENEA Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile	-	-	parziale (archivi distinti)	-
INAF Istituto nazionale di astrofisica	parziale (solo per dati raw)	SI (tramite Virtual Observatory)	SI (tramite Virtual Observatory)	Centro Archivi Nazionale (parzialmente)
INAPP Istituto nazionale per l'analisi delle politiche pubbliche	-	-	parziale (solo dati aggregati)	Servizio statistico
INFN Istituto nazionale di fisica nucleare	in approvazione	-	misto (misto dati/pubblicazioni)	Gruppo di lavoro Open Science e commissione calcolo e reti
INGV Istituto nazionale di geofisica e vulcanologia	SI	SI	parziale (archivi distinti)	Ufficio gestione dati
INRIM Istituto nazionale di ricerca metrologica	SI	-	SI (community su ZENODO)	Gruppo di lavoro Open Science
ISPRA Istituto superiore per la protezione e la ricerca ambientale	in approvazione	SI	parziale (archivi distinti)	SINA - Sistema Informativo Nazionale Ambientale
ISS Istituto superiore di sanità	SI	-	misto (misto dati/pubblicazioni)	Servizio Comunicazione Scientifica, Servizio Conoscenza, Servizio di Coordinamento e supporto alla Ricerca, Servizio Informatica
ISTAT Istituto nazionale di statistica	-	-	parziale (archivi distinti)	DCCI - Direzione centrale per la comunicazione, informazione e servizi ai cittadini e agli utenti
OGS Istituto nazionale di oceanografia e di geofisica sperimentale	parziale (solo per dati oceanografici)	-	parziale (archivi distinti)	Comitato per la proprietà intellettuale e la scienza aperta

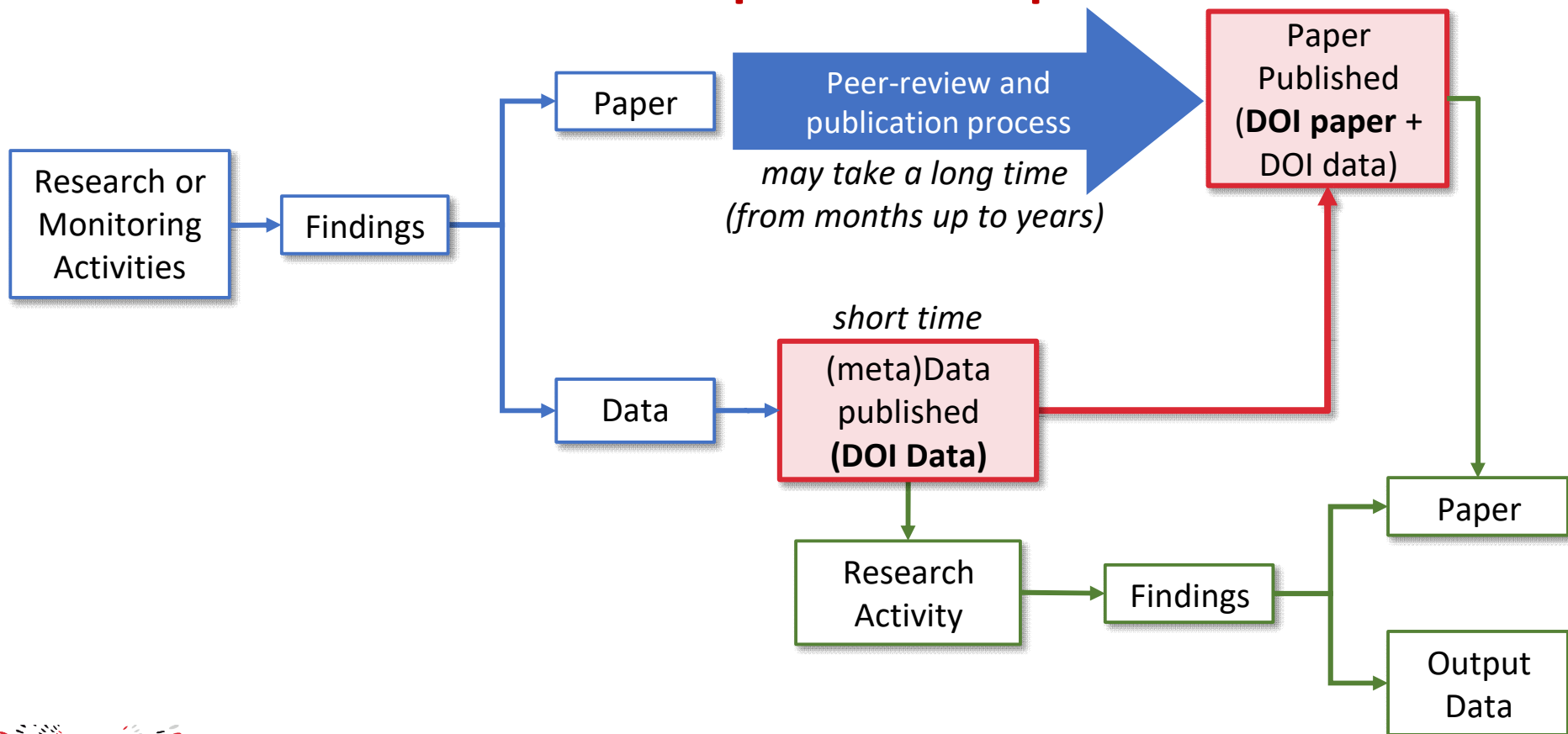
Canali di condivisione distinti per i diversi prodotti della ricerca



Canali di condivisione distinti per i diversi prodotti della ricerca



Canali di condivisione distinti per i diversi prodotti della ricerca



Lo scoglio della mancata valutazione dei dati

La condivisione dei dati in modo FAIR richiede **nuove competenze** e professionalità che sono oggi **scarsamente disponibili o assenti**.

Lo scoglio della mancata valutazione dei dati

La condivisione dei dati in modo FAIR richiede **nuove competenze** e professionalità che sono oggi **scarsamente disponibili o assenti**.

I ricercatori devono sopperire a questa assenza di figure specifiche e dedicate ma le attività necessarie **non sono valutate** ai fini dell'avanzamento di carriera.

Lo scoglio della mancata valutazione dei dati

La condivisione dei dati in modo FAIR richiede **nuove competenze** e professionalità che sono oggi **scarsamente disponibili o assenti**.

I ricercatori devono sopperire a questa assenza di figure specifiche e dedicate ma le attività necessarie **non sono valutate** ai fini dell'avanzamento di carriera.

Condividere i dati conformemente ai principi FAIR è **oneroso**

- per i **ricercatori** richiede dedizione, tempo e nuova formazione
- per le **infrastrutture** richiede l'adozione di nuovi standard tecnologici
- per l'**ente** richiede la sostenibilità nel lungo periodo, sia in termini di risorse umane che finanziaria

Mancano sistemi di valutazione dei dati

Non c'è un processo standardizzato di **peer-review** per la pubblicazione dei dati (*vedi Zenodo*).

Mancano sistemi di valutazione dei dati

Non c'è un processo standardizzato di **peer-review per la pubblicazione dei dati** (*vedi Zenodo*).

Non sono ancora disponibili metodi consolidati e standardizzati di valutazione dei dati.

Mancano sistemi di valutazione dei dati

Non c'è un processo standardizzato di **peer-review per la pubblicazione dei dati** (*vedi Zenodo*).

Non sono ancora disponibili metodi consolidati e standardizzati di valutazione dei dati.

Senza un sistema di metriche oggettive e facilmente misurabili (**Key Performance Indicators**) è arduo pretendere l'adozione di un sistema di valutazione dei dati.

Mancano sistemi di valutazione dei dati

Non c'è un processo standardizzato di **peer-review per la pubblicazione dei dati** (*vedi Zenodo*).

Non sono ancora disponibili metodi consolidati e standardizzati di valutazione dei dati.

Senza un sistema di metriche oggettive e facilmente misurabili (**Key Performance Indicators**) è arduo pretendere l'adozione di un sistema di valutazione dei dati.

Vista la crescente importanza di EOSC, è sempre più necessario poter misurare aspetti come:

- **il grado di FAIRness** (*valutazione ex-ante*)
- **l'impatto verso l'utente** (*valutazione ex-post*)
 - Quanti e quali utenti ne ricavano beneficio? Che tipo e grado di beneficio?
 - Le informazioni per un uso consapevole dei dati erano sufficienti e chiare?
 - L'uso dei dati è avvenuto coerentemente alla loro natura e nei limiti della loro attendibilità?
 - E' stato complicato usare i dati nel proprio workflow?
 - Ci sono state ricadute negative a causa della disponibilità dei dati?



Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Il legislatore (EU/ITA) sta regolamentando modalità e tempi della condivisione dei dati.

Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Il legislatore (EU/ITA) sta regolamentando modalità e tempi della condivisione dei dati.

Soddisfare le condizioni per condividere Open Data esige **molteplici azioni su piani diversi.**

Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Il legislatore (EU/ITA) sta regolamentando modalità e tempi della condivisione dei dati.

Soddisfare le condizioni per condividere Open Data esige **molteplici azioni su piani diversi**.

L'uso corretto dei dati è ancora poco diffuso (es. citazione dei DOI) nei processi della ricerca.

Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Il legislatore (EU/ITA) sta regolamentando modalità e tempi della condivisione dei dati.

Soddisfare le condizioni per condividere Open Data esige **molteplici azioni su piani diversi**.

L'uso corretto dei dati è ancora poco diffuso (es. citazione dei DOI) nei processi della ricerca.

Governare gli Open Data a livello istituzionale richiede **nuove professionalità e fondi dedicati**.

Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Il legislatore (EU/ITA) sta regolamentando modalità e tempi della condivisione dei dati.

Soddisfare le condizioni per condividere Open Data esige **molteplici azioni su piani diversi**.

L'uso corretto dei dati è ancora poco diffuso (es. citazione dei DOI) nei processi della ricerca.

Governare gli Open Data a livello istituzionale richiede **nuove professionalità e fondi dedicati**.

Il panorama degli enti di ricerca è variegato, alcuni sono più avanti su alcuni aspetti.

Conclusioni

EOSC sta creando le premesse per rendere i dati scientifici sempre più aperti e utilizzabili.

Il legislatore (EU/ITA) sta regolamentando modalità e tempi della condivisione dei dati.

Soddisfare le condizioni per condividere Open Data esige **molteplici azioni su piani diversi**.

L'uso corretto dei dati è ancora poco diffuso (es. citazione dei DOI) nei processi della ricerca.

Governare gli Open Data a livello istituzionale richiede **nuove professionalità e fondi dedicati**.

Il panorama degli enti di ricerca è variegato, alcuni sono più avanti su alcuni aspetti.

In questa fase è fondamentale condividere le esperienze:

serve uno sforzo per documentare il percorso di gestione degli Open Data in ciascun ente.

Grazie per l'attenzione!

Domande?

Per il GdL CoPER Open Science conper.openscience@lists.infn.it

mario.locati@ingv.it

agiorgetti@ogs.it

asarao@ogs.it

roberta.vigni@isprambiente.it

