



Settore Servizi per la ricerca e diffusione della conoscenza scientifica

U.O. Sviluppo performance dell'archivio istituzionale della ricerca e servi: ib SISTEMA BIBLIO

Questionario sulle abitudini dei ricercatori in materia di deposito dei dati della ricerca – giugno 2022

Analisi dei risultati

A cura del Settore Servizi per la ricerca e diffusione della conoscenza scientifica

Maria Stella Castiglia¹, Maria Francesca D'Asaro², Maria Concetta Stella³ e Vittorio Tranchina⁴

13 gennaio 2023

DOI: 10.5281/zenodo.7602705

1 ORCID: 0000-0003-4596-8755 2 ORCID: 0000-0001-6527-2662 3 ORCID: 0000-0003-2814-8288 4 ORCID: 0000-0002-5113-0759



CC-BY-NC-SA-4.0





U.O. Sviluppo performance dell'archivio istituzionale della ricerca e servi:

Iter di strutturazione del questionario sulle abitudini dei ricercatori in materia di deposito dei dati della ricerca

Nell'ambito delle attività dell'obiettivo "Far crescere e diffondere in Ateneo il livello di consapevolezza sul valore della condivisione aperta dei dati della ricerca", assegnato alla Responsabile del Sistema Bibliotecario e Archivio storico di Ateneo (SBA), incardinato tra le azioni "O.2.4 - Migliorare la capacità di collaborazione e di attrazione dei fondi di ricerca sul piano internazionale, nazionale, regionale e favorire lo sviluppo dell'Open-Science" del Piano strategico 2021-2023¹, si è proceduto con una indagine conoscitiva finalizzata alla definizione di un questionario da sottoporre agli autori della ricerca scientifica di Ateneo in merito alle loro abitudini in materia di deposito dei dati della ricerca.

Tale indagine conoscitiva preliminare ha previsto l'analisi di siti, questionari e pagine web, di strutture di ricerca italiane (vicine alla realtà dell'Ateneo palermitano come target di riferimento), dedicati alla tematica della scienza aperta, con particolare riguardo all'accesso aperto ai dati.

Tra gli Atenei che dedicano una sezione ampia e ben circostanziata al tema in questione, sono risultati particolarmente adeguati alle finalità dell'indagine in esame quelli di Bologna, Milano Statale, Padova, Trento. Le sezioni web dei suddetti Atenei (e altre analoghe di altri Atenei, presumibilmente ispirate a queste) dedicate al tema in oggetto contengono soprattutto informazioni di ampio respiro, necessarie ad inquadrare al meglio la problematica e comprendere lo stato dell'arte in merito.

Non è, invece, risultato semplice rintracciare informazioni in merito ad indagini appositamente dedicate ad acquisire previamente, da parte degli autori, informazioni sul loro uso dei dati della ricerca, sulle abitudini correlate in merito alla conservazione e condivisione dei dati, sul loro livello di consapevolezza delle problematiche correlate. Alcuni spunti su possibili elementi da tenere presenti sono stati tenuti in considerazione a partire da alcuni questionari più genericamente destinati ad indagare sulla autovalutazione della ricerca dei Dipartimenti, come ad esempio nel caso dell'Università della Campania "Luigi Vanvitelli".

Tra i pochi modelli rintracciabili in rete di questionario propedeutico alla strutturazione di un servizio di Ateneo (consulenza agli autori e problematiche di archiviazione, diffusione e condivisione) appositamente dedicato alla questione dell'accesso aperto ai dati della ricerca, è risultato di particolare interesse quello erogato nell'ambito del progetto "BISA – Bibliosan per la Scienza Aperta"³, indagine condotta dall'Istituto Superiore di Sanità e presentata al convegno "I dati

_

¹ https://www.unipa.it/ateneo/.content/documenti/Piano_strategico_2021_2023-NUOVA_VERSIONE.pdf

² https://www.unicampania.it/doc/AQ/Ricerca/Questionario autovalutazione dipartimenti.pdf

http://www.bibliosan.it/ftp/bisa atti 15052017/POLTRONIERI 15 mag 2017.pdf



Settore Servizi per la ricerca e diffusione della conoscenza scientifica U.O. Sviluppo performance dell'archivio istituzionale della ricerca e servi; ib SISTEMA BIBLIOTE



aperti, cemento della scienza" del 2017. Il questionario si pone come il punto finale di una più complessa indagine sui dati aperti della ricerca⁴ portata avanti da un apposito gruppo di lavoro dell'Istituto. Tra i punti maggiormente significativi, anche ai fini dell'obiettivo assegnato alla sottoscritta, si rileva un interessante quadro europeo sui dati aperti della ricerca e sul contesto italiano e un rimando al progetto e-infrastructures Austria⁵, pioniere in ambito europeo nell'analisi della gestione pratica dei dati della ricerca e delle infrastrutture e servizi correlati.

Infine, ulteriori spunti di riflessione utili alla strutturazione del questionario sono stati tratti dai contenuti di una tesi di master ("Il valore dei dati della ricerca: il caso della Loughborough University Library") in Digital Library Learning⁶ e di un webinar ("Open Science dalla A alla Z")⁷ tenuto dalla dott.ssa Elena Giglia, dell'Università di Torino, nell'Ateneo palermitano in data 25 e 27 maggio 2021, nell'ambito degli eventi di formazione interna.

L'indagine qui descritta brevemente, portata avanti dal Settore "Servizi per la ricerca e diffusione della conoscenza scientifica"⁸ del Servizio speciale Sistema Bibliotecario e Archivio storico di Ateneo (SBA)⁹, ha permesso l'elaborazione di un questionario che è stato rivolto (mediante una procedura telematica con invio di risposta) agli autori della ricerca scientifica dell'Ateneo di Palermo, per sondare le loro consapevolezze in merito all'importanza, all'uso e alla condivisione dei dati e le loro necessità (già note o in via di sviluppo) ai fini di una possibile infrastruttura di raccolta e gestione dei dati a livello di Ateneo.

Il questionario è diviso in sezioni: Dati generali, Dati della ricerca, Formato dei dati, Archiviazione dei dati, Aspetti etici e legali, Accessibilità e riuso e le risposte sono del tipo "chiuse", con possibilità di risposta unica o multipla secondo il quesito. Il questionario si chiude con la possibilità (campo testuale aperto) di indicare "Eventuali commenti o osservazioni".

https://www.iss.it/strumenti-di-riferimento/-/asset_publisher/GIDBUf2rmBr2/content/id/5181021

https://zenodo.org/record/34005#.YmgM8NpBzIV

https://books.openedition.org/ledizioni/1481?lang=it

https://zenodo.org/record/3902797#.YmgO-9pBzIU

https://www.unipa.it/amministrazione/direzionegenerale/sba/settoreserviziperlaricerca/

https://www.unipa.it/amministrazione/direzionegenerale/sba/



U.O. Sviluppo performance dell'archivio istituzionale della

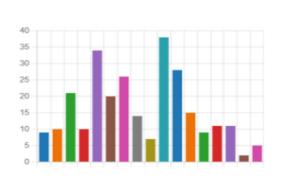


Analisi dei risultati del "Questionario sulle abitudini dei ricercatori in materia di deposito dei dati della ricerca"

Il sondaggio è stato compilato da 270 soggetti (46 professori ordinari, 108 professori associati, 60 ricercatori, 13 assegnisti, 43 dottorandi), in prevalenza di età compresa tra i 35 e 55 anni, distribuiti nelle varie aree disciplinari con una prevalenza di risposte nelle aree bibliometriche (nell'immagine sotto riportata viene rappresentata la distribuzione per aree disciplinari).

3. Area scientifica





La prima parte del questionario è stata finalizzata ad indagare il livello di conoscenza del tema e le abitudini consolidate, rispetto ai propri ambiti di ricerca, sulla produzione di dati e sulle loro caratteristiche.

Circa il 30% dei soggetti dichiara di non essere a conoscenza del fatto che alcuni programmi di finanziamento prevedono l'archiviazione dei dati raccolti/elaborati durante il progetto di ricerca; il 45% non conosce le caratteristiche che questi dati devono avere, mentre il 18,5% dichiara di non conoscere il significato dell'acronimo FAIR.

Il 57,4% degli intervistati dichiara di produrre e utilizzare dati di ricerca, il 23% di farlo "talvolta", il 6,3% "raramente", il 13,3% di non farlo "mai". Oltre la metà dei soggetti dichiara di utilizzare o di produrre prevalentemente "dati grezzi".



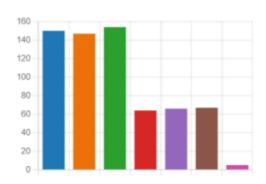




Si osserva che nei progetti di ricerca vengono generalmente prodotti dati in vari tipi di formato digitale, come evidenziato dal grafico seguente.

 In che tipo di formato digitale vengono raccolti ed elaborati i dati della ricerca nei tuoi progetti? [risposta multipla]





In questo caso si tratta di una domanda a risposta multipla e si può notare come vengono prodotti prevalentemente documenti di testo (150 occorrenze), fogli di calcolo (154 occorrenze), immagini (147 occorrenze). Si registrano, anche se in misura minore, risposte collegate a materiali multimediali (64 occorrenze), banche dati (66 occorrenze) e software (67 occorrenze). Di seguito quanto indicato da chi ha scelto la risposta "altro": elaborazioni su software statistici; dati numerici; codice, video, dati in formati vari (es. HDF5); immagini biomedicali (DICOM, NIFTI, etc.).

A seguire è stato richiesto se nell'identificazione del formato dei dati si tiene conto della possibilità di riuso da parte della comunità scientifica. La maggior parte dei soggetti (circa il 50% del totale) dichiara di tenerne conto; significativo appare che il 30% non si sia mai posto il problema e che il 7,5% non se ne preoccupi.

Molto interessante è quanto emerge dalla domanda, a risposta multipla, legata alle abitudini circa il salvataggio dei dati della ricerca. Qui è evidente che la maggior parte dei ricercatori salva i dati nel proprio pc (215 occorrenze); 109 soggetti hanno dichiarato di salvare i dati in un'applicazione cloud, 108 soggetti su un hard disk esterno, 58 soggetti su una chiavetta USB, 39 in un server della propria istituzione, 1 su un dvd riscrivibile.

Di seguito si elencano i dettagli delle risposte per le quali era richiesta ulteriore specifica:

- chi ha selezionato "archivio istituzionale", ha indicato: Drive, IRIS (7 occorrenze), YouUnipa e Drive, Drive community e Cloud unipa, Community Gmail Unipa, computer aziendali, Cloud universitario, Drive Unipa.
- chi ha selezionato "archivio di dati ad accesso aperto", ha indicato: OSF (2 occorrenze), Zenodo (3 occorrenze), github (3 occorrenze), dispositivo NAS, Orchid, Mendeley Data, ResearchGate, ArXiv.





Settore Servizi per la ricerca e diffusione della conoscenza scientifica U.O. Sviluppo performance dell'archivio istituzionale della ricerca e servi:

- chi ha selezionato "archivio di dati ad accesso riservato", ha indicato: Google drive (4 occorrenze), https://osf.io/, Dropbox (4 occorrenze), banche dati di progetto, archivi in dotazione di strumenti clinici, Forms, Google drive (3 occorrenze), Overleaf, Wetransfer, Server NAS, Dropbox (2 occorrenze).
- chi ha selezionato la voce "altro", ha indicato: Google drive (2 occorrenze), allegati elettronici riviste.

Le risposte fornite offrono diversi spunti di riflessione in merito al livello di conoscenza e/o consapevolezza degli strumenti a disposizione.

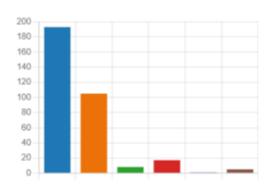
Riguardo alle dimensioni dei dati prodotti, la maggioranza dei soggetti (38 %) risponde che il volume medio dei dati prodotti in un anno non supera i 50 GB. Una quota consistente di soggetti non ha un'idea chiara della quantità di dati prodotti in un anno (circa il 18%) né della misura del singolo dataset (circa il 26 %). Il 40% dei soggetti dichiara di produrre singoli dataset di volume inferiore a 10 GB.

Relativamente alla definizione di ruoli e responsabilità nella gestione/conservazione dei dati di ricerca, la maggioranza dei soggetti (oltre il 52 %) dichiara che questi non sono codificati.

Nel grafico seguente si rappresentano le risposte fornite alla domanda che indaga su chi si occupa della gestione/conservazione dei dati di ricerca.

Nei tuoi progetti di ricerca, chi si occupa della gestione / conservazione dei dati di ricerca? [risposta multipla]





Si tratta di una domanda a risposta multipla dalle cui risposte si può notare come nella maggioranza dei casi è lo stesso ricercatore ad occuparsi della gestione e della conservazione dei dati (193 occorrenze). 105 soggetti dichiarano che è il responsabile del progetto o del gruppo di lavoro ad occuparsi di questo aspetto.

A seguire viene richiesto se solitamente viene individuato uno o più soggetti responsabile/i della validazione dei dataset gestiti nel progetto di ricerca: anche in questo caso i ricercatori potevano indicare più di una voce. Le risposte fornite evidenziano una netta prevalenza di ricercatori che dichiarano che non



U.O. Sviluppo performance dell'archivio istituzionale della



è previsto un momento formale di validazione del dataset finale (145 occorrenze); in 66 occorrenze viene dichiarato che il validatore è lo stesso soggetto responsabile della conservazione; in 19 occorrenze viene dichiarato che nel progetto è prevista una figura di informatico/statistico che valida il dataset finale, mentre solo in 5 occorrenze emerge che nel progetto è prevista una figura di esperto del dominio dei dati

Si elencano le risposte associate alla scelta "altro":

- guide

che valida il dataset finale.

- dipende dal progetto
- io stesso
- per lo più il provider del progetto ma non sempre esiste questa figura
- Earthchem
- non produciamo set di dati così vasti da necessitare trattamenti statistici
- facciamo e rifacciamo reazioni e ne valutiamo la riuscita
- i dati, quasi tutti o quelli consistenti appaiono poi nella relativa pubblicazione scientifica.

Un gran numero di soggetti (55,5 %) dichiara che nell'ambito dei progetti di ricerca non è previsto un documento per la gestione della documentazione e dei dataset o dichiara di non saperlo (20 %). L'8% dei soggetti dichiara di gestire la nomenclatura di dataset e documenti e il versioning di entrambi, mentre solo l'1,5% dichiara di gestire il versioning dei dataset.

È interessante quanto emerge a proposito del quesito che indaga su eventuali perdite fortuite dei dati trattati nei progetti di ricerca: il 73% dei soggetti dichiara di non essere mai incorsi nel problema. Per chi ha dichiarato di esservi incorso, la causa principale che lo ha determinato viene imputata a guasti, rotture o danneggiamento (24 occorrenze), cancellazione accidentale (17 occorrenze) , obsolescenza del formato o del supporto (15 occorrenze).

Un'altra sezione del questionario è focalizzata sugli aspetti legati all'accessibilità ed al riuso dei dati. Mettendo a confronto le risposte ai due quesiti che indagano su quali sono i soggetti che possono accedere ai dati durante il suo svolgimento e poi a progetto concluso, si possono rilevare alcune differenze, come è visibile dai due grafici sotto riportati.

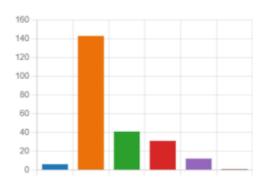






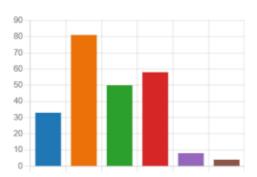
30. Chi può accedere ai dati di ricerca durante il progetto?





32. Chi può accedere ai dati di ricerca a fine progetto?



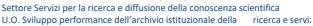


Risulta evidente che la voce più selezionata è quella legata alla condivisione dei dati con il proprio gruppo di ricerca; tuttavia, a fine progetto cresce leggermente sia la numerosità di chi ha selezionato la voce "chiunque", sia quella legata a "tutti i partner del progetto" che quella dell'"invio via posta su richiesta". Resta comunque una certa quota di risposte che testimonia come nessun altro possa accedere ai dati di ricerca, né durante la fase progettuale né a progetto concluso.

Chi ha scelto di specificare ha puntualizzato in entrambe le domande che questo dettaglio dipende dal progetto e che non è possibile generalizzare.

Il quesito successivo, a risposta multipla, è dedicato alla modalità di condivisione dei dati. Appare confortante il basso numero di soggetti che non condividono i dati (11 occorrenze). Molti sono invece i ricercatori che hanno un'abitudine a rendere disponibili i dati tramite cloud (115 occorrenze) e tramite e-mail (133 occorrenze); sembra invece non proprio radicata l'abitudine ad utilizzare strumenti cooperativi come il collaborative notebook (7 occorrenze).

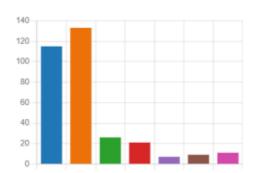






34. Con quale modalità rendi disponibili i dati della tua ricerca? [risposta multipla]





Solo alcuni dei 21 soggetti che hanno dichiarato di utilizzare un archivio di dati hanno specificato:

- Google drive (3 occorrenze)
- https://osf.io/
- *Dropbox* (3 occorrenze)
- *Iris* (4 occorrenze)
- Mendeley Data
- strumenti clinici
- cloud
- OneDrive
- hard disk

Chi ha scelto la voce "altro" ha indicato:

- PC utilizzato come server connesso alla rete LAN del dipartimento
- Pubblicazioni scientifiche
- Li mando su richiesta
- Pubblico i risultati della ricerca e l'intera documentazione raccolta durante la ricerca
- Mediante invio e-mail
- Google drive, riviste scientifiche, stampa cartacea o open access
- Trasferimento dati via wetransfer.

Anche in questo caso si rileva una generalizzata mancanza di conoscenza e/o consapevolezza degli strumenti.

Coloro i quali non condividono i dati della propria ricerca hanno fornito le seguenti motivazioni nel campo libero disponibile:

- temo che possa esserne fatto un uso improprio o non corretto
- temo di violare le norme sulla riservatezza dei dati
- temo che possano essere sfruttati per uso commerciale





 non posso sostenere i costi e i tempi per corredarli delle informazioni che permettano di renderli riutilizzabili

U.O. Sviluppo performance dell'archivio istituzionale della

- non lo ritengo utile
- non lo ritengo necessario
- li condivido con altri ricercatori in funzione di comuni obiettivi.

Le risposte fornite qui inducono alcune riflessioni sul livello di conoscenza del tema, ad esempio in materia di licenze d'uso.

Il 70% di soggetti risponde positivamente al quesito successivo che indaga l'abitudine dei ricercatori di utilizzare dati esterni (ad es. provenienti da altre ricerche; condivisi da ricercatori esterni al progetto...). Il restante 30% dichiara di non utilizzare dati esterni per le seguenti motivazioni:

- non esistono dati disponibili (19 occorrenze)
- esistono ma non si possono utilizzare (6 occorrenze)
- non so se esistono (35 occorrenze)
- non so come citare correttamente i dati esterni e non voglio avere problemi di proprietà (8 occorrenze).

Di seguito le risposte dei 13 soggetti che hanno scelto di indicare altre motivazioni: non servono; nel mio settore non si utilizzano dati di ricerca; non è mai stato necessario; wasn't needed; uso dati nuovi; le mie ricerche si basano su dati originali raccolti nel corso della stessa ricerca; non si rende necessario; nel mio settore disciplinare si usano bibliografie e prodotti scientifici elaborati nell'ambito di altri progetti, non dati; nella ricerca qualitativa la condivisione di dati è più complessa in quanto non è possibile cumulare i dati per eventuali generalizzazioni dei risultati; non mi servono; non riguardano la mia disciplina; faccio ricerca di base; in ambito giuridico raramente rilevano.

Oltre il 66% dei soggetti dichiara di non avere mai avuto problemi legali utilizzando dati esterni; dalle altre risposte si rileva comunque un certo timore alla condivisione, probabilmente collegata alla sopra citata mancanza di conoscenza in materia di licenze d'uso.

A seguire è stato richiesto il livello di utilizzo da parte dei ricercatori di "dati sensibili": la maggioranza (circa il 45,5 %) ha dichiarato di non produrne o utilizzarne "mai", il 27 % di farlo "raramente", il 22,6 di farlo "talvolta" e il 4,8 % di farlo "sempre".

La parte finale del questionario è dedicata ad una riflessione generale sui vantaggi che porterebbe la condivisione dei dati nella ricerca e su aspetti legati al tipo di supporto e alle azioni che occorrerebbero ai ricercatori.

Alla domanda (a risposta multipla) su quali vantaggi porterebbe la condivisione dei dati della ricerca si può notare una netta preponderanza di soggetti che riconoscono un valore positivo a questo aspetto. Si registrano infatti 189 occorrenze per la risposta "possibilità di riutilizzo per il progresso della ricerca scientifica"; 195 occorrenze per la risposta "possibilità di intraprendere nuove collaborazioni di ricerca";





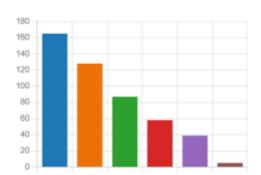
Settore Servizi per la ricerca e diffusione della conoscenza scientifica U.O. Sviluppo performance dell'archivio istituzionale della ricerca e servi:

117 occorrenze per la risposta "creare valore, evitando sprechi e/o duplicazioni di ricerche già svolte maggiore visibilità e impatto della mia ricerca"; 50 occorrenze per la risposta "è importante per il finanziamento e/o la valutazione del mio progetto"; solo 12 occorrenze per la risposta "nessuno".

Altrettanto significativo sembra il fatto che i ricercatori dichiarano di avere necessità di un'infrastruttura tecnica dedicata, corsi di formazione, assistenza legale e servizio di helpdesk. Esiste anche un certo numero di soggetti che afferma di non avere necessità di supporto o servizi particolari.

46. Che tipo di supporto/servizio ti occorrerebbe per gestire i dati? [risposta multipla]





Di seguito le risposte di chi ha scelto la voce "altro":

- personale informatico
- nel caso di dati biomedicali, una più rapida interazione con Commissione Etica e/o Commissione Bioetica dell'Ateneo
- Research assistant che si occupano della gestione dei dataset.

L'ultima domanda proposta ai ricercatori (a risposta multipla) mira a comprendere quali siano le aspettative dei ricercatori circa le azioni dell'Ateneo su queste tematiche. Le risposte registrate testimoniano che un corposo numero di ricercatori si aspetta delle azioni concrete da parte dell'Istituzione. Nello specifico sono state registrate 183 occorrenze per la risposta "formazione ai ricercatori sul concetto di dati, sulla corretta gestione, sulle tecniche di diffusione e condivisione"; 146 occorrenze per la risposta "linee guida/politiche di indirizzo"; 114 occorrenze per la risposta "previsione di professionalità ad hoc". Interessanti sono anche le risposte di chi ha selezionato la voce "altro": creazione di un dataset condiviso cui eventualmente accedere per integrare i propri dati nel rispetto della privacy e delle proprietà intellettuali; infrastrutture; zero burocrazia; laboratori; libertà; creazione strutture informatiche di conservazione dati e formazione ai ricercatori; acquisire nuove banche dati da rendere fruibili a tutti i ricercatori di Ateneo; maggiore valorizzazione dei risultati intermedi emergenti dai processi di ricerca; maggiore valorizzazione delle pubblicazioni prodotte durante il progetto di dottorato; l'accesso gratuito a software di analisi dati (GraphPad, MatLab ecc...); aiuto nel superamento di alcune "barriere" registrate da parte di enti esterni a condividere dati già disponibili per una loro eventuale diffusione; una interfaccia e un cloud che siano di immediato utilizzo e caricamento oltre che di gestione.



U.O. Sviluppo performance dell'archivio istituzionale della



In generale, l'analisi dei risultati del questionario ha permesso di rilevare alcuni dati estremamente significativi e di confermare alcune delle direzioni che il Settore Servizi per la ricerca e diffusione della conoscenza scientifica e l'U.O. Sviluppo performance dell'archivio istituzionale della ricerca e servizi bibliometrici hanno già intrapreso, nei mesi passati, con l'obiettivo di migliorare i servizi offerti agli autori della ricerca presso Unipa.

In sintesi, si evince un certo interesse dei ricercatori Unipa verso i temi della conservazione e della condivisione dei dati della ricerca, ma un livello di conoscenza non abbastanza approfondito e radicato. Oltre la metà degli intervistati dichiara di produrre dati e altrettanti dichiarano di non essere a conoscenza delle caratteristiche che i dati devono avere in seno ad alcuni programmi di finanziamento. Poco consolidate sono l'abitudine alla redazione di documentazione a supporto e la cura del versioning, probabilmente collegata al limitato utilizzo del Data Management Plan. Altrettanto poco utilizzata è la prassi della "validazione" dei dati prodotti. Si registra invece un'abitudine abbastanza diffusa alla condivisione dei dati di ricerca nel proprio gruppo di lavoro, prevalentemente con l'utilizzo di cloud, così come una spiccata sensibilità a riconoscere i vantaggi che porterebbe una maggiore condivisione dei dati della ricerca. Va però sottolineato che alcuni ricercatori evidenziano timori legati a possibili usi impropri dei dati condivisi.

Molto significative sono le richieste espresse dai ricercatori nella parte conclusiva del questionario indirizzate per lo più sulla richiesta di "formazione ai ricercatori sul concetto di dati, sulla corretta gestione, sulle tecniche di diffusione e condivisione", sull'approvazione di "linee guida/politiche di indirizzo" e sulla "previsione di professionalità ad hoc" in seno all'Ateneo. Ciò consente di confermare l'utilità della nuova sezione del portale, implementata qualche mese fa, dedicata ad una serie di FAQ sull'Open Science.