

Research Data

I Big Data della ricerca

Dati della ricerca: cosa sono?

- Sono fatti, osservazioni o esperienze sulle quali è costruita e/o verificata una teoria. I dati possono essere numerici, descrittivi, sonori o visivi.
- Vengono raccolti, osservati e/o creati, per scopi di analisi e soprattutto per produrre risultati di ricerca originali



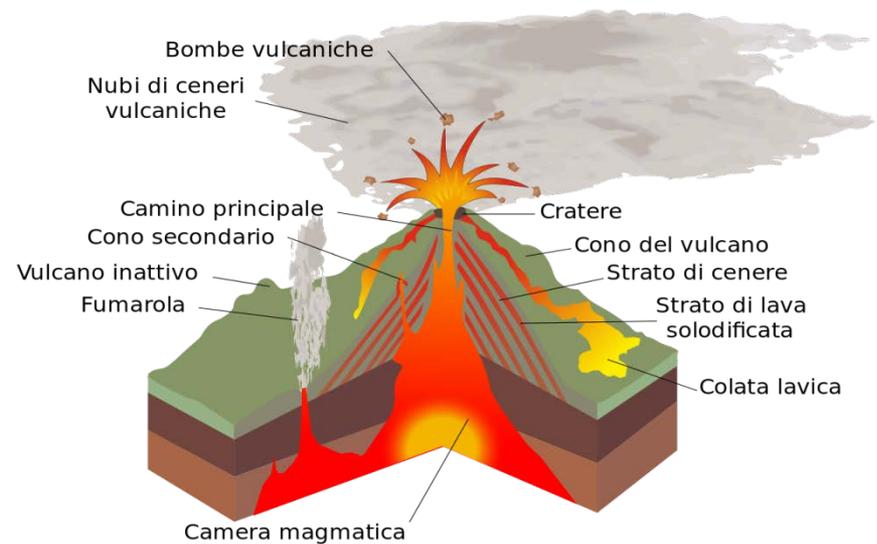
Sono parte essenziale di ogni processo di ricerca

Gestirli e amministrarli in modo consapevole e sistematico è un imperativo dal quale ogni istituzione di ricerca pubblica o privata non può prescindere



La ricerca sempre più «data intensive»

- Costante crescita di dati della ricerca in formato digitale in moltissime discipline
- Non solo nelle scienze (astronomia, fisica, climatologia, vulcanologia, medicina, chimica, matematica ecc..)
- Ma anche nelle scienze sociali e umane (statistica, economia, psicologia, archeologia, storia dell'arte ecc..)



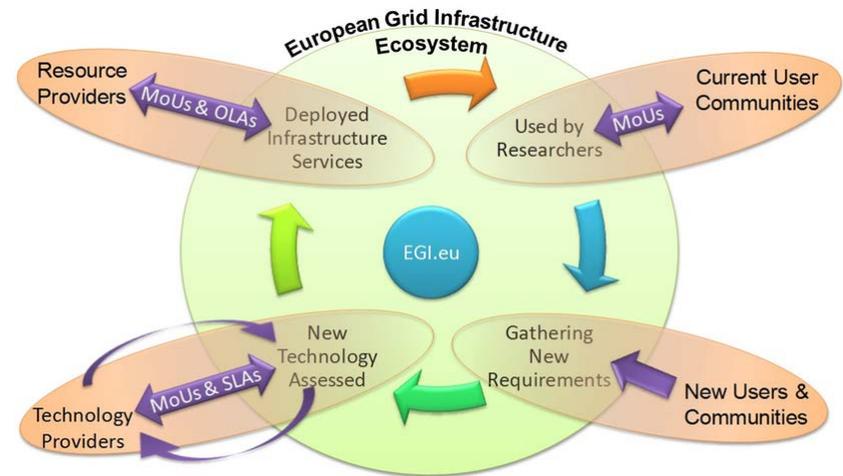
Il contesto «data intensive»

eScience

eResearch

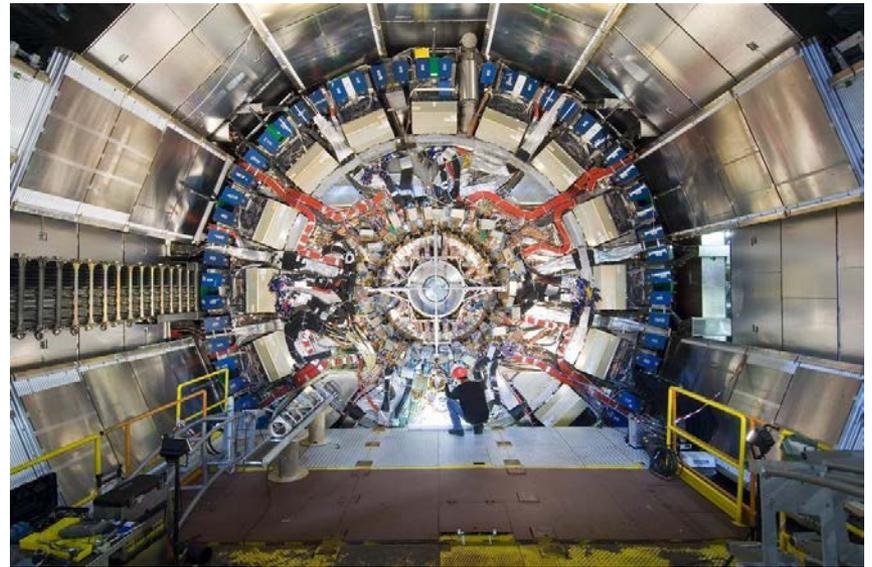
eScience

- La scienza basata su sistemi di calcolo molto potenti e performanti distribuiti in rete che grazie alla loro potenza e alla memoria sono in grado di analizzare grandi quantità di dati (large data sets)
- Esempio: European Grid infrastructure
 - un'infrastruttura distribuita di elaborazione dati per supportare decine di migliaia di ricercatori di molte discipline, con capacità di analisi dei dati senza precedenti.



eResearch

Collaborazione tra ricercatori che avviene tramite la rete indipendentemente dalla distanza, tramite l'uso di risorse di calcolo/computer distribuite per la produzione, elaborazione dati condivisi



Tipologia dati della ricerca

Dati strutturati

- **Matrici di dati**
- **Dati transazionali**

Grafi

- **Web e Social network**
- **Strutture molecolari**

Dati Ordinati

- **Dati spaziali**
- **Serie storiche**
- **Sequenze**
- **Sequenze genetiche**

Dati non strutturati

- **Documenti testuali**
- **Immagini**
- **Audio e Video**

Perché è importante gestire e amministrare i dati della ricerca

Per il bisogno dei ricercatori di avere un supporto

- tecnico (per il 60%)
- legale (per il 40%)
 - Dati risultati da varie indagini nei paesi europei tra queste nelle università austriache e britanniche nel 2015

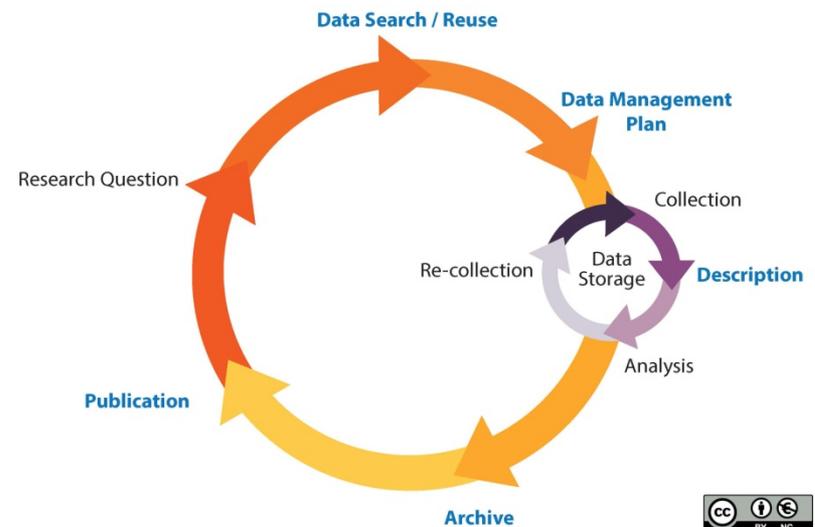
Strategie per l'organizzazione / gestione dei dati di ricerca

- **Data Management Plan (DMP)**
 - è un documento formale che illustra come i dati devono essere gestiti sia durante un progetto di ricerca che dopo il completamento del progetto.
- **L'obiettivo di un piano di gestione dei dati è quello di considerare i molti aspetti della gestione dei dati**
 - della generazione di metadati
 - della conservazione dei dati e dell'analisi prima dell'inizio del progetto
- **Questo assicura la «Data Curation»**
 - Cioè che i dati siano ben gestibili nel presente e siano preparati per la conservazione in futuro

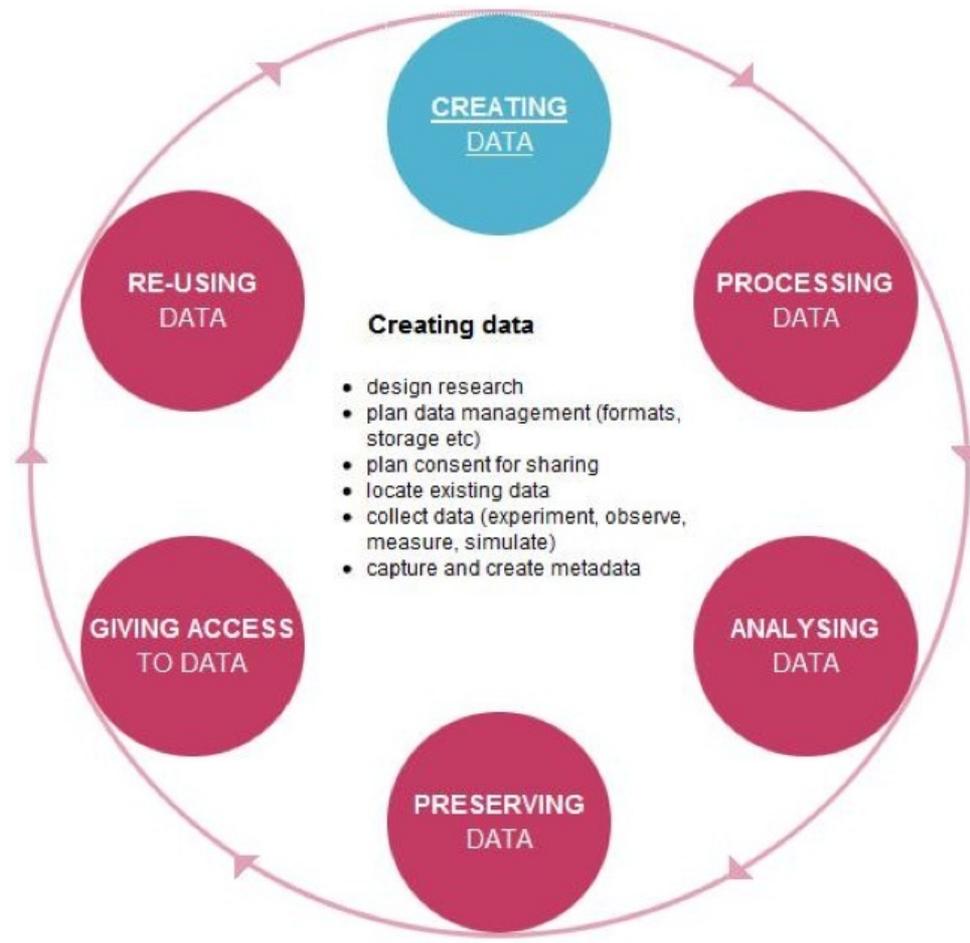
Data Curation

Secondo l'Università dell'Illinois *Graduate School of Library and Information Science*, "data curation è la gestione dei dati al fine di scoprire, recuperare, mantenere, aggiungere valore, permettere il riutilizzo dei dati, garantendone contemporaneamente l'autenticazione, l'archiviazione, la gestione e la preservazione"

The Research Data Curation Lifecycle



Ciclo di vita dei dati



La proprietà intellettuale dei dati della ricerca

- **Il progresso scientifico e tecnologico dipende dal controllo e gestione dei diritti di proprietà intellettuale nella ricerca**
 - Circa il 90% delle produzioni intellettuali prodotte dal sistema di ricerca internazionale è chiuso entro piattaforme editoriali con accesso a pagamento
- **Quindi lo sfruttamento efficace dei risultati di ricerca dipende da una corretta gestione della proprietà intellettuale, che dovrebbe essere parte integrante nella gestione complessiva della conoscenza generata nell'ambito di un progetto**

Gestione della proprietà intellettuale dei dati della ricerca

- È necessario identificare il proprietario dei dati e le condizioni per il loro utilizzo: chi produce dati, chi li manipola, chi li finanzia
- La proprietà (ownership) e altri diritti di proprietà intellettuale sui dati spesso sono incerti: le problematiche sui diritti – dovute anche alle possibili diverse giurisdizioni - vanno risolte a monte attraverso accordi tra le parti
- Necessario avere a monte adeguate policy istituzionali sulla proprietà dei dati
- Nei progetti finanziati o in collaborazione con altri partner magari privati i ricercatori devono sempre accertarsi se hanno il diritto di aprire i dati alla condivisione
- Se si stanno riutilizzando dati di proprietà di terze parti è necessario negoziare con il produttore eventuali permessi di uso e riuso
- La gestione e amministrazione dei diritti è funzionale alla loro rivendicazione, e può aiutare a difendersi in caso di frode scientifica

Research Data: progetti in corso in Europa

OpenAIRE e OpenAIRE plus

- un portale per l'accesso ai [risultati e ai dati della ricerca finanziata nel 7° ProgrammaQuadro](#)



Research Data: progetti in corso in Italia

- **SmartData**

- Progetto finanziato dal CINECA per gestire le nuove problematiche “BigData” ed HPC (supercalcolo) fornendo, accanto all'infrastruttura per la produzione, l'archiviazione e la fruizione dei dati, un servizio di analisi avanzata sui dati prodotti dagli utenti che ne favorisce il riuso anche da parte di altre comunità scientifiche



Research Data: progetti in corso in Italia

- **MAPPA** - Metodologia Applicate alla Predittività del Potenziale Archeologico
 - Il primo open digital archaeological archive italiano, che rende accessibili tutti i dati pubblici delle indagini archeologiche
 - Uso di sistemi e procedure standardizzate per la redazione e la gestione dei dati archeologici

The screenshot shows the MAPPA website interface. At the top, there are navigation links: "home | partner | contatti | sitemap | credits". On the right, there are social media icons for Facebook, Twitter, YouTube, and LinkedIn, along with language options "italiano | english". The main header features the MAPPA logo and the tagline "metodologie applicate alla predittività del potenziale archeologico". Below the logo, there is a menu with items: "Il progetto", "Chi siamo", "Pubblicazioni", "Biblioteca", "News", "Multimedia", "MAPPAGIS", "MAPPAopenDATA", and "Servizi". To the right of the menu, there are logos for "REGIONE TOSCANA", "FAS Fondo Area Sottosviluppate 2007-2013", and the "UNIVERSITÀ DI PISA" logo. Below the logos, the text reads: "UNIVERSITÀ DI PISA Dipartimento di Scienze Archeologiche Dipartimento di Scienze della Terra Dipartimento di Matematica". A search bar is located below the text. The main content area features a large image of a 3D map with a red "Perche'" button and a navigation bar with "Come", "Chi", "Quando", and "Dove" buttons. Below the image, there is a section titled "Mappa Gis" with the text: "Pubblicare, interrogare, analizzare, conoscere. Il WebGIS di MAPPA è Carta Archeologica e Carta del Potenziale Archeologico di Pisa: uno strumento interattivo e aperto, per la tutela, la governance e la ricerca." To the right of this section is a red box titled "Mappa Open Data" with the text: "Conservare, disseminare, collaborare. L'archeologia libera e aperta è la nuova frontiera della ricerca." At the bottom of the page, there is a footer with the text: "Prossimi eventi: Seminario sui dati aperti della ricerca, Vercelli 26 ottobre 2012 MAPPAproject partecipa a: 'Il futuro open della ricerca. Iniziative e...' -->