

BUONE PRATICHE PER OTTENERE LA FAIRNESS: UNA PROPOSTA PER INAF (ISTITUTO NAZIONALE DI ASTROFISICA)

Francesca Martines - INAF - Osservatorio Astronomico di Palermo
francesca.martines@inaf.it



Oggi la scienza produce enormi quantità di dati, che devono essere gestiti correttamente in tutte le fasi della loro esistenza, anche dopo la conclusione del progetto che li ha generati. C'è inoltre una spinta molto forte verso l'Open Science, basata sul principio che i risultati della ricerca pagata con fondi pubblici devono essere disponibili. È quindi necessario convergere il più possibile verso dati FAIR (Findable, Accessible, Interoperable and Reusable).

Le competenze dei bibliotecari possono dare un utile contributo alla gestione dei dati della ricerca, poiché l'essenza del lavoro di un bibliotecario è organizzare le informazioni in modi standard e consentirne la scoperta e l'uso. Per raggiungere questo obiettivo, le informazioni devono essere strutturate e descritte correttamente, come accade per i materiali di cui si occupano tradizionalmente i bibliotecari. Le competenze dei bibliotecari sono quindi molto importanti nella gestione dei dati di ricerca (RDM), con l'implementazione di standard e la consulenza per la standardizzazione e la semantica dei descrittori dei dati.

Dunque, la «FAIRness» è un obiettivo. Ma... come raggiungerla? E nel nostro caso, come raggiungerla all'interno dell'Istituto Nazionale di Astrofisica?

QUAL È LA SITUAZIONE IN ASTRONOMIA?

L'astronomia presenta alcune specificità che la distinguono da altre discipline e che la pongono, per molti versi, in una posizione più favorevole alla cosiddetta «FAIRificazione» dei dati. Per inquadrare meglio la situazione, è opportuno per prima cosa schematizzare «come funziona» il lavoro degli astronomi. La maggior parte di essi presenta delle proposte di osservazione presso strutture (telescopi) nazionali e internazionali, a terra e in orbita, e quelle approvate danno luogo alle osservazioni vere e proprie. I risultati di tali osservazioni sono prodotti sotto forma di dati grezzi (raw data) archiviati a cura della struttura osservativa, e per un periodo che può andare dai 6 ai 24 mesi sono accessibili solo dal responsabile della proposta (PI/Principal Investigator) e dal suo gruppo; trascorso tale periodo i dati diventano generalmente pubblici. I ricercatori lavorano sui dati effettuando riduzioni, simulazioni, filtraggi, spesso anche generando software specifici: in base a ciò che ottengono, solitamente producono una o più pubblicazioni basate sui risultati del proprio lavoro.

Possiamo quindi sintetizzare le componenti essenziali del lavoro della maggior parte degli astronomi come segue:

I dati (osservativi), i cosiddetti «**dati grezzi**».

I dati ottenuti come risultato delle elaborazioni ad opera dei ricercatori, ma anche il software utilizzato per queste operazioni: chiameremo tutti questi «**dati derivati**».

Le **pubblicazioni**, basate sui dati e sui processi di cui sopra.

Se esaminiamo questi elementi alla luce dei principi FAIR, il quadro che si presenta è il seguente:

Dati grezzi

- Sono strutturati in modo (più o meno) standard (ad esempio, FITS).
- La conservazione è garantita dalle strutture osservative che li generano e li archiviano.
- Sono (moderatamente) reperibili ma... devono essere citati nelle pubblicazioni!
- Sono dotati di metadati ad opera di chi gestisce l'archivio.
- Sono accessibili: più o meno aperti, generalmente dopo un periodo di embargo.

Dati derivati

- Sono archiviati sui PC dei singoli ricercatori (non è garantita la conservazione).
- Non sono reperibili.
- Non sono dotati di metadati.
- Non sono accessibili.

Pubblicazioni

- Sono conservate in archivi ad accesso aperto (open access repository), quindi la conservazione è garantita.
- Sono reperibili.
- Sono dotate di metadati.
- Sono accessibili.

C'È CHIARAMENTE UN GROSSO ANELLO DEBOLE IN QUESTA CATENA...

- I **dati grezzi** e le **pubblicazioni** sono generalmente strutturati, provvisti di metadati, conservati e resi disponibili (archivi dei telescopi, depositi OA, ecc.) e sono «sufficientemente FAIR».

- I **dati derivati**, in assenza di uno specifico piano di gestione dei dati (DMP) che stabilisca dove debbano essere conservati e come debbano essere strutturati e dotati di metadati, sono generalmente memorizzati sui PC dei singoli ricercatori, o su spazi di archiviazione generici o su piattaforme «pubbliche» che non garantiscono la conservazione degli oggetti (per non parlare della FAIRness). Questi dati non sono quindi reperibili, accessibili, interoperabili o riutilizzabili.

Un altro problema è che i dati vengono citati - quando va bene - nei modi più vari all'interno delle pubblicazioni: è evidente che una standardizzazione in questo processo renderebbe i dati molto più facilmente ricercabili e, soprattutto, agevolerebbe il collegamento tra i dati e le pubblicazioni su di essi basati. Attualmente non esistono standard formali per la citazione dei dati, sebbene molto lavoro sia stato svolto da varie organizzazioni (RDA, DCC UK, etc.) Recentemente è stato pubblicato sull'*Astrophysical Journal Supplement* un articolo* che suggerisce una serie di best practices per la pubblicazione di dati nella letteratura astronomica: tali indicazioni costituiscono un'eccellente base di partenza per la redazione di una policy in tal senso, di cui la comunità astronomica nel suo insieme, ed INAF nello specifico, ha certamente bisogno.

**Best Practices for Data Publication in the Astronomical Literature*, Tracy X. Chen et al 2022 ApJS 260 5,
<https://iopscience.iop.org/article/10.3847/1538-4365/ac6268>

DI COSA HA BISOGNO INAF PER RAGGIUNGERE LA FAIRNESS?

- Diffondere la cultura FAIR all'interno dell'ente e farne comprendere ai ricercatori l'importanza e l'utilità per la comunità scientifica.
- Avere una policy istituzionale sui dati FAIR che indichi ai ricercatori che i loro dati devono essere FAIR e perché. E' in corso di creazione un gruppo di lavoro dedicato.
- Redigere un Data Management Plan che dia indicazioni specifiche e operative su come gestire i dati.
- Offrire uno spazio di archiviazione con la possibilità di attribuire un DOI agli «oggetti» selezionati dai ricercatori: il processo di richiesta del DOI comporta l'implementazione di tutta una serie di metadati in ottica FAIR. Poiché questa esigenza è presente quasi esclusivamente per i dati derivati, in questo modo essi vengono depositati, ne viene garantita la conservazione e possono essere resi FAIR. Questa sezione è già in avanzata fase di realizzazione.
- Avere personale dedicato (Data Scientists e Data Steward).
- Infine, ma non meno importante, disporre di linee guida per la citazione dei dati (per i dati grezzi e derivati).