



The banner features a row of six icons: a globe, a book, a handshake, a money bag with a Euro symbol, a scale of justice, and a bicycle. Below the icons, the text reads 'AIUCD 2021' in large black letters, followed by 'DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale' and '10° congresso annuale PISA 19-22 gennaio'. On the right side, a list of topics is displayed in colored text: 'DIGITAL PUBLIC HUMANITIES' (red), 'OPEN CULTURE' (orange), 'RETI SOCIALI' (yellow), 'TECH ECONOMY' (green), 'E-PARTICIPATION' (blue), and 'TECNOLOGIE ASSISTIVE' (purple). The background includes binary code and a classical building facade.

AIUCD 2021

DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale

10° congresso annuale **PISA** 19-22 gennaio

DIGITAL PUBLIC HUMANITIES
OPEN CULTURE
RETI SOCIALI
TECH ECONOMY
E-PARTICIPATION
TECNOLOGIE ASSISTIVE

Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

Di cosa parliamo quando parliamo di FAIR?

Francesca Frontini¹, Anas Fahad Khan²

¹ Istituto di Linguistica Computazionale "A. Zampolli", Italia – francesca.frontini(«»)ilc.cnr.it

² Istituto di Linguistica Computazionale "A. Zampolli", Italia – fahad.khan(«»)ilc.cnr.it

ABSTRACT

Nel 2016 un consorzio di scienziati afferenti a diverse istituzioni e discipline enuncia i principi FAIR; in questi quattro anni l'importanza e la portata del programma FAIR è divenuta sempre più evidente. L'adesione a tali principi nelle discipline umanistiche sembra farsi largo, ma non senza difficoltà e interrogativi. Questo lavoro propone una riflessione sulle implicazioni della proposta FAIR per la gestione dei dati scientifici, confrontandola con la sua effettiva ricezione nella comunità delle DH in Italia e in Europa.

PAROLE CHIAVE

Principi FAIR, Open Data, dati della ricerca, politiche della ricerca, EOSC.

1. INTRODUZIONE

Nel marzo 2016, un consorzio di scienziati afferenti a diverse istituzioni e discipline pubblica "FAIR Guiding Principles for scientific data management and stewardship" (Wilkinson et al. 2016), una serie di principi e linee guida identificati per la gestione dei dati scientifici.

Quello che, in un clima sempre più favorevole agli *open data* e alla condivisione dei risultati scientifici, poteva a prima vista sembrare l'ennesimo richiamo a già note buone pratiche di data curation, si rivela ben presto come un programma più concreto e ambizioso. L'acronimo è accattivante e se vogliamo anche rassicurante - FAIR sta per *findable* (trovabile), *accessible* (accessibile), *interoperable* (interoperabile), *reusable* (riutilizzabile), e l'assonanza con la parola inglese *fair*, richiama evidentemente le categorie semantiche di equità, giustizia, pari condizioni (si pensi ad espressioni come *fair play*, o *fair trade*) (Amiel et al 2020). Tuttavia, i principi FAIR¹ enunciati nel manifesto stilano una lista di requisiti di stampo tecnico ben precisi e stringenti; l'iniziativa appare poi da subito dotata di un notevole sostegno istituzionale.

Le FAIR *data policies* sono infatti ormai diventate anche una parte importante dell'iniziativa European Open Science Cloud (EOSC) della Commissione Europea, che mira a sviluppare un'infrastruttura di servizi al fine di promuovere la scienza aperta. Ad esempio, il progetto FAIRsFAIR², finanziato dall'UE, ha l'obiettivo di sviluppare un'infrastruttura della conoscenza e implementare soluzioni pratiche per assistere gli scienziati (e qui possiamo intendere il termine "scienziato" in un senso estremamente ampio, che comprende i ricercatori che lavorano nelle scienze sociali e umanistiche) nel rendere i loro dati più FAIR. Vi è poi EOSCFair4, uno dei gruppi di lavoro EOSC, e parte integrante della sua struttura di governance, che è dedicato a "implementare i principi dei FAIR data" e ha l'obiettivo esplicito di promuovere "l'interoperabilità interdisciplinare". Già oggi, nei *data management plan* dei progetti europei e nazionali (in alcuni paesi) è di fatto necessario far riferimento alla FAIRizzazione dei risultati della propria ricerca.

Tutto questo ci fa capire che i principi FAIR sono destinati a diventare imprescindibili per il ciclo di vita dei dati scientifici, anche per chi lavora nel settore delle scienze umane. Tuttavia, in questo ambito l'adesione ai principi FAIR sembra più lenta. Una possibile motivazione si trova nel recente documento pubblicato dal gruppo di lavoro EOSCFair dal titolo "Six Recommendations for Implementation of FAIR Practice"³ che rileva gli atteggiamenti delle varie comunità disciplinari; i risultati mostrano che molti ricercatori vedono i principi FAIR come adatti solo ai dati quantitativi e digitali⁴. Questo è forse anche dovuto al fatto che, almeno all'inizio, la comprensione delle implicazioni pratiche e tecnologiche dietro ai principi astratti è stata parziale e limitata ad una lettura superficiale e generica dei principi espressi. Nonostante questo ritardo, osserviamo come oggi un certo numero di progetti e pubblicazioni in ambito DH facciano riferimento ai principi FAIR, e pensiamo che sia utile cercare di studiare la ricezione di questo nuovo paradigma all'interno della nostra comunità.

Nella nostra presentazione, delineaeremo prima di tutto in breve le idee chiave dei principi FAIR, partendo da un esempio concreto, quello proposto da Barend Mons e riguardante i dati biomedici, mostrando prima di tutto la differenza tra i principi FAIR nella loro formulazione attuale e, ad esempio, pratiche più generiche di *Open Data*. Ripercorreremo in

¹ <https://www.force11.org/group/fairgroup/fairprinciples>

² <https://www.fairsfair.eu/>

³ Deniz Beyan et al 2020. Disponibile anche come Google doc: <https://docs.google.com/document/d/1VKbI29Dc5jHhkVVGZkUTD-L8UeqIllvsSjiNeurjc6k/edit?usp=sharing>

⁴ FAIR "was made for quantitative data (e.g., social sciences and humanities) or that it is not suitable for the study of real world objects because that is different from the study of digitised objects (e.g., humanities but much less in natural history collections)".

seguito la ricezione dei principi FAIR analizzando alcune pubblicazioni dal 2016 in poi, in Italia e all'estero. Trarremo infine alcune conclusioni preliminari al fine di aprire una discussione sulle conseguenze dei cambiamenti in atto per la nostra comunità scientifica.

2. FAIR DATA & SERVICES, FAIR DIGITAL OBJECTS, KNOWLETS - COSA VUOL DIRE FAIR NELLE SCIENZE BIOMEDICHE

Non passeremo qui in rassegna i principi uno per uno, ma già ad una prima lettura si nota il riferimento a protocolli e standard informatici di stoccaggio, accesso e metadato dei dati, nonché alle licenze di utilizzo. Inoltre, e soprattutto, il riferimento alle macchine, oltre che agli umani, come utenti finali di tali dati.

Il fatto che non si tratti semplicemente dell'esposizione di idee astratte risulta visibile anche nel sostegno che da subito l'iniziativa ha ricevuto a livello istituzionale e scientifico. Ne è la prova la creazione del consorzio internazionale GoFAIR⁵, nel cui *steering committee* siedono rappresentanti dei governi francese, olandese e tedesco, e che si prefigge di implementare concretamente i principi FAIR nei vari settori disciplinari. Ed è infatti in Europa che la promozione dei principi FAIR sta ricevendo il grande impulso, anche in connessione al già citato EOSC.

Ma qual è l'obiettivo ultimo del programma FAIR? A prima vista esso sembra avvicinarsi a idee propuginate dai fautori dell'*Open Data* - che punta a rendere disponibili con licenze aperte i dati della ricerca, ma non è esattamente così. FAIR non vuol dire esattamente *open*; i dati FAIR possono anche essere dati chiusi, o aperti solo per alcuni. Alla base del programma FAIR vi è la possibilità di interrogare in maniera standardizzata e combinare grandi set di dati di provenienza diversa, con l'obiettivo di imprimere un'accelerazione alle scoperte scientifiche in vari ambiti. Il paradigma FAIR ha dunque una forte attinenza ai principi del Web Semantico e dei *Linked Data* (non necessariamente Open), in quanto raccomanda la descrizione di metadati e dati secondo modelli formali e vocabolari condivisi.

Forse l'ambito in cui questo programma trova la sua applicazione più chiara è quello delle scienze bio-mediche. Uno dei più noti sostenitori dei principi FAIR è Barend Mons⁶ professore al Leiden University Medical Center, nei Paesi Bassi. Nella sua recente presentazione a EOSS 2020 dal titolo "The FAIR Guiding Principles in Times of Crisis"⁷ egli illustra il suo programma per un riuso più efficace dei dati scientifici nel suo settore, ed in particolare per l'uso dei big data per identificare i trattamenti più efficaci per curare i sintomi del COVID-19. Considerando la mole di pubblicazioni sull'argomento e il fatto che una ricerca ha bisogno di essere confermata indipendentemente più e più volte prima di essere considerata un dato scientifico assodato, è facile capire come in questo settore al momento la semplice lettura di articoli scientifici da parte di umani non sia più sufficiente. Secondo il prof. Mons la visione FAIR dovrebbe permettere di costruire un vero e proprio *Internet of FAIR Data & Services*⁸, composto di *FAIR Digital Objects*⁹. In particolare, i risultati della ricerca dovrebbero essere pubblicati non solo sotto forma testuale, ma anche sotto forma di grafi di conoscenza, in cui ogni informazione, ad esempio la proprietà di un dato farmaco, o i suoi effetti secondari, possano essere trasformati in *knowlets*, unità di conoscenza strutturata, standardizzata, FAIR. Esplorando un enorme grafo di conoscenza composto dagli studi sull'uso dell'idrossiclorochina per la cura delle complicazioni del COVID-10, Mons mostra come sia possibile "vedere" che, nel complesso, gli articoli che riscontrano effetti positivi sono pochi.

Ascoltando la presentazione ci si rende conto che, almeno nelle scienze della vita, i principi FAIR non sembrano rappresentare solo buone pratiche di gestione dei dati, ma un programma ambizioso che mira a modificare ed in qualche modo "automatizzare" il modo di fare ricerca. Va comunque detto che anche su questa visione a lungo termine i sostenitori di FAIR non sono tutti d'accordo; nel già citato "Six Recommendations for Implementation of FAIR Practice"¹⁰ si sottolinea che il FAIR per le macchine, per quanto importante, rappresenta un obiettivo molto difficile da raggiungere.

3. RICEZIONE DEI PRINCIPI FAIR NELLE SCIENZE UMANE - UN'ANALISI DI CORPUS

A nostro avviso la portata del programma FAIR è ben lungi dall'essere stata pienamente compresa e interiorizzata nelle scienze umane e sociali, dove prevale una visione dei principi FAIR forse ancora un po' "ingenua", che in buona sostanza si riduce alla adesione a buone pratiche di *data curation* già note ad alcuni, ma ora *à la page* e riassunte da uno slogan accattivante. Al fine di verificare questa intuizione, ci serviremo di una analisi di articoli ed abstract sottomessi alle conferenze AIUCD e DH (ADHO) a partire dal 2016¹⁰, anno della pubblicazione dei principi FAIR. I risultati preliminari (si veda la Tabella 1) mostrano la penetrazione dei principi FAIR nella comunità DH (almeno europea) a partire dal 2017.

⁵ <https://www.go-fair.org/>

⁶ https://en.wikipedia.org/wiki/Barend_Mons

⁷ <https://eoss2020.inf.unibz.it/programme/>

⁸ <https://www.go-fair.org/resources/internet-fair-data-services/>

⁹ <https://fairdo.org/>

¹⁰ Questo lavoro preliminare è stato condotto sugli atti delle conferenze, disponibili online, tuttavia intendiamo approfondire la ricerca utilizzando l'Index of Digital Humanities conferences, reso disponibile di recente (Eichmann-Kalwara et al 2020)

In quell'anno un ruolo importante è stato svolto dal progetto PARTHENOS, che ha contribuito a calare i principi nel contesto delle humanities¹¹. Come settore trainante emergono le scienze del patrimonio (*Digital Cultural Heritage*); in quest'ultimo ambito l'esistenza di ontologie formali in RDF come CIDOC-CRM ha probabilmente facilitato l'adesione a vocabolari condivisi, indispensabili ad una vera FAIRizzazione dei (meta)dati. Tuttavia, negli altri ambiti il riferimento ai *Linked Data* in relazione FAIR è raro e per molti progetti essi non costituiscono un principio ispiratore. Il riferimento a principi FAIR (magari solo a un principio) appare in alcuni casi chiosa *ex post*, che serve a giustificare delle scelte fatte a monte. Un articolo può ad esempio dichiarare l'adesione ai principi di riusabilità e interoperabilità sottolineando la scelta di uno standard di codifica del testo o una licenza aperta, tralasciando però quelli di trovabilità e accessibilità che richiederebbero il deposito del dataset in un centro dati riconosciuto (“data is deposited in a trusted repository”) che permette l'*harvesting* di metadati.

In DH sono quindi ancora rari i lavori nei quali un dataset viene presentato argomentando punto per punto l'adesione ai principi FAIR sulla base delle scelte tecnologiche. Più frequenti invece gli articoli, seminari o panel dove la discussione intorno a FAIR avviene nel contesto di grandi infrastrutture di ricerca o progetti infrastrutturali (si vedano ad esempio Dumouchel e Giglia 2019 o Wessel 2020), cosa che sottolinea come l'implementazione di tali principi non possa essere demandata ai singoli progetti, ma necessiti di interventi strutturali e strumenti ben precisi.

Menzioni dei principi FAIR	2016	2017	2018	2019	2020
AIUCD articoli ed abstract (inclusi panel)	-	-	2 (terminologia; open data e infrastrutture)	2 (cultural heritage; Linguistic Linked Data)	3 (principi FAIR per DH; edizioni digitali; geolinguistic)
DH abstract (paper, panel, seminari)	-	1	1	11	4

Tabella 1 – Paper e abstract delle conferenze AIUCD e DH che menzionano i principi FAIR, con indicazioni delle tematiche.

4. CONCLUSIONI PRELIMINARI

Nel problematizzare la ricezione dei principi FAIR da parte della comunità DH, non intendiamo proporre un'evangelizzazione acritica dei riluttanti umanisti al programma FAIR, ma piuttosto mettere in evidenza anche le contraddizioni e le difficoltà che da più parti cominciano ad essere espresse in particolare nel campo delle DH. L'implementazione dei requisiti FAIR, con tutta la loro complessità tecnica, richiede investimenti importanti, che tendono a favorire i grandi centri di ricerca dei paesi più ricchi, dotati di infrastrutture e finanziamenti ad hoc. In questo senso FAIR non vuole sempre dire *fair*, come ha sottolineato Antonio Rojas Castro nella sua recente presentazione proprio sulla difficoltà di applicarli da in progetti e dataset provenienti dal *Global South* (BBAW 2020).

Ci sono comunque ragioni per essere ottimisti. L'adozione del paradigma dei *Linked (Open) Data* in una serie di discipline di SSH mostra che ci sono aree in cui una forte convergenza verso linguaggi, modelli e vocabolari standard aiuta a supportare il tipo di interoperabilità sintattica e semantica - all'interno e tra discipline - che è alla base dei principi FAIR. Questo è vero in campi come la linguistica e aree correlate come le risorse linguistiche e la lessicografia, ma anche negli studi medievali, per i dati geospaziali e quelli del patrimonio culturale ecc. A questo si aggiunge l'uso ormai diffuso delle linee guida TEI che aiutano a codificare diversi tipi di risorse utilizzando un modello unificato basato su XML. Come sottolinea Mons, i lavori pionieristici sulla standardizzazione dei metadati delle risorse linguistiche sono stati all'avanguardia anche rispetto alle scienze naturali. Questo mostra che esiste una solida base di culture e pratiche su cui FAIR può essere implementato nelle SSH.

RINGRAZIAMENTI

Questo abstract è ispirato dai lavori svolti dagli autori nel contesto del progetto MSH-SUD Montpellier CommonData, e dei progetti H2020 SSHOC (grant agreement 823782) e TRIPLE (grant agreement 863420). Ringraziamo i nostri colleghi di ILC-CNR per le stimolanti discussioni sulle tematiche trattate.

¹¹ Si veda ad esempio il 'PARTHENOS – Joint PARTHENOS-CLARIN-DARIAH Workshop on Applying the FAIR Principles PARTHENOS', svoltosi nel 2017, <http://www.parthenos-project.eu/joint-parthenos-clarin-dariah-workshop-on-applying-the-fair-principles/>, oltre che i riferimenti in bibliografia.

BIBLIOGRAFIA

- [1] Amiel, Philippe, Francesca Frontini, Pierre-Yves Lacour, and Agnès Robin. 'Pratiques de gestion des données de la recherche : une nécessaire acculturation des chercheurs aux enjeux de la science ouverte ?'. Résultats d'une enquête exploratoire dans le bassin montpellierain (juin 2018). *Cahiers Droit, Sciences & Technologies*, no. 10 (11 May 2020): 147–68. <https://doi.org/10.4000/cdst.2061>.
- [2] Barbuti, Nicola. 'Ripensare i Dati Come Risorse Digitali: Un Processo Difficile?' In *AIUCD 2020 – Book of Abstracts*. Milano, 2020.
- [3] BBAW. *Antonio Rojas Castro: FAIR Enough? Building DH Resources in an Unequal World*, 2020. <https://vimeo.com/445147368>.
- [4] Deniz Beyan, Oya, Neil Chue Hong, Stefano Cozzini, Marta Hoffman-Sommer, Rob Hooft, Liisi Lembinen, Juuso Marttila, and Marta Teperek. 'Seven Recommendations for Implementation of FAIR Practice'. Zenodo, 23 June 2020. <https://doi.org/10.5281/zenodo.3904140>.
- [5] Dumouchel, Suzanne, and Elena Giglia. 'CO-OPERAS IN: Integration And Cooperation To Face Fragmentation And Address Complexity In The SSH'. DataverseNL, 2019. <https://doi.org/10.34894/AJF13Q>.
- [6] Eichmann-Kalwara, Nickoal, Scott B. Weingart, Matthew Lincoln, and et al. 'The Index of Digital Humanities Conferences'. Carnegie Mellon University, 2020. <https://dh-abstracts.library.cmu.edu>. <https://doi.org/10.34666/k1de-j489>.
- [7] ICCD. 'Linee guida per l'applicazione dei principi FAIR alla gestione e al riuso dei dati - ICCD - Istituto Centrale per il Catalogo e la Documentazione'. Accessed 5 September 2020. <http://www.iccd.beniculturali.it/it/150/archivio-news/4722/linee-guida-per-l-applicazione-dei-principi-fair-alla-gestione-e-al-riuso-dei-dati>.
- [8] PARTHENOS, Hella Hollander, Francesca Morselli, Frank Uiterwaal, Femmy Admiraal, Thorsten Trippel, and Sara Di Giorgio. 'PARTHENOS Linee guida per l'applicazione dei principi FAIR alla gestione e al riuso dei dati', 2 September 2019. <https://doi.org/10.5281/zenodo.3363243>.
- [9] Sbarbati, Stefano. 'The FAIR Principles and the EOSC Concept in PARTHENOS Community – PARTHENOS Project'. Accessed 5 September 2020. <http://www.parthenos-project.eu/the-fair-principles-and-the-eosc-concept-in-parthenos-community>.
- [10] Wessel, Leon. 'CLARIN for DH Scholars'. Presented at the DH2020 Book of Abstracts, 2020. https://dh2020.adho.org/wp-content/uploads/2020/07/348_CLARINforDHScholars.html.
- [11] Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (15 March 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.