

Infrastrutture, terminologie e policy per la ricerca umanistica: note per un confronto interdisciplinare

Annarita Liburdi, Cristina Marras, Ada Russo

CNR-ILIESI, Istituto per il Lessico Intellettuale Europeo e Storia delle Idee

Abstract. L'intervento riguarda le infrastrutture digitali e la Policy dei dati della ricerca. Descrive l'esperienza del gruppo di ricerca del CNR-ILIESI nel progetto PARTHENOS. Si concentra sull'open access e sull'open science nella preparazione, pubblicazione, gestione, accesso e riuso di documenti/testi/informazioni su piattaforme interoperabili. Guarda con attenzione alla terminologia e alla necessità di un Glossario dedicato nella stesura di linee guida per la gestione e regolamentazione di questa interoperabilità. Presenta le specificità della ricerca storico filosofica e umanistica digitale e alcuni problemi e prospettive aperti dal confronto con la costruzione di infrastrutture digitali e di workflows per la ricerca, facendo emergere come inevitabile un confronto con le pratiche e i protocolli delle altre discipline.

Keywords. Interdisciplinarietà, Open access, Open Science, Scienze Umane, Glossari

Introduzione

Il programma scientifico dell'Istituto per il Lessico Intellettuale Europeo e Storia delle Idee (ILIESI) del Consiglio Nazionale delle Ricerche (CNR) fa dello studio della terminologia di cultura la chiave d'accesso alla storia del pensiero filosofico e al costituirsi dell'identità culturale europea. I risultati delle attività di ricerca sono rappresentati da studi critici, archivi digitali, metodologie digitali per l'analisi testuale. Obiettivo dell'ILIESI è integrare i risultati di 50 anni di ricerca alle nuove tecnologie, confrontandosi con esperienze interdisciplinari nazionali e internazionali. In particolare vengono testati strumenti e processi di standardizzazione per la creazione di ambienti integrati e interoperabili; strumenti di annotazione semantica a supporto dello scambio e creazione di nuova conoscenza; modelli per l'organizzazione della conoscenza e la ricerca interdisciplinare; strumenti per la gestione e la valorizzazione del multilinguismo. In questo contesto si innesta la partnership con il progetto Horizon 2020 Parthenos, Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies e la specificità dei tasks assegnati al gruppo di ricerca ILIESI (di cui fanno parte oltre alle autrici di questo intervento Giancarlo Fedeli e Hansmichael Hohenegger).

Parthenos vede insieme 16 partners al fine di realizzare una infrastruttura di ricerca europea che si ponga come cluster capace di rafforzare e promuovere la collaborazione tra i diversi settori delle scienze umane. Per questo uno dei suoi obiettivi principali è la definizione di standard comuni e di azioni condivise, oltre a strumenti e applicazioni per una armonizzazione delle policies.

In particolare, le attività si sono concentrate sulla stesura di "Guidelines for Common

Policies Implementation”, che costituiscono un ponte tra differenti campi e stakeholders nelle scienze umane. La ricerca ha previsto a tal fine una mappatura delle policies riguardanti il data management e la qualità dei dati, metadati, repository e IPR (Intellectual Property Rights), open data e open access e ha individuato linee di integrazione, i principi FAIR, per la qualità dei dati. L'idea di fondo è “make your data as FAIR as possible”.

1. Infrastrutture digitali e problemi di terminologia

Uno degli aspetti della riflessione condotta dal gruppo ILIESI è legato alla terminologia condivisa in Parthenos, progetto che pone tra i suoi obiettivi la condivisione di dati. Nella redazione dei documenti di progetto è emersa infatti la difficoltà a convergere su una terminologia di riferimento univoca e condivisa. Da ciò la necessità di creare un glossario per definire i termini, sciogliere gli acronimi e documentare standard e tecnologie.

Nella piattaforma SSK “Standardization Survival Kit”, una piattaforma dedicata alla promozione di un uso diffuso degli standard nelle scienze umane, è stata realizzata una tassonomia ordinata gerarchicamente con un approccio sintetico per macro temi, all'interno dei quali sono contenuti i termini. Ma tale approccio non facilita l'accesso ai termini, non valorizza la molteplicità di domini di utilizzo, e costringe a percorsi di ricerca predefiniti non sempre esaustivi. Inoltre, pur se tassonomie e glossari sono vocabolari controllati, nelle tassonomie la classificazione delle parole è intesa principalmente per la ricerca, mentre nei glossari è orientata ai contenuti e al loro significato.

Da qui la nostra riflessione sull'opportunità di un glossario collaborativo, strutturato non per macro temi ma per termini che funzionano come chiavi d'accesso e che permettono la costruzione dinamica di raggruppamenti. Per esempio l'acronimo TEI (Text Encoding Initiative) è nello stesso tempo un acronimo che va sciolto, uno standard che va documentato e l'istituzione che gestisce lo standard. Non deve dunque comparire in modo ridondante in macro temi diversi, ma può assumere il ruolo di chiave di accesso per questa molteplicità. È importante poi ricostruire il contesto di ogni termine per renderlo il più completo possibile: definizione, fonte della definizione, e, là dove opportuno, sitografia e

Fig. 1
Problemi di
terminologia

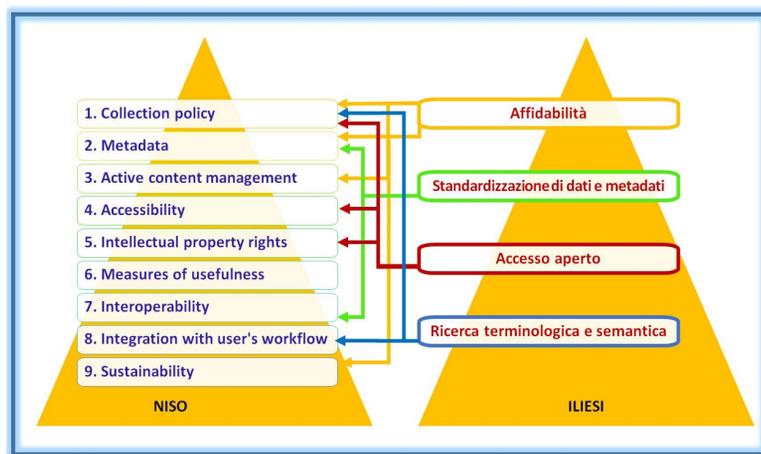


istituzioni di riferimento. Essenziale è, inoltre, assicurare la correttezza dell'informazione, attraverso ad esempio la cronologia delle modifiche che contestualizzano l'entrata del glossario. Non ultimo è il problema del multilinguismo: la terminologia di riferimento per la comunità scientifica è l'inglese, ma il glossario deve essere capace di poter parlare a una comunità plurilingue stabilendo tuttavia una 'nomenclatura' univoca e condivisa (Fig. 1). A differenza di un thesaurus, il glossario non prevede gerarchie. In questo modo è possibile far emergere la molteplicità tematica di appartenenza di un singolo termine, di mettere in evidenza le possibili associazioni e intersezioni tra ambiti senza essere incanalati in percorsi precostituiti, e si favorisce la dimensione euristica della ricerca terminologica.

2. Open science: teorie e pratiche

Gli archivi digitali dell'ILIESI costituiscono un esempio di infrastruttura ad accesso aperto che raccoglie materiali eterogenei, ma coerenti rispetto a tematiche di carattere filosofico e/o terminologico. Gli archivi si basano su programmi open-source e codifiche standard testuali e per i metadati, usano strumenti per l'arricchimento semantico e tecnologie per la rappresentazione della conoscenza. La loro finalità è favorire e supportare l'attività di ricerca, disseminare i risultati, consentirne il riuso. Le linee guida seguite dall'ILIESI per la realizzazione degli archivi digitali rispecchiano i requisiti definiti a livello nazionale e internazionale per le collezioni digitali e riguardano soprattutto la necessità di garantirne l'affidabilità e la sostenibilità nel tempo; di condividere standard per dati e metadati per favorire l'uso e il ri-uso; di assicurare l'accesso aperto agli archivi (Fig. 2, Liburdi et al. 2015).

Fig. 2
Linee guida
ILIESI per una
collezione
digitale



Nella realizzazione degli archivi digitali, sono stati centrali gli aspetti legati alla condivisione dei dati della ricerca e all'accesso aperto alle pubblicazioni, perché indispensabili per la realizzazione della scienza aperta. È evidente come anche in un ambito di ricerca umanistica specifico il problema della policy e del rapporto tra teoria e pratica è cruciale. Il CNR, ente pubblico di ricerca, per esempio è impegnato a sostenere i principi della libera circolazione della conoscenza, ha aderito alla "Berlin declaration" e ha firmato il "Position

Statement” sull’accesso aperto ai risultati della ricerca in Italia. Tuttavia, non si è dato una policy che guidi in modo chiaro e trasparente le pratiche dei ricercatori dell’Ente negli aspetti legali degli open data e dell’open access. In qualche modo anche nell’ambito della nostra esperienza di Istituto sentiamo l’esigenza di avere delle best practices condivise che pur tenendo conto delle differenze disciplinari - come per esempio la differenza del periodo di embargo per la pubblicazione - contemplate dalla legge italiana (Legge 7/12/2013, n. 112), tutelino il ricercatore di ogni ambito disciplinare nell’attuazione dell’open access e dell’open science.

3. Conclusioni

Se l’accesso aperto alle pubblicazioni scientifiche cresce e diventa una pratica sempre più consolidata e regolamentata, quello relativo ai dati della ricerca è diventato centrale solo di recente. Per lo più si esita a renderli pubblici per il timore di vederli riportati impropriamente, di non vederli attribuire correttamente o di avvantaggiare eventuali competitors.

Una protezione eccessiva dei dati si scontra con il modello collaborativo della scienza aperta, inibendo una possibile dialettica virtuosa tra la salvaguardia/protezione della paternità di un dato/idea/manufatto e della sua creazione, e la libertà di riutilizzarlo e migliorarlo. L’accesso aperto può inoltre entrare in conflitto con l’IRP, o meglio con una interpretazione della proprietà intellettuale della ricerca piuttosto restrittiva.

Attualmente la ricerca pone l’accento su collaborazione e condivisione anziché sui tradizionali concetti di ‘proprietà intellettuale’ e pubblicazione. Indubbiamente l’accesso aperto ai dati è uno dei pilastri della moderna metodologia di ricerca che si basa sempre più sulla cooperazione nel lavoro e su nuove modalità di distribuzione e diffusione della conoscenza attraverso le tecnologie digitali.

Su queste basi lo scambio dinamico di open data innesca processi di innovazione e riproducibilità creando contesti scientifici sempre più interattivi e integrati; in tali contesti, e nelle infrastrutture scientifiche che ad essi fanno riferimento, il linguaggio scientifico e tecnico è ampiamente condiviso e la terminologia di dominio diventa un indicatore sensibile dello stato e sostenibilità delle infrastrutture stesse. In questa prospettiva la policy assume un ruolo fondamentale di guida per i ricercatori, nell’attuazione di open access e open data e per l’identificazione di perimetri comuni (terminologia, standard, etc..) nell’ambito dei quali esercitare la scienza aperta.

Riferimenti bibliografici

Archivi Digitali ILIESI: <http://www.iliesi.cnr.it/risorsedigitali>

Decreto Legge 8 agosto 2013, n. 91: <http://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legge:2013;91>

FAIR data principles: <https://libereurope.eu/wp-content/uploads/2017/12/LIBER-FAIR-Data.pdf>

Istituto per il Lessico Intellettuale Europeo e Storia delle Idee: <http://www.iliesi.cnr.it>

Liburdi, A., Russo, A., Tardella, M. (2015): Ricerca filosofica e archivi digitali: quale im-

patto sulla diffusione della cultura scientifica e sui beni culturali?, Poster. IV Convegno AIUCD, Digital Humanities e beni culturali: quale relazione? Torino, 17-19/12/ 2015.

Position Statement CNR: <https://www.cnr.it/it/position-statement>

Progetto PARTHENOS: <http://www.parthenos-project.eu>

Autrici



Annarita Liburdi - annarita.liburdi@cnr.it

Dal 1982 è nei ruoli del Consiglio Nazionale delle Ricerche e dal 2006 è Primo Tecnologo assegnato all'Istituto per il Lessico Intellettuale Europeo e Storia delle idee. Ha lavorato presso l'Istituto di Analisi dei Sistemi ed Informatica, la Biblioteca Centrale del CNR, il Comitato Nazionale di consulenza per le Scienze Matematiche e presso l'Istituto di Scienze e Tecnologie della Cognizione. Le sue aree di interesse sono: biblioteche e archivi digitali; accesso aperto; ipertesto.

Cristina Marras - cristina.marras@cnr.it

Ricercatrice presso l'Istituto per il Lessico Intellettuale Europeo e Storia delle Idee del Consiglio Nazionale delle Ricerche, affianca la sua ricerca in filosofia, filosofia del linguaggio e umanistica digitale con attività di valorizzazione del dialogo interdisciplinare attraverso l'esplorazione dei diversi linguaggi e delle tecnologie che favoriscono la condivisione di metodi, pratiche e risultati della ricerca.



Ada Russo - ada.russo@cnr.it

Tecnologo presso l'Istituto per il Lessico Intellettuale Europeo e Storia delle Idee del Consiglio Nazionale delle Ricerche, si occupa di trattamento automatico di dati documentali e testuali; realizzazione di strumenti specifici per l'elaborazione, l'organizzazione e reperimento di informazioni di tipo linguistico-lessicografico; uso di thesauri terminologici strutturati come chiave di accesso a materiali lessicografici; analisi di sistemi di gestione di banche di dati documentali e testuali.