

Un modello per la compilazione di Data Management Plan per l'archeologia

Sara Di Giorgio¹, Paola Ronzino²

¹ICCU, Istituto Centrale per il Catalogo Unico delle biblioteche italiane per le informazioni bibliografiche

²PIN Scrl Servizi didattici e scientifici per l'Università di Firenze

Abstract. Il contributo presenta il template per la creazione di Data Management Plan (DMP), sviluppato dal progetto PARTHENOS con l'obiettivo di offrire uno strumento che soddisfi i requisiti degli istituti finanziatori dei progetti di ricerca, fornendo al contempo uno strumento che permetta ai ricercatori di documentare la propria ricerca. Il modello, che trae origine dall'iniziativa europea dell'Open Science e dai principi FAIR, è stato adattato alle esigenze dei ricercatori nel settore archeologico, ed è stato testato da un gruppo di esperti del dominio. Il contributo presenta gli ultimi aggiornamenti apportati al template e include una panoramica sugli strumenti implementati per facilitarne la compilazione e produrre un modello che sia machine-actionable.

Keywords. Data Management Plan, FAIR, Open Science, archeologia

Introduzione

I Data Management Plan (DMP) sono degli strumenti che permettono di raccogliere informazioni su come i dati saranno creati, definendone i piani per la loro condivisione e conservazione nel tempo, indicando eventuali restrizioni da applicare ai dati di ricerca nelle varie fasi del progetto.

La politica lanciata dalla Commissione Europea in materia di Open Innovation e Open Science ha dato luogo a una serie di iniziative con l'obiettivo di rendere i dati di ricerca scientifica accessibili e di garantire il riutilizzo, la redistribuzione e la replicabilità della ricerca (1).

In questo contesto, nel 2015 sono stati sviluppati i FAIR principles (2, 3), i quali fungono da linea guida per i ricercatori che intendano migliorare la riusabilità dei loro dati fornendo una serie di indicazioni affinché i dati siano reperibili, findable, accessibili, accessible, interoperabili, interoperable e riutilizzabili, reusable.

Per perseguire tale obiettivo, il programma Horizon 2020 ha pubblicato le linee guida sulla gestione dei dati FAIR e ha fornito ai beneficiari di finanziamenti H2020 il modello FAIR del DPM di H2020 (4), da presentare entro i primi sei mesi dall'inizio di un progetto. Partendo dalle linee guida di H2020 sulla gestione dei dati FAIR, PARTHENOS ha elaborato un modello per la definizione di un DMP specificatamente pensato per la comunità archeologica (5).

I DMP si presentano generalmente sotto forma di documenti compilati dai ricercatori

e nella maggior parte dei casi la procedura di creazione è eseguita manualmente oppure mediante l'utilizzo di strumenti online (6). I file ottenuti utilizzando le soluzioni digitali si presentano spesso in formati testuali, come file PDF o DOC, orientati all'human readability, cioè ottimizzati per essere letti in modo naturale dalle persone. Solo in alcuni casi gli strumenti disponibili permettono di ottenere in uscita un formato machine readable, come JSON o XML, cioè i cui dati sono codificati per essere processati in modo automatico dai calcolatori.

Il presente contributo include una panoramica sugli ultimi aggiornamenti del template e degli strumenti sviluppati per rendere il modello facilmente utilizzabile e machine-actionable.

1. PARTHENOS DMP per la Open Research in archeologia

PARTHENOS (7) è un progetto H2020 che mira a rafforzare la coesione della ricerca nel settore degli studi linguistici, dei beni culturali, della storia, dell'archeologia attraverso un gruppo tematico di infrastrutture di ricerca europee, favorendo l'interoperabilità dei dati e formulando standard e politiche comuni in materia di gestione dei dati.

In PARTHENOS, la comunità archeologica è rappresentata da ARIADNE (8), una rete europea che è stata finanziata tra il 2013 e il 2017, e che ha sviluppato un'infrastruttura di ricerca per l'integrazione dei dati archeologici distribuiti negli archivi digitali di tutta Europa, consentendone la scoperta e l'accesso per mezzo del suo portale (9).

Un'indagine condotta tra gli esperti del consorzio e successivamente estesa ad altri esperti del dominio, ha fornito al gruppo di lavoro di PARTHENOS un quadro completo di standard e buone pratiche per la creazione, l'archiviazione e la condivisione dei dati utilizzati dalla comunità archeologica, i quali sono stati utilizzati come punto di partenza per preparare l'elenco dei requisiti e linee guida offerte ai ricercatori per la compilazione del DMP.

Il modello del DMP di PARTHENOS mantiene la stessa struttura di quello di H2020, in modo da utilizzare un modello oramai familiare ai ricercatori. Rispondendo alle domande contenute nel DMP, i ricercatori forniscono informazioni su:

- i dati prodotti dalla ricerca;
- gli standard usati per strutturare i dati;
- il sistema di identificazione unica e persistente dei dati;
- la metodologia di archiviazione dei dati;
- la metodologia di condivisione dei dati con informazioni sulle condizioni di accesso e di riutilizzo dei dati.

Rispetto al DMP di H2020, nel modello DMP di PARTHENOS sono forniti degli elenchi in cui è possibile selezionare standard e flussi operativi adottati in archeologia, offrendo spunti per la compilazione.

La necessità di supporto alla compilazione di un DMP è stata fortemente sottolineata da un gruppo di esperti, i quali hanno risposto a un'indagine condotta dal team di Data Management del progetto OpenAIRE e dal gruppo di esperti sui dati FAIR (10). L'obiettivo di questo sondaggio è stato quello di raccogliere feedback da un gruppo di

esperti che hanno valutato il template DMP di H2020 al fine di identificare eventuali lacune e raccogliere suggerimenti per il miglioramento.

2. Sviluppi tecnici del DMP di PARTHENOS

Una delle attività in fase di finalizzazione prevede la creazione di collegamenti tra le linee guida del PARTHENOS DMP e gli altri strumenti sviluppati da PARTHENOS e già disponibili online sul sito del progetto. In particolare, saranno possibili i collegamenti con i contenuti del Wizard e lo Standardization Survival Kit (SSK). Si tratta di due strumenti online realizzati per offrire un supporto pratico ai ricercatori che non hanno familiarità con il processo di gestione dei dati e che così possono conoscere in modo interdisciplinare, come i dati dovrebbero essere raccolti, elaborati, archiviati e condivisi con altri ricercatori. Il Wizard offre una serie di documenti di policy sulla gestione dei dati organizzati per aree tematiche di ricerca e lo SSK, attraverso l'uso di scenari di ricerca, guida i ricercatori nella selezione e nell'uso degli standard più appropriati per le loro particolari discipline e flussi di lavoro.

Il PARTHENOS DMP offrirà quindi una vasta gamma di elementi da utilizzare per la compilazione del DMP, adatti anche a utenti meno esperti, grazie ai ricchi contenuti offerti dai vari strumenti.

Per agevolare la compilazione del DMP di PARTHENOS, è stato sviluppato un applicativo ad hoc, la cui progettazione ha tenuto in considerazione sia le necessità pratiche dei ricercatori, sia l'attuale evoluzione tecnologica che i documenti digitali stanno subendo.

L'obiettivo finale è quello di ottenere dei machine-actionable DMP (11), cioè dei DMP le cui informazioni possano essere elaborate e rese comprensibili automaticamente dai calcolatori, e che siano allo stesso tempo dei documenti interoperabili, aggiornabili e condivisibili all'interno della comunità degli stakeholder. L'interfaccia messa a disposizione online è stata studiata per agevolare la compilazione del DMP mediante l'utilizzo di soluzioni intuitive e user friendly.

Le domande a cui il ricercatore è invitato a rispondere sono suddivise in pagine successive, dotate di una barra di avanzamento comune che si presenta come il punto di riferimento principale per l'utilizzatore. La visualizzazione complessiva delle varie parti che compongono il modello guida passo a passo l'utente, indicando approssimativamente il tempo necessario per finirli.

Ogni pagina raggruppa domande tematiche simili, distinte in obbligatorie e facoltative, arricchite da pop-up informativi di aiuto alla compilazione. Nel caso in cui alcuni dei punti obbligatori per la sottomissione del DMP non siano stati compilati, il loro numero sarà visualizzato in rosso nella barra di avanzamento. Al termine della procedura di compilazione sarà possibile scaricare le informazioni in esso contenute in formato PDF, oppure in JSON.

Il file JSON è fondamentale all'interno dell'applicativo poiché offre all'utente la possibilità di salvare una versione del proprio lavoro. La compilazione del questionario, infatti, può essere interrotta in qualsiasi momento scaricando il file JSON contenente i dati

correnti. Tale file potrà essere ricaricato all'interno dell'interfaccia online per proseguire e terminare il lavoro.

Se invece la compilazione del questionario è definitiva, i dati contenuti nel file costituiranno una versione del DMP utile per eventuali successive revisioni o aggiornamenti.

La progettazione degli sviluppi futuri dell'applicativo è finalizzata a rendere i dati contenuti all'interno dei DMP condivisibili e interoperabili tra quelle comunità di ricerca che adotteranno delle soluzioni comuni per agevolare la cooperazione tra i propri sistemi. Rendere dei documenti interoperabili significa fare in modo che le informazioni in essi contenuti possano essere scambiate tra sistemi diversi in maniera completa e affidabile. Per questo è necessario considerare sia l'aspetto sintattico sia quello semantico dei dati. I calcolatori possono trattare e gestire in maniera sintattica la maggior parte delle informazioni, se queste sono codificate in formati standard come XML o JSON, ma non sono in grado di interpretarle e "comprenderle" se queste non sono modellate utilizzando dei vocabolari controllati e degli standard condivisi.

I DMP generati nella prima versione dell'applicativo soddisfano già i requisiti per l'interoperabilità sintattica, grazie alla codifica in formato JSON. Saranno successivamente adeguati a livello semantico, mediante l'uso di vocabolari controllati, di standard e di modelli di dati aperti e condivisi tra gli appartenenti alla comunità di ricerca, in particolare il modello semantico CRMpe, sviluppato all'interno di PARTHENOS (12).

Questo consentirà al DMP di avvicinarsi ulteriormente alle caratteristiche dei machine-actionable DMP e offrirà ai ricercatori la possibilità di trarre beneficio dalla condivisione delle informazioni.

3. Conclusioni

Il modello DMP di PARTHENOS si basa su quello di H2020, con l'aggiunta di un utile supporto nella compilazione del documento grazie alle linee guida e ai riferimenti incrociati creati con i contenuti degli altri strumenti sviluppati dal progetto (ad es. Wizard e SSK). Il lavoro condotto da PARTHENOS anticipa quelle che sono le raccomandazioni e le azioni promosse dal "European Commission Expert Group on FAIR data" (13) fornendo uno strumento di supporto alla compilazione del documento.

Uno strumento online realizzato dal PIN all'interno del progetto è attualmente in fase di sviluppo per essere incorporato nel sito web di PARTHENOS per fornire ai ricercatori un tool semplice per creare il proprio DMP. Lo strumento attualmente consente la compilazione del modulo, grazie a un'interfaccia intuitiva e la possibilità di scaricare una copia del documento in PDF e in JSON.

Altre attività in previsione per il prossimo futuro includono la traduzione del modello DMP di PARTHENOS in diverse lingue per fornire versioni nazionali a quei paesi che non hanno ancora sviluppato un proprio modello (ad es. italiano, spagnolo, greco, tedesco, ecc.).

Al momento, nella comunità di PARTHENOS solo il modello archeologico è stato sviluppato e testato. In futuro, altre comunità di ricercatori delle discipline umanistiche, che sono disposte a contribuire con le loro pratiche quotidiane, saranno coinvolte nella

definizione di linee guida mirate.

Al fine di promuovere e diffondere il modello DMP di PARTHENOS e le sue linee guida alle comunità di ricerca archeologiche, si prevede l'organizzazione di seminari di formazione e servizi di consulenza per garantire una coerente diffusione dei risultati di PARTHENOS, allo scopo di aumentare la consapevolezza sulla ricerca aperta in archeologia.

Riferimenti bibliografici

- (1) European Commission, Research and Innovation, Open Science, <https://ec.europa.eu/research/openscience/index.cfm> (ultima consultazione 8/12/2018)
- (2) FORCE 11, Guiding Principles for Findable, Accessible, Interoperable and Reusable Data Publishing Version B1.0, <https://www.force11.org/fairprinciples> (ultima consultazione 8/12/2018)
- (3) M. D. Wilkinson, M. Dumontier et al., 2016, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3:160018, <https://doi.org/10.1038/sdata.2016.18> (ultima consultazione 8/12/2018)
- (4) Data Management Template in Horizon 2020
http://ec.europa.eu/research/participants/docs/h2020-fundingguide/cross-cutting-issues/open-access-data-management/datamanagement_en.htm#A1-template (ultima consultazione 8/12/2018)
- (5) S. Bassett, S. Di Giorgio, et al., 2017, A DMP template for Digital Humanities: the PARTHENOS model, pp. 74-77, Proceedings of GARR Conference "The Data Way to Science", Venezia, 15-17 Novembre
- (6) Data Curation Centre, <https://dmponline.dcc.ac.uk/> (ultima consultazione 8/12/2018)
- (7) PARTHENOS (Pooling Activities, Resources and Tools for HeritageE-research Networking, Optimization and Synergies) <https://www.parthenos-project.eu> (ultima consultazione 8/12/2018)
- (8) ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) www.ariadne-infrastructure.eu (ultima consultazione 8/12/2018)
- (9) ARIADNE Portal, <http://portal.ariadne-infrastructure.eu/>
- (10) M. Grootveld, E. Leenarts, et al., 2018, OpenAIRE and FAIR Data Expert Group survey about Horizon 2020 template for Data Management Plans (Version 1.0.0 Data set). Zenodo. <http://doi.org/10.5281/zenodo.1120245> (ultima consultazione 8/12/2018)
- (11) T. Miksa, S. Simms, et al., Ten simple rules for machine-actionable data management plans, <https://zenodo.org/record/1172673#.XBOIOGhKiUk> (ultima consultazione 8/12/2018)
- (12) G. Bruseker, M. Doerr, et al., Report on the common semantic framework, PARTHENOS project deliverable 5.1, 2017 http://www.parthenosproject.eu/Download/Deliverables/D5.1_Common_Semantic_Framework_Appendices.pdf (ultima consultazione 8/12/2018)
- (13) S. Hodson, et al., 2018, FAIR Data Action Plan: Interim recommendations and

actions from the European Commission Expert Group on FAIR data (Version Interim draft). <http://doi.org/10.5281/zenodo.1285290> (ultima consultazione 8/12/2018)

Autori



Sara Di Giorgio - sara.digiorgio@beniculturali.it

Collabora dal 2004 con l'Istituto centrale per il Catalogo Unico delle biblioteche italiane del Ministero dei beni e delle attività culturali, per lo sviluppo di CulturalItalia, il portale della cultura italiana del MiBACT, e di numerosi progetti europei per la digitalizzazione, la fruizione e la conservazione a lungo termine del patrimonio culturale digitale. Partecipa al progetto Parthenos, coordinando le attività per la definizione dei requisiti per lo sviluppo di policy sul data lifecycle. Coordina i Task 'Managing FAIRness of archaeological data and IPR' e 'Innovating services with industries' nell'ambito del progetto ARIADNE Plus. È membro del Council di Europeana, la digital library dell'Unione Europea.

Paola Ronzino - p.ronzino@gmail.com

Paola Ronzino, è una ricercatrice post-doc presso il PIN. Ha conseguito un Master in Conservazione dei Beni Culturali e un Dottorato in Science and Technology for Cultural Heritage. I suoi interessi di ricerca riguardano lo sviluppo di ontologie e standard di metadati per la documentazione archeologica e il patrimonio culturale, con interesse specifico per la documentazione degli edifici archeologici. Ha partecipato a diversi progetti europei coordinati dal PIN sul patrimonio culturale digitale, ed è stata attivamente coinvolta nelle attività di ricerca del progetto ARIADNE. Attualmente, partecipa alle attività del progetto PARTHENOS, in particolare nella definizione delle esigenze degli utenti e l'elaborazione del Data Management Plan per le discipline umanistiche. È Project Manager del progetto ARIADNEplus all'interno del quale si occupa delle attività relative all'integrazione di dataset geografici.

