



Università  
Ca' Foscari  
Venezia

Corso di Laurea Magistrale  
in Storia e gestione del patrimonio  
archivistico e bibliografico  
ordinamento ex D.M. 270/2004

Tesi di Laurea

# Digital Ecosystems for Open Science

Il caso Università di Vienna, l'esempio di Phaidra

**Relatore**

Ch.ma Prof.ssa Dorit Raines

**Correlatore**

Ch. Prof. Riccardo Ridi

**Laureando**

Davide Montanaro  
Matricola 855470

**Anno Accademico**

2016 / 2017



## **Indice:**

<b>Introduzione</b>	<b>5</b>
<b>Capitolo 1 – Un ecosistema digitale</b>	<b>9</b>
1.1 Il progetto e la sua genesi	9
1.2 Phaidra, processi ed infrastrutture	12
1.3 I destinatari	19
1.4 I servizi offerti da Phaidra	22
<b>Capitolo 2 - Modello di digital workflow e soggetti coinvolti</b>	<b>25</b>
2.1 Il produttore di dati	27
2.2 Le principali entità che offrono i servizi	30
A. La biblioteca	30
B. Il centro di calcolo	33
C. I servizi legali	35
<b>Capitolo 3 - La struttura interna di Phaidra</b>	<b>39</b>
3.1 La gestione del sistema	41
3.2 L'architettura del sistema	46
3.3 La gestione elettronica dei processi lavorativi	53
<b>Capitolo 4 – Il digital workflow: il documento e le fasi di lavoro</b>	<b>61</b>
4.1 La preparazione (pre-ingest)	63
4.2 Il caricamento (ingest)	68
4.3 L'archiviazione e la gestione (management)	73
4.4 L'estrazione e la visualizzazione (visualization and re-use)	78
<b>Capitolo 5 - La gestione delle immagini: presupposti e formati presi in esame</b>	<b>83</b>
5.1 I presupposti (FAIR Principles)	84
5.2 Le immagini	88
5.3 Le mappe geografiche	100
5.4 I libri	103
Appendice “immagini che si muovono”	106
<b>Conclusioni</b>	<b>109</b>
<b>Bibliografia</b>	<b>113</b>



## Introduzione

Sfogliando la corrispondenza tra gli umanisti Poggio Bracciolini e Francesco Barbaro si può notare come, il 6 luglio del 1417, quest'ultimo ringrazia il primo per le importanti ricerche con il quale aveva contribuito alla “*Respublica literaria*”. In quel periodo infatti, Poggio Bracciolini era segretario apostolico al concilio di Costanza (1414-1418) e approfittando dell'occasione si era dedicato a importanti ricerche nel campo della filologia. Sono molto affascinato da questo aneddoto in quanto è la prima volta che appare il concetto di “*repubblica delle lettere*”, espressione poi che si diffonderà e prenderà piede in modo particolare durante l'illuminismo. Credo riesca a trasmettere perfettamente l'ideale di condivisione e apertura del sapere il quale, tramite il contributo di tutta la comunità intellettuale, spinge verso la ricerca della verità e il bene comune. È apprezzabile il valore universale che fa da cardine a questa repubblica virtuale dove, non contando le divisioni politiche, ideologiche o religiose, l'unica battaglia consentita è quella contro chi vuole limitare la libertà di partecipazione e discussione.

Rivedendo questo concetto in una chiave più moderna è chiaro come attualmente, grazie all'evoluzione tecnologica, si sono aperte numerose nuove possibilità di condivisione del sapere. L'essere facilmente connessi, tramite la rete, a persone che vivono dall'altra parte del mondo rende quasi banale la comunicazione e la richiesta di informazioni. Ciò che spesso manca però è la mediazione tra chi detiene una determinata risorsa e chi ne vorrebbe fruire. A tappare questo buco ci stanno pensando le biblioteche che, fedeli custodi della conoscenza ormai da millenni, si stanno reinventando anche tramite interfacce elettroniche. Ci sono varie problematiche che il personale bibliotecario si trova ad affrontare ogni giorno, come la cronica carenza di spazio per i volumi, l'aumentata sensibilità per le tematiche ambientali o il mantenersi all'avanguardia per dare un servizio all'altezza agli utenti. Siamo ad un punto di transizione in cui queste istituzioni creano sempre più spesso delle piattaforme informative per facilitare la gestione. Queste hanno svariati compiti che puntano diretti ai due principali obiettivi delle biblioteche: la preservazione e la libera fruizione del proprio patrimonio, finalità che purtroppo a volte si scontrano tra loro. Per far sì che la fruizione diventi sempre più proficua si è puntato sempre di più a far risparmiare il tempo allo studioso con delle interessanti novità: dal rendere accessibili le proprie risorse da qualsiasi postazione alla possibilità di condividerle. Lo stesso utente, se opportunamente validato, in certi casi può essere colui che carica i contenuti.

Fasce sempre più ampie e disparate della società sono ogni giorno alla ricerca di nuove fonti e informazioni: non solo ricercatori ma studenti, educatori, politici, giornalisti, organizzazioni pubbliche e private, fino ad arrivare ai semplici cittadini interessati. Ciò porta a considerare il fatto che i processi di produzione, organizzazione, diffusione e mantenimento della memoria siano sempre

più complessi. A livello europeo si sta puntando molto sulla gestione condivisa delle informazioni e non viene nascosto quanto sia fondamentale esserne consapevoli. *“Because research in genomics, pharmacology or the fight against cancer increasingly depends on the availability and sophisticated analysis of large data sets. Sharing such data means researchers can collaborate, compare, and creatively explore whole new realms. We cannot afford for access to scientific knowledge to become a luxury, and the results of publicly funded research in particular should be spread as widely as possible.”* (Neelie Kroes, Vice President European Commissioner responsible for the Digital Agenda)<sup>1</sup>. Il collegamento di tutte le risorse presenti nelle biblioteche europee darebbe sicuramente risultati sorprendenti, su una via che appare tracciata. Un sistema intuitivo, di facile utilizzo e aperto a chiunque, che si adatti facilmente ai continui cambiamenti richiesti dalla scienza e dalla ricerca. Anche se le incognite sono sempre infinite, a partire dalla quantità di fondi, la presenza di personale qualificato, l'impegno e la passione per portare avanti un progetto condiviso.

Ho avuto la fortunata possibilità, grazie al progetto Erasmus, di svolgere questo mio secondo anno di laurea magistrale presso l'università di Vienna. Essendo in loco ho potuto studiare più da vicino il sistema Phaidra, ideato e sviluppato a partire dalla biblioteca d'ateneo in collaborazione con il centro di calcolo. In quanto piattaforma che ha come obiettivo la conservazione a lungo termine dei documenti digitali e la condivisione dei dati di ricerca, si inserisce all'interno di un ecosistema che racchiude tutto ciò che ruota attorno all'Università viennese. Attualmente si inserisce nell'orbita dell'European Open Science Cloud, meta dichiarata da parte dell'apposita commissione creata dell'Unione Europea e proprio per questo può offrire la sua esperienza decennale nel campo ed entrare in un'ottica di interoperabilità e di condivisione con altre piattaforme a livello continentale. Per via della sua vocazione internazionale Phaidra è stato esportato anche in altre università europee come Ca' Foscari e l'Università di Padova. Nato nel 2008 proprio all'Università di Vienna è stato sviluppato con occhio di riguardo per la ricerca, l'insegnamento e l'amministrazione accademica. Studenti e docenti possono liberamente caricare quanto ottenuto tramite gli sforzi delle ricerche. Questi risultati grazie alle peculiarità del software, sono messi a disposizione in tutto il mondo con varie possibilità di ricerca. Si tratta di più tipologie di dati, che spaziano dai testi ai filmati, dalle immagini ai file audio; dotati di un collegamento permanente, in modo che sia sempre citabile. Gli oggetti, nel caso siano inerenti ad una stessa categoria, possono anche essere associati in collezioni creando una sorta di mini portale tematico.

Grazie al materiale e al prezioso aiuto che mi è stato offerto dai gentilissimi responsabili del progetto,

---

<sup>1</sup> *Opening up Europe: from Common Standards to Open Data*, OpenForum Europe Summit 2011, Brussels. Disponibile a: < [http://europa.eu/rapid/press-release\\_SPEECH-11-596\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-11-596_en.htm) > [Ultimo accesso 25/09/2017]

Paolo Budroni e Susanne Blumesberger, il percorso alla scoperta della piattaforma è stato ben guidato e mi ha condotto fruttuosamente alla stesura della presente tesi di laurea. Volgendo lo sguardo ai contenuti e procedendo per gradi, la mia intenzione è quella di partire con un'introduzione al mondo di Phaidra scoprendo cosa può offrire questo sistema, a partire dai servizi e dai soggetti che vengono coinvolti. A seguire non potrà mancare un'analisi della struttura interna del software passando attraverso la sua gestione e la sua architettura; il tutto grazie anche all'ausilio di workflows della biblioteca universitaria. Nella parte centrale l'obiettivo è procedere osservando come il documento digitale, a partire dalla sua creazione, venga caricato, gestito e infine visualizzato tramite interfaccia. Negli ultimi due capitoli si entrerà più nel vivo della ricerca: dalle fasi di lavoro che una risorsa da caricare attraversa, alla comprensione della tipologia di oggetto digitale di cui voglio parlare più specificatamente, ossia le immagini digitali, fino ad arrivare alle conclusioni, dove l'attenzione si focalizzerà sull'uso della piattaforma presso il mio Ateneo. L'obiettivo sarà quello di coglierne le peculiarità in relazione ai concetti precedentemente sviluppati, quali politiche, governance, servizi, dati e infrastruttura.





## Capitolo 1 – Un ecosistema digitale

### 1.1 Il progetto e la sua genesi:

A partire da una breve introduzione su “cos'è un sistema eco-digitale”, in questa sezione si tratterà di esporre le linee guida del progetto Phaidra: la sua nascita, la sua filosofia, gli obiettivi e la sua evoluzione a pari passo con quella della tecnologia.

*“Encourage the development of e-infrastructures as a service, by making sure that the various ICT components are aligned and provide a joint service. This should be done by building on the success of existing systems, while reducing current fragmentation through the creation of an ECOSYSTEM of infrastructures.”<sup>2</sup>*

Questa citazione è tratta dal documento “Amsterdam Call for Action on Open Science”, il quale sintetizza i contenuti espressi durante l'omonima conferenza nella capitale olandese il 5-6 aprile 2016. In maniera diretta e schematica viene illustrata, da parte degli esperti europei, l'attuale evoluzione a livello teorico e pratico dell'idea di scienza aperta, rispondendo con delle ipotetiche soluzioni ad alcune problematiche, come quelle in questo caso relative allo sviluppo di infrastrutture comuni. L'interrogativo che viene posto riguarda le nuove modalità di analisi e comunicazione scientifica, le quali attualmente necessitano di un ambiente sicuro e intuitivo per gestire e sfruttare senza perdita l'imponente quantità di dati potenzialmente utili. La mia intenzione è di introdurre, tramite queste prime righe, il concetto di *ecosistema* menzionato nella citazione qui sopra. È uno dei tanti casi in cui un'espressione, presa da un'altra disciplina, trova un nuovo utilizzo in un contesto diverso ma per qualche aspetto assimilabile. Se pensiamo alla biologia, nominando un ecosistema ci viene naturale pensare ad un determinato ambiente fisico, dentro il quale si svolgono interazioni, flussi, scambi; il tutto in un equilibrio dinamico che evolve continuamente. Ci sono molti elementi che vanno a comporre un'unica entità, ognuno con un suo ruolo, una sua funzione, un suo significato. Allo stesso tempo, un ecosistema digitale vuole proporsi come un ambiente che offre strumenti e procedure per un'efficace gestione quotidiana dei dati caricati dai soggetti preposti. Questi ultimi, all'interno dell'ecosistema, interagiscono tra di loro in un contesto sicuro e conosciuto scambiandosi informazioni e creando un clima favorevole di arricchimento in termini di conoscenza, abilità e contatti.

Nel primo report “Realising the European Open Science Cloud” (2016), realizzato dalla commissione “High Level Expert Group” dell'Unione Europea, vengono esposti degli elementi chiave per un effettivo sviluppo del progetto EOSC, il quale “needs to be developed as a data infrastructure commons, that is an eco-system of infrastructures”. Risulta evidente come le vecchie piattaforme,

---

<sup>2</sup> AMSTERDAM CALL FOR ACTION ON OPEN SCIENCE, basato sulla conferenza *Open Science – From Vision to Action*, tenutasi ad Amsterdam il 4 - 5 aprile 2016. Disponibile a:  
<<http://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf>>  
[Ultimo accesso 12/09/2017]

singolarmente, non possano supportare adeguatamente questa mole crescente di dati, sia per problemi legati allo spazio, sia per questioni di interconnessione: “A complex eco-system of infrastructures: it may seem counter-intuitive but the challenges of ever bigger data can no longer be solved only by ever bigger infrastructure. Next to advanced computer science, which will bring innovative computing and storage and new advanced algorithms for knowledge extraction from data, we need fundamentally to rethink infrastructure as we know it.”<sup>3</sup>

Spesso i dati crescono addirittura più velocemente di quanti ne siano trasferibili, o semplicemente sono troppo pesanti per essere spostati. In altri casi il problema riguarda invece la licenza non troppo flessibile, la quale non permette lo spostamento in ambienti diversi dall'originale.

La via tracciata è quella dove un'infrastruttura non contiene tutti i dati ma, essendo questa collegata in maniera efficiente a tante altre infrastrutture affini e avendo tutte assieme un pacchetto applicativo in grado di gestire i collegamenti, offre infinite possibilità di ricerca con grande beneficio per lo spazio di archiviazione.

Phaidra<sup>4</sup>, il progetto che si pone al centro di questo studio, va inserito nel contesto di un ecosistema digitale che fa riferimento all'Ateneo che gli ha dato la luce, ossia l'università di Vienna. Questo ecosistema universitario si occupa per l'appunto di tutte le categorie di dati digitali che rientrano nell'orbita di questo genere di istituzione, come potrebbero essere libri antichi e moderni, manoscritti, fotografie, mappe, learning objects, filmati e altro ancora. Nonostante apparentemente possa essere legato esclusivamente ad un ambito bibliotecario e archivistico, in realtà si rifà ad altri campi accademici quali la ricerca scientifica. Vediamo ora di cosa si tratta, partendo dalla scelta del nome che ricalca l'omonimo mito di origine greco antica. *Permanent Hosting, Archiving and Indexing of Digital Resources and Assets*, questo l'acronimo derivato dal nome che già lascia intendere lo scopo per cui questa piattaforma è stata pensata e realizzata nel 2008. Nata tramite la collaborazione di biblioteca e centro di calcolo (ZID – Zentraler Informatikdienst) universitari e basata sul software open source Fedora, si è posta sin da subito chiari obiettivi strettamente connessi con l'esponenziale sviluppo che l'evoluzione tecnologica ha portato al mondo della conoscenza. Questi puntano ad offrire alle categorie interessate (insegnanti, studenti, ricercatori, amministrazione) la possibilità di memorizzare, documentare e archiviare in maniera permanente le loro pubblicazioni rendendole disponibili a livello globale attraverso la rete Internet. A tale scopo, il software si occupa di archiviare e rendere al contempo fruibili le risorse con caratteristiche diverse in base al contenuto delle stesse. I

---

<sup>3</sup> EUROPEAN COMMISSION, *Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud*, 2016. Disponibile a: <<https://ec.europa.eu/digital-single-market/en/news/first-report-high-level-expert-group-european-open-science-cloud>> [Ultimo accesso 12/09/2017]

<sup>4</sup> <https://phaidra.univie.ac.at/> [Ultimo accesso 25/09/2017]

processi in questione sono pensati per categorizzare i Digital Objects con parole chiave, descrizioni, abstracts, thesauri, comprendendo al proprio interno sia dati digitalizzati, sia dati nativi digitali. Questi possono anche essere creati direttamente con un apposito strumento offerto dalla piattaforma. Una volta caricati e resi disponibili, i files possono essere localizzati ovunque e, compatibilmente col tipo di licenza, scaricati. La licenza viene scelta dall'utente che si occupa dell'upload, il quale deve essere consapevole della normativa vigente onde evitare infrazioni di copyright.

Uno degli elementi chiave, che rende Phaidra una piattaforma valida nel perseguire gli scopi sopra indicati, sta nel fatto che ogni oggetto possiede un link permanente e un identificatore persistente. Questi ne consentono la citabilità, che altrimenti è messa costantemente a repentaglio dai continui cambiamenti che subiscono le pagine web più comuni. Inoltre ogni risorsa archiviata viene opportunamente descritta tramite specifici metadati i quali, con le loro peculiarità, rendono il file trovabile tramite i motori di ricerca e relazionato ad altri file con una qualsiasi affinità. Senza dimenticare che c'è la possibilità di legare tra loro i documenti in collezioni digitali. Questo tipo di approccio relazionale permette di favorire l'interdisciplinarietà, promuovendo la ricerca anche tramite l'apporto di contributi altrimenti poco considerati.

Al fine di mantenere viva la vocazione globale del progetto, è stata promossa una politica di interoperabilità con altre piattaforme tra cui spiccano progetti europei della portata di Europeana<sup>5</sup>, OpenAIRE<sup>6</sup> e OpenAIREplus. Infatti, come accennato in apertura di capitolo, c'è molto fermento a livello europeo per quanto riguarda le opportunità che l'applicazione di un "European Open Science Cloud" potrebbero offrire allo sviluppo. Phaidra, inserendosi in questo contesto, può fornire il suo contributo tramite la sua esperienza ormai decennale rispondendo positivamente alle linee guida indicate dall'UE. Tra queste si può segnalare ad esempio l'uso dei *principi FAIR* (Findable, Accessible, Interoperable, Reusable)<sup>7</sup> o del *Data Management Plan*<sup>8</sup>, di cui si parlerà approfonditamente a seguire. Importante inoltre l'approccio multilinguale della piattaforma, al momento disponibile in italiano, tedesco, inglese e serbo.

Attualmente si può dire che Phaidra dal 2008 abbia compiuto notevoli passi avanti, uscendo dai confini dell'Ateneo viennese e venendo implementata in numerosi istituti universitari europei (in Austria, Italia, Serbia, Montenegro e Bosnia-Herzegovina) tra i quali, nel 2014, l'università Ca' Foscari di Venezia dove sto compiendo il mio percorso di studi.

---

<sup>5</sup> <http://www.europeana.eu/> [Ultimo accesso 12/09/2017]

<sup>6</sup> <https://www.openaire.eu/> [Ultimo accesso 12/09/2017]

<sup>7</sup> <https://www.force11.org/group/fairgroup/fairprinciples> [Ultimo accesso 12/09/2017]

<sup>8</sup> MIKSA T., SCHRAUF C., SANCHEZ SOLÍS B., *Data Management Plans (DMP)*, 2016. Disponibile a: <<http://phaidra.univie.ac.at/o:441308>> [Ultimo accesso 12/09/2017]

### 1.2 Phaidra, processi ed infrastrutture:

L'obiettivo che l'Università di Vienna si è data a livello ideale e per il quale è nata una piattaforma come Phaidra, è stato quello di fornire ai membri dell'ateneo di ogni livello (ricerca, studio, insegnamento e amministrazione) l'opportunità di salvare in sicurezza i propri documenti e renderli disponibili in tutto il mondo tramite i propri servizi; creare quindi, un sistema solido in grado di ospitare Oggetti Digitali (Digital Objects). Le fondamenta, concettualmente parlando, si possono riassumere attraverso le idee di *Open Science*, *Open Data* e *Open Access*, che convergono in un semplice repository per gestire i dati. Vale la pena spendere due parole per comprendere al meglio cosa si intende con il termine repository, il quale è entrato a far parte in maniera sempre più ricorrente del lessico nel campo della gestione documentale<sup>9</sup>. Letteralmente si può tradurre con deposito o ripostiglio, ma in questo caso archivio potrebbe essere la parola più opportuna<sup>10</sup>. Infatti non si tratta di altro che di un contenitore che raccoglie dati e informazioni digitali, archiviati e resi di facile individuazione tramite l'uso di metadati e di tabelle relazionali che agevolano il lavoro dei motori di ricerca. Queste sue peculiarità permettono ai repositories di accogliere e gestire efficientemente anche imponenti volumi di dati che, nelle fasi di lavoro che vanno dal caricamento (ingest) all'estrazione e visualizzazione (visualization and re-use), seguono tutte una serie di regole che vedremo nel corso della tesi.

Sin dal primo momento, la libertà di accesso è stata una delle motivazioni più importanti che ha accompagnato lo sviluppo di Phaidra e per questo tutti i membri dell'università, inclusi gli studenti, sono stati invitati ad utilizzare il repository. La tecnologia è stata inoltre fornita ad altre università ed istituzioni di ricerca in cinque stati diversi<sup>11</sup>, andando a creare quello che viene chiamato “Phaidra Network”. Dopo una prima fase di utilizzo e con la crescita del numero di utenti della piattaforma, risultò piuttosto evidente come il sistema dovesse essere più flessibile ed aperto a ulteriori esigenze. Per questo motivo ci fu una fase in cui venne ripensata l'intera impostazione del programma, tramite un processo di riprogettazione. Fondamentale risultò la comunicazione con i beneficiari del sistema e il confronto tra un ampio range di dati caricati e di casi d'uso. L'obiettivo divenne chiaro, ossia indirizzare quanto più possibile i servizi ai bisogni degli utilizzatori. Due furono le decisioni prese

---

<sup>9</sup> Vedi anche: Confederation of Open Access Repositories (COAR). Homepage: <https://www.coar-repositories.org/> [Ultimo accesso: 25/09/2017]

<sup>10</sup> È interessante a riguardo anche la novità concettuale relativa al termine “documento”, il quale oggi può essere qualsiasi oggetto digitale; si tratta di una distinzione ancora non del tutto seguita nel nostro linguaggio dove usiamo “documento” per il settore archivistico, “testo” o “libro” per il settore bibliografico e dove il repository equivale nella mente delle persone ad una biblioteca e non un archivio. Vedi anche: RIDI R., *Il mondo dei documenti : cosa sono, come valutarli e organizzarli*, Roma ; Bari : GLF editori Laterza, 2010.

<sup>11</sup> AUSTRIA: FH St. Pölten, Fonds zur Förderung wissen. Forschung, Kunstuniversität Linz, Landesbibliothek Vorarlberg, Universität für angewandte Kunst Wien, Universität für Musik und darstellende Kunst Graz, Universität Wien. BOSNIA ED ERZEGOVINA: Univerzitet u Banjoj Luci, Univerzitet u Sarajevu, Univerzitet u Tuzli. ITALIA: Università Ca' Foscari, Università Iuav di Venezia, Università di Padova. MONTENEGRO: Univerzitet Crne Gore. SERBIA: Univerzitet u Beogradu, Univerzitet u Kragujevcu, Univerzitet u Nišu. Fonte: <http://phaidra.org/community/phaidra-partners/> [Ultimo accesso: 25/09/2017]

per raggiungere lo scopo: per prima quella di riconsiderare la struttura tecnica dei repositories verso una micro-services architecture e, successivamente, furono ripensati i modelli per la gestione dei dati in modo da renderli utilizzabili per diversi casi d'uso.<sup>12</sup>

Basata su software open source Fedora<sup>13</sup>, Phaidra rappresenta la base tecnologica per l'implementazione delle politiche relative all'Open Access dell'università di Vienna, in linea con la “*Berlin Declaration on Open Access to Scientific Knowledge*”<sup>14</sup> firmata nel 2010 da Georg Winckler, allora rettore dell'Ateneo viennese. Questa dichiarazione, firmata da circa 300 istituzioni, biblioteche, archivi, musei e governi da tutto il mondo, ha come obiettivo principale quello di formare un fronte per sfruttare al meglio internet come risorsa per diffondere la conoscenza. L'impegno è quindi quello di promuovere politiche per un web che sia “sostenibile, interattivo e trasparente. I contenuti ed i mezzi di fruizione dovranno essere compatibili e ad accesso aperto.”

Fondamentale ai fini della comprensione del profilo della piattaforma sta il modello di riferimento OAIS (Open Archival Information System)<sup>15</sup>, su cui è oggi basato gran parte del mondo degli archivi e dei depositi digitali. Pubblicato nel 2002 e aggiornato nel 2014, ha un approccio piuttosto generale, infatti non si occupa né di architetture (le quali possono essere molto diverse), né di costi, bensì di concetti e principi riguardanti la terminologia e le funzionalità richieste in questo ambito.

Attualmente, lo sviluppo del software ha luogo nelle comunità di ricerca, dove è in atto una continua sfida che riguarda le infrastrutture digitali e il loro coordinamento. Ma, nel momento in cui viene sviluppato un software da un gruppo di ricerca, il quale ha tra gli obiettivi la preservazione dei dati digitali, può esso essere mantenuto dopo la fine del progetto? Ed eventualmente come? Questa sfida mostra chiaramente come il bisogno di consultazione tra ricercatori, dovrebbe avere luogo sin dall'inizio dei lavori.

Un grande cambiamento concettuale è stato quello avvenuto nel passaggio da semplice repository a Eco-Sistema digitale<sup>16</sup>. Si è passati dunque da un semplice archivio, utile per salvare i dati della

---

<sup>12</sup> GANGULY R., BUDRONI P., SÁNCHEZ SOLÍS B., "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud." In *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*. ELPUB. Limassol, Cyprus, 2017, pp. 203-210.

<sup>13</sup> Vedi anche: <https://getfedora.org/it/> [Ultimo accesso: 25/09/2017] e [https://it.wikipedia.org/wiki/Fedora\\_\(informatica\)](https://it.wikipedia.org/wiki/Fedora_(informatica)) [Ultimo accesso 25/09/2017]

<sup>14</sup> <https://openaccess.mpg.de/Berlin-Declaration> [Ultimo accesso 25/09/2017]

<sup>15</sup> THE CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM PRACTICES, *Reference Model for an Open Archival Information System (OAIS)*. Magenta Book, 2012. Disponibile a: <<https://public.ccsds.org/pubs/650x0m2.pdf>> [Ultimo accesso 12/09/2017]. Vedi anche: LAVOIE B., *Open Archival Information System (OAIS) Reference Model, 2*, (Traduzione italiana a cura di F. Cancedda, C. Consonni, A. Galeffi, A. Marchitelli, L. Sardo, con il coordinamento scientifico di Paul Gabriele Weston); Roma: Associazione italiana biblioteche, 2016.

<sup>16</sup> GANGULY R., BUDRONI P., SÁNCHEZ SOLÍS B., "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud." In *Expanding Perspectives on Open Science: Communities,*

propria istituzione, ad una struttura più moderna funzionante su più livelli: Governance, Service e Data (contenente il repository). La figura 1 mostra il caso Phaidra e come questo sviluppo porti a nuove possibilità su più fronti, le quali verranno affrontate dettagliatamente nei prossimi capitoli. Per quanto riguarda le prospettive future, il piano è quello di integrare un servizio per i DMPs (Data Management Plans) basato sul *DMP Online Tool*<sup>17</sup> del DCC (Digital Curation Center, presso l'università di Edimburgo) e sulle raccomandazioni a riguardo di RDA<sup>18</sup> (Research Data Alliance). Questi DMPs sono forniti in modo che siano leggibili sia dagli umani che dalle macchine.

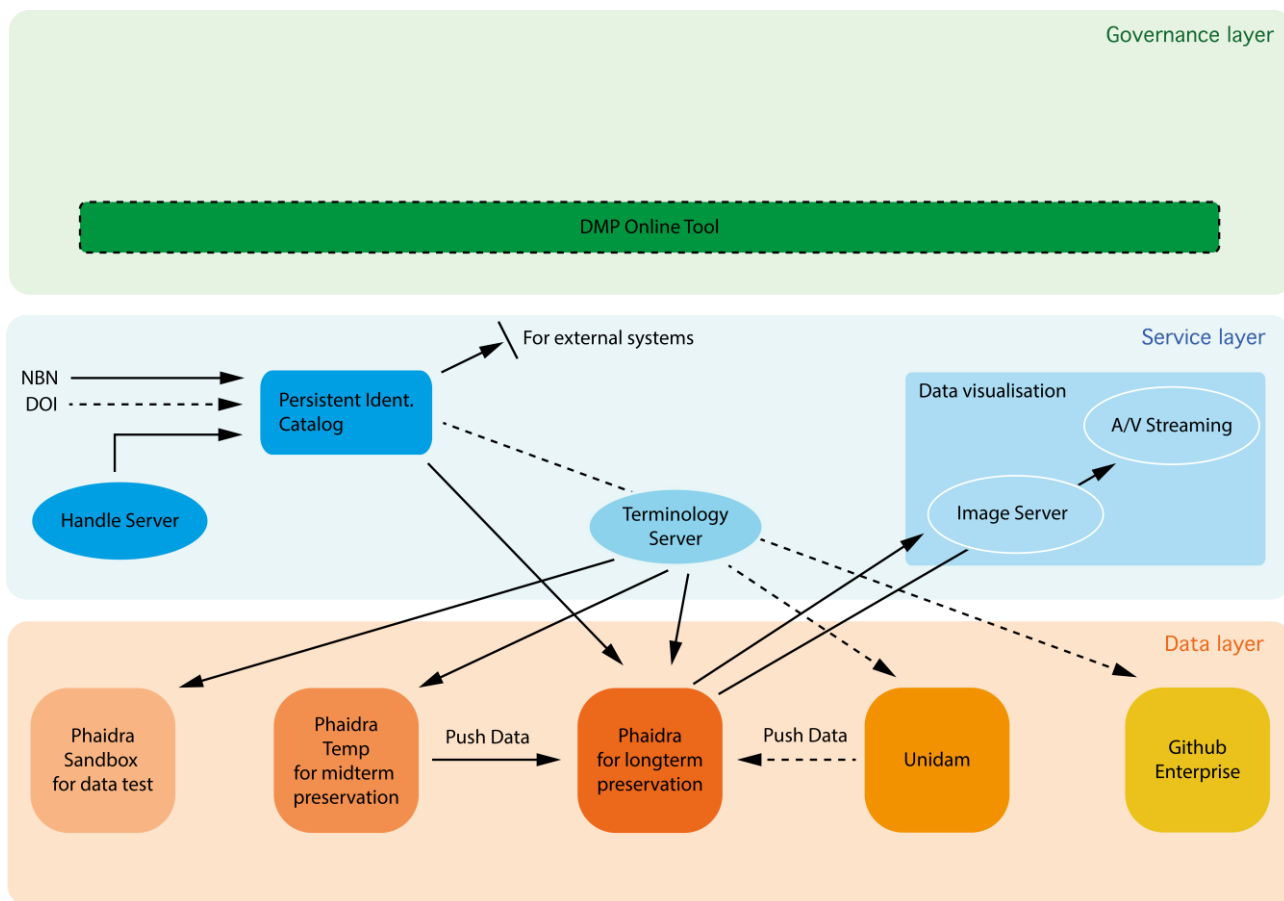


Figura 1- Ecosystem of Phaidra

CC BY 4.0 Raman Ganguly. <http://phaidra.univie.ac.at/o:527326>

Nella fase di creazione di un Eco-sistema digitale va tenuto conto del fatto che, non solo è necessario fornire una buona infrastruttura funzionale e funzionante, ma questa deve essere connessa con la comunità di ricerca e deve mantenere collegamenti con gli altri progetti affini (nel caso di Phaidra potremmo fare l'esempio di OpenAire ed Europeana) e rimanere all'interno delle linee guida dell'European Open Science Cloud (EOSC). Uno dei rischi più concreti da scongiurare nello sviluppo

*Cultures and Diversity in Concepts and Practices*. ELPUB. Limassol, Cyprus, 2017, pp. 203-210.

<sup>17</sup> <https://dmponline.dcc.ac.uk/> [Ultimo accesso 25/09/2017]

<sup>18</sup> <https://www.rd-alliance.org/> [Ultimo accesso 25/09/2017]

di un'idea a visione globale come questa, è quello di ragionare senza una piena consapevolezza dei mezzi e delle possibilità offerte dalla tecnologia, abbandonandosi alla fretta e/o all'approssimazione. Questo potrebbe portare alla creazione di un sistema che nasce già vecchio o mal funzionante, con tutto ciò che ne deriva. È necessario quindi sottolineare l'importanza della condivisione e della collaborazione tra esperti europei del campo per rimanere sempre aggiornati e uniti sul fronte comune. Come già anticipato in queste prime pagine, Phaidra e l'Università di Vienna rientrano all'interno di piani futuri molto più ampi; ma ovviamente non sono le uniche entità a cui si fa riferimento nel momento in cui si pensa a piattaforme per l'archiviazione dei dati digitali usate nei progetti europei, nonostante la piattaforma venga usata in altre università austriache ed europee. È bene dunque, ai fini della comprensione globale dell'elaborato, spiegare in due parole altre iniziative correlate su diversi piani; queste possono essere dei partner paritari o magari progetti superiori volti a stabilire una certa uniformità a livello di ideali, di processi, di strutture. Ecco a seguire un elenco sintetico di elementi che rientrano in questa sfera:

- ◆ **Kulturpool<sup>19</sup>**: Si tratta di un progetto che offre un accesso centrale alle risorse del patrimonio culturale austriaco digitale e fa riferimento anche ad oggetti archiviati in Phaidra. Tra i destinatari di questi oggetti ci sono anche le scuole.
- ◆ **OpenAIRE (Open Access Infrastructure for Research in Europe)<sup>20</sup>**: Insieme ad altri 37 partner europei, la biblioteca dell'Università di Vienna è coinvolta in un progetto pionieristico, attraverso il quale si punta allo sviluppo di stretti rapporti non solo con altre *open access repositories*, ma anche con editori e progetti che riguardano l'interdisciplinarietà in Europa. Il progetto OpenAIRE sostiene la Commissione europea nell'attuazione di un programma pilota di accesso aperto nell'ambito del settimo *Research Framework Program*. Il suo obiettivo è quello di realizzare una piattaforma in grado rendere disponibili gli studi in corso.
- ◆ **EUROPEANA LIBRARIES<sup>21</sup>**: L'Università di Vienna è una delle 26 istituzioni partecipanti a questo progetto, il quale ha lo scopo di fornire preziosi contributi culturali sotto forma di oggetti digitali; questi provengono da vari enti culturali come musei e archivi sparsi sul territorio europeo e che sono destinati ad aumentare nel prossimo futuro.
- ◆ **Tempus<sup>22</sup>**: Nel contesto del programma dell'Unione Europea chiamato *Tempus IV (2007-2013)*, che sostiene la modernizzazione dell'istruzione superiore in alcuni paesi est europei, dell'Asia centrale, dei Balcani orientali e dell'area mediterranea principalmente tramite la

---

<sup>19</sup> <http://kulturpool.at/display/kupo/Home> [Ultimo accesso 25/09/2017]

<sup>20</sup> <https://www.openaire.eu/> [Ultimo accesso 25/09/2017]

<sup>21</sup> <http://pro.europeana.eu/project/europeana-libraries> [Ultimo accesso 25/09/2017]

<sup>22</sup> <http://eacea.ec.europa.eu/tempus/> [Ultimo accesso 25/09/2017]

cooperazione universitaria, è stato approvato un progetto di nome “Development of new library services and repositories at the Universities of Western Balkan”. Promosso dalla *Education, Audiovisual and Culture Executive Agency* e presentato dalla Biblioteca Universitaria di Belgrado il 15 gennaio 2010, ha ottenuto l'effetto di coinvolgere quest'ultima istituzione con Phaidra, assieme alle università di Sarajevo, Montenegro, Maribor e il Middlesex Library Services di Londra.

- ◆ **Europeana Local<sup>23</sup>**: Sostenuta dalla Commissione Europea, si tratta di una biblioteca digitale europea che fornisce un accesso facile e veloce a libri, documenti ed opere d'arte continentali. Grazie ad un semplice click del mouse questi files (documenti archivistici, film, dipinti, libri digitalizzati, giornali, suoni, ...) possono venire raggiunti e visualizzati comodamente dalla propria città.
- ◆ **OpenAIREplus**: Il precedente progetto OpenAIRE, che ha fornito risultati sul fronte del progetto pilota per l'UE Open Access, ha portato ad un ampliamento e ad uno sviluppo in varie direzioni. È stata creata una Open Access Infrastructure di seconda generazione con l'obiettivo principale di creare un servizio robusto e collaborativo per il *cross-linking* delle pubblicazioni scientifiche *peer-reviewed*. Questo grazie ad una infrastruttura digitale capace di raccogliere, arricchire e conservare i metadati dei set di dati scientifici. Oltre a questo vengono affrontate questioni di "diritti di proprietà intellettuale, modelli di finanziamento efficienti e standard". Nel 2015 inoltre è stato lanciato OpenAire2020.
- ◆ **E-Infrastructures Austria<sup>24</sup>**: Il progetto e-Infrastructures Austria è nato nel 2014 con il proposito di affrontare le opportunità e le sfide che ogni giorno sorgono nella fase di gestione degli oggetti digitali tramite le nuove tecnologie, includendo più istituzioni di ricerca austriache possibili. Si tratta di un progetto federale, il cui obiettivo è lo sviluppo e l'espansione coordinata dei sistemi in grado di ospitare dati digitali in tutta l'Austria. Si occupa quindi, di archiviazione sicura e di disponibilità a lungo termine dei Digital Objects nei campi della ricerca e dell'insegnamento, offrendo una cornice ideale per discutere e valutare le strategie di conservazione con esperti austriaci e internazionali. Viene ritenuto fondamentale coinvolgere le comunità di ricerca, le biblioteche e i servizi IT di tutte le istituzioni scientifiche dello stato, sia nell'ambito accademico, sia in quello non universitario, al fine di lavorare insieme per promuovere lo sviluppo di repository locali, raccogliere i risultati e renderli disponibili a tutte le istituzioni partecipanti al progetto. Durante i lavori, tutte le istituzioni partner utilizzano un server locale per i documenti. Contemporaneamente, vengono realizzate infrastrutture per la ricerca di dati e altri oggetti digitali (ad esempio contenuti e-learning) e

---

<sup>23</sup> <http://pro.europeana.eu/project/europeanalocal> [Ultimo accesso 25/09/2017]

<sup>24</sup> <http://e-infrastructures.at/> [Ultimo accesso 25/09/2017]



vengono determinati futuri servizi utili per la ricerca. Tutto questo è stato reso possibile grazie ai finanziamenti del Ministero della scienza, ricerca e commercio (BMFWF).

- ◆ **LEARN (LEaders Activating Research Networks)<sup>25</sup>**: Progetto biennale volto alla creazione della LERU Roadmap for Research Data<sup>26</sup> (tabella di marcia per i dati della ricerca), ossia una guida che sensibilizzi le istituzioni culturali a gestire al meglio i dati e le risorse digitali. Spesso infatti succede che a causa di incuranza o cattiva programmazione ci sia una grande dispersione di dati. Questo può essere evitato tramite adeguate politiche di gestione dei dati che vanno dalla formazione ai costi. Learn quindi punta, tramite incontri e Workshops, a sistemare le frammentazioni tra diverse infrastrutture digitali e a massimizzare le potenzialità di una condivisione globale della ricerca.
- ◆ **Horizon2020<sup>27</sup>**: Programma realizzato dall'Unione Europea, ha come scopo quello di finanziare tutte quelle ricerche volte all'innovazione e al miglioramento della qualità di vita. È stato fortemente sostenuto dai leader e parlamentari d'Europa, concordi che questo tipo di investimento è fondamentale per il futuro del continente. Si basa principalmente su tre settori chiave: eccellenza scientifica, leadership industriale e sfide per la società. Partito nel 2014, mette sul piatto numerosi fondi tramite un piano di sette anni.

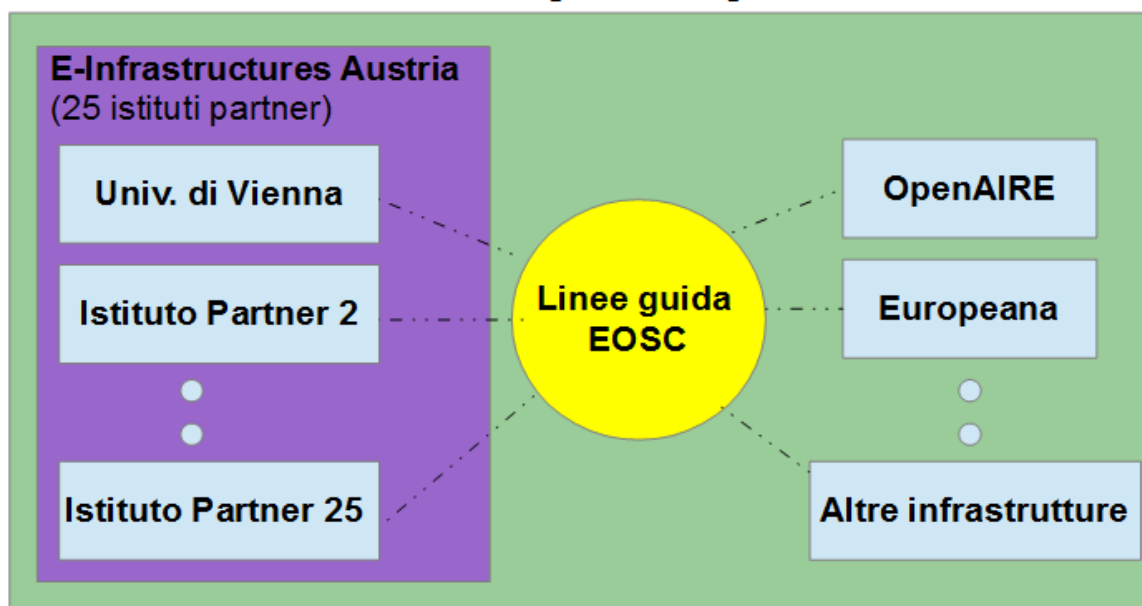
---

<sup>25</sup> <http://learn-rdm.eu/en/about/> [Ultimo accesso 25/09/2017]

<sup>26</sup> LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 2013. Disponibile a: <[http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf)> [Ultimo accesso 12/09/2017]

<sup>27</sup> <https://ec.europa.eu/programmes/horizon2020/> [Ultimo accesso 25/09/2017]

## EOSC – European Open Science Cloud



© 2017 D. Montanaro

Figura 2- Infrastrutture europee nell'ambito dell'EOSC

### 1.3 I destinatari:

Al momento della nascita di Phaidra lo staff ha avuto modo di confrontarsi, a partire da un punto di vista teorico, riguardo le politiche di accesso alla piattaforma e ai suoi contenuti. È stata data particolare attenzione alle parole del premio nobel americano Harold Varmus, il quale sostiene che i risultati della ricerca dovrebbero essere accessibili al pubblico, in quanto il grosso della spesa viene finanziato dalle tasse. Oltre a questo concetto, sono stati sottolineati alcuni postulati espressi durante la Budapest Open Access Initiative del 2002 in definizione di *Open Access*<sup>28</sup>:

- *By "open access" to this literature, we mean its free availability on the public internet,*
- *permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose,*
- *without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.*

Nonostante tutto però, sempre stando ai risultati dei lavori di Budapest, non si vuole necessariamente obbligare un fornitore di contenuti a rendere disponibile senza restrizioni l'accesso al suo lavoro. Deve sempre avere la possibilità di scegliere la licenza d'uso senza obblighi burocratici.

Questa breve introduzione concettuale, vuole essere una porta d'accesso per meglio comprendere a chi è rivolta una piattaforma come Phaidra. Stiamo parlando di un sistema aperto, non solo perché basato su un sistema open source, ma anche perché permette l'accesso a chiunque sia fornito di una connessione internet nonostante qualche necessaria limitazione per i non membri.

Per comodità si può dividere i destinatari in due grandi categorie<sup>29</sup>, gli utenti passivi e gli utenti attivi; questi ultimi si possono ulteriormente dividere in membri effettivi dell'università ed esterni a cui è stata fornita un' autorizzazione.

Gli **utenti passivi**, ossia tutti i soggetti in grado di accedere al sistema, possono compiere qualsiasi tipo di ricerca tra gli oggetti e le collezioni caricate. Se il tipo di licenza prevista lo permette, possono inoltre visualizzare, scaricare e stampare le risorse trovate. Il tutto senza necessità di una registrazione alla piattaforma.

Gli **utenti attivi** invece, solo coloro abilitati a caricare contenuti nel sistema e per questo devono assolutamente avere un account legato all'Ateneo. Escludendo al momento i soggetti che operano quotidianamente all'interno dell'istituzione (professori, studenti, ricercatori e corpo amministrativo), l'accesso può essere fornito ad utenti esterni una volta che questi abbiano ottenuto un account con relativa e-mail. Potenzialmente si parla di categorie di persone che hanno o hanno avuto

---

<sup>28</sup> <http://www.budapestopenaccessinitiative.org/read> [Ultimo accesso 25/09/2017]

<sup>29</sup> <http://datamanagement.univie.ac.at/en/about-phaidra/phaidra-portals/> [Ultimo accesso 25/09/2017]

collaborazioni con qualche dipartimento universitario, società interessate alle attività accademiche o anche singoli cittadini interessati al mondo culturale e scientifico. In ogni caso nel momento in cui si ottiene un account, bisogna prendere atto di ciò che esso comporta a livello di regolamento e seguire le istruzioni fornite dal personale autorizzato. È chiaro che, essendo un sistema aperto con determinati fini di conservazione e ricerca, non sia possibile usare un account a fini commerciali, pubblicitari, privati, etc.<sup>30</sup>

È giunto ora il momento di esporre quelle che sono le categorie più direttamente coinvolte, ossia i membri effettivi dell'università. Phaidra è stato sviluppato per la ricerca, l'insegnamento e l'amministrazione presso l'Università di Vienna, in modo da offrire vantaggi e nuove possibilità, come vedremo a seguire attraverso ogni singolo ambito:

- **Ricerca:** poniamo il caso in cui un professore sia intenzionato a pubblicare una propria ricerca con i relativi articoli e risultati. Può caricare diverse tipologie di oggetti come, se necessario, fotografie con relative descrizioni anche multilingua; può caricare i testi degli articoli trasformandoli in PDF e decidere con chi condividerli; può inoltre, caricare il testo del resoconto finale sfruttando le potenzialità del *Phaidra Importer*, il quale permetterà poi di scaricare il documento o parte di esso oppure visualizzarlo con il *Book Viewer*.

Questi oggetti possono essere poi riuniti in collezioni e, a discrezione dell'autore, condivisi tramite un link permanente: col mondo intero, con qualche persona in particolare, con il proprio Ateneo o con la facoltà di riferimento. Un'altra funzionalità interessante consiste nel poter creare dei modelli per le descrizioni delle risorse: in questo modo, il produttore di dati ha la possibilità di condividere la struttura dei metadati con i colleghi e fare un lavoro pulito che implica guadagno di tempo e riutilizzo di schemi preimpostati.

Una volta che il professore in questione va in pensione, sarà possibile per lui trasferire la proprietà degli oggetti ad un collega il quale, se vorrà, potrà continuare a dare il suo contributo alla ricerca.

Un altro caso da prendere in considerazione è quello in cui un ricercatore, facente parte del corpo dell'Università Di Vienna, stia lavorando ad un progetto con colleghi provenienti da altri paesi ed istituzioni. Affinché tutti i partecipanti alla ricerca possano contribuire mediante l'accesso a Phaidra, il sistema offre due diverse opzioni: la prima, tramite la quale lo studioso dell'Ateneo viennese assegna un account ospite attivo ad ognuno dei partecipanti esterni, affinché questi diventino responsabili personalmente degli oggetti caricati e dei relativi metadati. L'altra via invece, presuppone la creazione di un unico account utilizzabile da tutti i membri del gruppo di ricerca; di conseguenza, tutti hanno accesso a tutti gli oggetti

---

<sup>30</sup> [https://phaidra.univie.ac.at/terms\\_of\\_use/show\\_terms\\_of\\_use](https://phaidra.univie.ac.at/terms_of_use/show_terms_of_use) [Ultimo accesso 25/09/2017]

disponibili.

- **Insegnamento:** una situazione comune in un'istituzione come un'università, è quella in cui un professore vuole rendere dei documenti accessibili esclusivamente ai suoi studenti. Questo caso può essere risolto creando un gruppo di persone autorizzate su Phaidra, dargli un nome e impostare l'accessibilità in maniera esclusiva nei loro confronti. Chiaramente i formati sono sempre tutti disponibili: documenti, audio, video, immagini, etc. Fino al momento in cui le impostazioni non vengono cambiate, solo il gruppo in questione avrà titolo di scaricare le risorse. Se il documento che si desidera condividere è solo uno, basta crearlo e condividerlo con quelli che saranno i possessori del link. Le collezioni, una volta create, possono sempre venire espanse con l'aggiunta di nuove pubblicazioni e le autorizzazioni all'accesso saranno sempre riassegnabili.

Un docente ha anche la possibilità di salvarsi una presentazione PowerPoint importante su Phaidra e averla sempre disponibile globalmente anche se salvabile solo come oggetto sconosciuto; ciò implica che il sistema non lo ritiene di un formato adatto per la conservazione a lungo termine e per questo rientra in una categoria a parte chiamata "Data".

- **Studio:** Phaidra ha interessanti funzionalità anche dal punto di vista degli studenti. Ognuno di essi, alla propria iscrizione, entra in possesso di un account con un gigabyte di memoria utilizzabile per le proprie attività: salvare le relazioni e dare loro la visibilità desiderata, inserire i risultati di ipotetici test di laboratorio, ricercare materiale rilevante per gli studi e metterlo assieme tramite collezioni. In quest'ultimo caso Phaidra mostra le sue potenzialità di portale di conoscenza infatti, una volta riuniti i file, questo agglomerato può essere utilizzato come portale tematico espandibile al quale viene assegnato un nome.
- **Amministrazione:** la piattaforma risulta comoda anche nel momento in cui si desidera archiviare i documenti amministrativi i quali, dopo essere stati descritti con gli opportuni metadati, vengono condivisi tramite licenza a chi di dovere (es. Singole persone, membri dell'Ateneo, altre istituzioni, ...). Questi oggetti, nel momento in cui diventano non più attuali, possono essere aggiornati con una nuova versione mentre alla vecchia viene impedita la visualizzazione. Anche in questo caso i gruppi di oggetti affini per via di qualsiasi motivo possono essere riuniti in collezioni, le quali sono condivisibili in blocco tramite un unico link permanente.

#### 1.4 I servizi offerti da Phaidra:

A partire dalla comprensione del soggetto “destinatario” nella scorsa sezione, l'obiettivo qui sarà quello di capire quali sono i servizi che vengono offerti. Dalla conservazione a medio e a lungo termine dei documenti digitali e dei dati, fino alle potenzialità che possono essere espresse da una piattaforma che sostiene l'interdisciplinarietà e l'approccio multilinguale. Prendendo come obiettivo di questa tesi il caso offerto dall'università di Vienna vediamo di capire velocemente come, l'ufficio che si occupa della conservazione degli Oggetti Digitali, ha strutturato i repositories per offrire il servizio ai più disparati destinatari. Attualmente, dopo un recente cambiamento, è stato istituito il dipartimento chiamato PHAIDRA-UNIDAM che mette insieme le omonime piattaforme. Oltre alla pagina principale che dà il nome all'intero progetto, vediamo cosa succede nelle pagine parallele che hanno tutte caratteristiche diverse in base all'uso previsto.

- **Phaidra** (<https://phaidra.univie.ac.at/>): si parla in questo caso del portale principale che, dal 2008, è disponibile ai membri dell'Università di Vienna come un sistema di gestione delle risorse digitali a livello universitario con funzioni di archiviazione a lungo termine. Facilita la documentazione e l'archiviazione a lungo termine di oggetti digitali. Tutti gli oggetti ottengono una firma digitale permanente, mentre metadati multilingue consentono una descrizione accurata di ogni oggetto. È inoltre possibile utilizzare licenze e riutilizzare gli oggetti. Gli sforzi di ricerca sono pertanto resi visibili e le conoscenze globali e transdisciplinari possono essere più facilmente trasferite. I dati digitali sono archiviati in modo sicuro, sono sempre disponibili e sono sempre pronti per il riutilizzo. Supporta varie funzioni, tra cui: pubblicazione elettronica, archiviazione delle immagini, visualizzazione delle collezioni, creazione di eBooks, salvataggio e streaming di video e molto altro ancora.
- **Phaidra Temp** (<https://phaidra-temp.univie.ac.at/>): utile nel momento in cui una risorsa deve essere conservata per un tempo non per forza tendente all'infinito. Infatti Phaidra Temp serve fondamentalmente per la conservazione di dati a medio termine. Questi poi, se fosse il caso e se il produttore lo ritiene opportuno, potranno essere spostati in Phaidra con i cambiamenti conseguenti.
- **Phaidra Sandbox** (<https://phaidra-sandbox.univie.ac.at/>): sviluppato con l'idea di essere uno strumento adatto a prendere confidenza con l'interfaccia e le funzionalità di Phaidra, senza avere alcuna responsabilità su quanto caricato. Infatti è pensato a scopo di test e per fare pratica con la piattaforma, senza che gli oggetti caricati diventino di dominio pubblico. È accessibile esclusivamente tramite la rete dell'università, a meno che non si usi un'apposita VPN (Virtual Private Network).
- **Phaidra National Depot** (<https://depot.phaidra.at/>): si tratta di un progetto volto ad offrire le funzionalità di Phaidra ad enti e progetti non universitari.

- **u:Scholar** (<https://uscholar.univie.ac.at/>): è un servizio di pubblicazione gestito dall'Università di Vienna. Si tratta di un repository istituzionale il cui scopo è quello di raccogliere e presentare i risultati ottenuti dal personale universitario. Ciò include articoli di giornale e monografie, come anche relazioni di lavoro o derivate da congressi. In linea di massima è possibile utilizzare u:Scholar anche per archiviare dati di ricerca. Per essere in grado di soddisfare i requisiti speciali necessari per questi ultimi, u:Scholar riceverà ulteriori aggiornamenti funzionali, come ad esempio specifici metadati per descrivere i dati di ricerca, nonché una gestione appropriata della versione. In quanto parte dell'iniziativa Open Access della biblioteca universitaria, u:Scholar è stato progettato come un contributo verso l'accesso globale dei risultati della ricerca accademica, specie se questa è stata possibile grazie ai finanziamenti pubblici. Molti editori consentiranno una seconda pubblicazione in un repository (Green Road to Open Access)<sup>31</sup>. Il personale accademico è invitato a utilizzare u:Scholar per consentire il libero accesso ai propri risultati di ricerca.

Il progetto è in stretta collaborazione con la squadra di documentazione della biblioteca (u:Cris<sup>32</sup>) in modo da poter determinare quali contenuti possono essere archiviati in u:Scholar senza l'espressa autorizzazione degli autori sulla base della rispettiva licenza. Ciò avviene nel caso di tutti i contributi pubblicati come Gold Open Access. La raccolta dei metadati da parte della squadra di documentazione di ricerca è stata regolata dalla legge nella Wissensbilanz-Verordnung per le università austriache.

Tecnicamente, u:Scholar è un'interfaccia utente specifica per Phaidra. Questa applicazione si basa sul software open source Fedora Commons per archiviare oggetti digitali e metadati.

La collaborazione con il Centro di Calcolo universitario è continua, allo scopo di garantire la disponibilità e lo sviluppo continuo di u:Scholar.

- **Unidam** (<http://unidam.univie.ac.at/>): attualmente è in uso la versione 4 di fine 2009; basato sull'EasyDB system, sviluppato a partire dal 2003 dalla “Berlin company Programmfabrik GmbH”, può venire utilizzato dai ricercatori per salvare e condividere le loro risorse per l'insegnamento e la ricerca. Viene supervisionato e portato avanti dal dipartimento Phaidra-Unidam. È in corso una conversione alla versione 5 di EasyDB, in cui il CMS (Content management system), è stato completamente riprogrammato negli ultimi anni ed ora è stato perfettamente adattato ai bisogni dell'Università di Vienna, anche se siamo ancora in una fase pilota (a partire da maggio 2017).

---

<sup>31</sup> Esistono due strategie principali per garantire pubblicazioni ad accesso aperto: la prima, definita Green road, sostiene la pratica dell'auto-archiviazione da parte degli autori. Le copie dei loro articoli di conseguenza, vengono conservate in archivi istituzionali o disciplinari, se non nei loro siti web personali. La seconda, chiamata Gold road, si riferisce ad una pubblicazione di riviste i cui articoli sono direttamente accessibili tramite l'Open Access.

<sup>32</sup> <https://ucrisinfo.univie.ac.at/>





## Capitolo 2 – Modello di digital workflow e soggetti coinvolti

Leggendo i più recenti report nell'ambito dell'Open Science, come ad esempio quelli redatti dal progetto LEARN, capita spesso di trovarsi a confronto con il concetto di Research Data Management (RDM)<sup>33</sup>, il quale viene utilizzato ormai costantemente dagli enti e istituti di ricerca. È fondamentale capire nel dettaglio a cosa si riferisce, dal momento che sarà spesso presente nel corso dell'elaborato. Il termine Research Data Management (RDM) si riferisce all'archiviazione, all'accesso e alla conservazione dei dati prodotti da un determinato progetto. Le pratiche di gestione dei dati riguardano l'intero loro ciclo di vita, dalla pianificazione del progetto alla sua conduzione, fino al backup dei dati quando vengono creati e poi riutilizzati. Senza dimenticare la conservazione a lungo termine dei dati finiti dopo la fine del progetto.

Ecco a seguire un accenno di quelle attività specifiche e problemi che sono incluse all'interno del Data Management: in modo corretto indicare, tramite un nome appropriato e riconoscibile, i nomi dei files; controllarli e garantirne la qualità; controllarne gli accessi; documentarli su vari livelli; archivarli e conservarli; dividerli e riutilizzarli; mantenerli integri e sicuri; garantirne la privacy e i diritti; creare metadati e vocabolari controllati che siano appropriati.

A garantire che tutte queste tematiche siano affrontate nel modo più opportuno, esistono una serie di soggetti, i quali sono coinvolti nelle fasi che vanno dalla produzione dei dati fino al vero e proprio servizio offerto. Verrà quindi tracciato un profilo del produttore di dati, dando la possibilità di capire chi può svolgere questo servizio e come. Mentre per quanto riguarda l'offerta ci sono tre entità principali, in stretta collaborazione tra loro, dalle quali non si può prescindere: la biblioteca, il centro di calcolo e i servizi legali. Queste entità sono fondamentali nelle loro competenze, perché vengono messe insieme per offrire un prodotto funzionale e sicuro che si rivolge a tutti gli ambiti universitari. Nella figura 3, è possibile vedere il modello di digital workflow per la gestione dei dati di Phaidra, grazie al quale ci si rende conto di come funzioni la piattaforma. Le tematiche poi relative alle singole fasi di lavoro verranno trattate approfonditamente nel corso del capitolo 4.

---

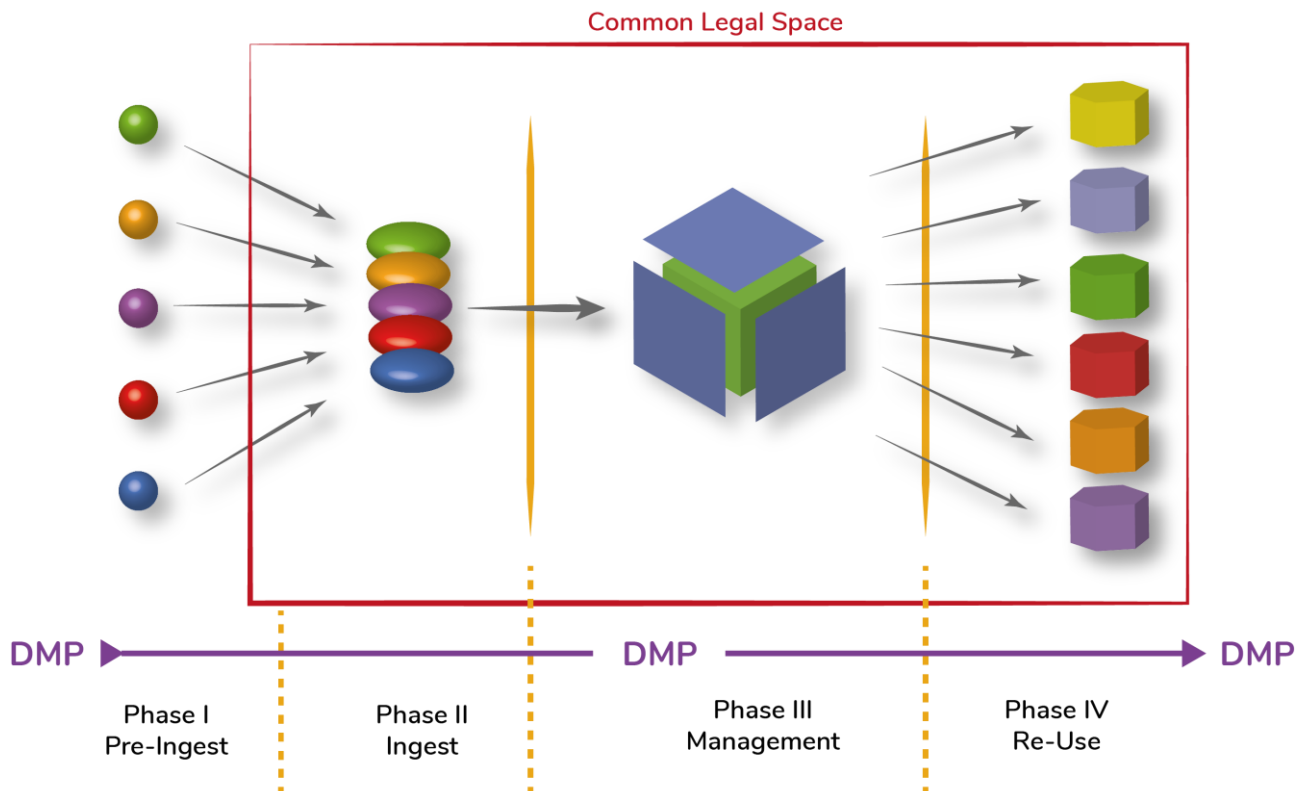
<sup>33</sup> <http://learn-rdm.eu/en/glossary/research-data-management/> [Ultimo accesso 25/09/2017] Vedi anche su RDM in biblioteche: Cox, AM and Pinfield, S (2013) *Research data management and libraries: Current activities and future priorities*. Journal of Librarianship and Information Science, Published Online before print, June 28, 2013. Disponibile a: [http://eprints.whiterose.ac.uk/76107/7/WRRO\\_76107.pdf](http://eprints.whiterose.ac.uk/76107/7/WRRO_76107.pdf) [Ultimo accesso 25/09/2017] e Sheila Corral, Mary Anne Kennan, and Waseem Afzal, *Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research*, LIBRARY TRENDS, Vol. 61, No. 3, 2013 ("Research into Practice," edited by Sheila Corral and Barbara Sen), pp. 636–674. Disponibile a: <https://www.ideals.illinois.edu/bitstream/handle/2142/46046/61.3.corral02.pdf?sequence=2&isAllowed=y> [Ultimo accesso 25/09/2017]

Data from various sources

Data conversion and enrichment

Data Preservation

Data Re-use



Digital workfolw model for data management  
Raman Ganguly, University of Vienna 2017

Figura 3 - Modello di Digital Workflow

<http://phaidra.univie.ac.at/o:527220>

### 2.1 Il produttore di dati:

Il produttore di dati è uno dei ruoli chiave all'interno del ciclo di vita dei dati, infatti è colui che detiene la responsabilità di assicurare che gli oggetti che carica siano il risultato di una corretta pratica di preparazione degli stessi, in linea con quanto progettato e definito dall'istituzione di ricerca e dall'RDM. Spesso, essendo il produttore un soggetto universitario, potrebbe poi trovarsi nella situazione di visualizzare, riutilizzare e citare i contenuti. Infatti i maggiori fruitori di Phaidra all'interno dell'Ecosistema dell'Università di Vienna sono appunto professori, ricercatori e studenti. Diverso il caso dell'Università di Padova dove, ad esempio, gli studenti non sono ammessi alla fase di caricamento. Nella sezione 1.3, in cui è stato affrontato brevemente il tema dei destinatari della piattaforma, veniva specificata la differenza tra utenti attivi e passivi; il ruolo dei primi, il quale corrisponde con il produttore di dati, è ciò che ci interessa e che verrà approfondito nelle prossime righe.

Un'attitudine fondamentale che l'utente attivo deve fare propria senza esitazione, è la *consapevolezza* del proprio lavoro. Il produttore infatti, deve essere consapevole di ciò che si prevede di fare con i dati in suo possesso, essendo a conoscenza di quali servizi e strumenti sono a sua disposizione. E qui entra in campo un altro aspetto fondamentale: la *formazione*. I campi su cui questa va svolta sono ampi e spaziano tra proprietà, sicurezza, etica, responsabilità, ri-uso, archiviazione e condivisione di dati. Si tratta dell'unica via utilizzabile per sfruttare al meglio questa nuova tecnologia e dare una svolta positiva alla questione del Data Management. Nel caso di Phaidra, la biblioteca organizza formazioni e workshop regolarmente oltre a fornire le informazioni più importanti tramite la propria pagina web; anche se ovviamente i materiali online hanno una potenzialità d'impatto minore. È presente inoltre un servizio di supporto che risponde alle necessità sia tecniche che pratiche, oltre che la possibilità di ottenere appuntamenti su richiesta.

Bisogna assicurare quindi che il valore di un'ideale come quello dell'Open Data sia pienamente realizzato tramite abilità pratiche e di supporto per riutilizzare i dati e assicurarne la qualità<sup>34</sup>. Un potente fattore di lancio della condivisione dei dati sta nella loro affidabilità. Per questo i soggetti produttori come studenti e ricercatori dovranno essere istruiti al fine di poter sfruttare al meglio queste nuove potenti possibilità.

Ci sono molte questioni da considerare e del quale il produttore di dati deve essere a conoscenza come, ad esempio, la limitatezza e il costo dello spazio d'archiviazione, le possibilità legali di accesso alla risorsa e le sue responsabilità a riguardo, la conoscenza degli standard descrittivi delle risorse che permettono loro di essere trovabili.

---

<sup>34</sup> Vedi anche: PARK J.-R. – TOSAKA Y., *Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms*, "Cataloging & Classification Quarterly", 48, 8 (2010), pp. 696-715.

Ma cosa succede quindi quando un utente carica un oggetto digitale e qual è il suo ruolo?

Premettiamo che per accedere l'utente deve essere in possesso di una casella di posta elettronica o di un account U:net forniti dall'Università di Vienna. Tramite l'accesso, reso possibile da questi mezzi, l'utente accetta le condizioni d'utilizzo<sup>35</sup> valide per tutti coloro che caricano qualche oggetto sulla piattaforma. Queste sono state redatte con lo scopo di cautelarsi da ogni possibile problematica derivata da un uso improprio della piattaforma e per offrire quindi un miglior servizio. Per quanto riguarda la durata, le condizioni sono valide finché il fornitore del servizio (ossia Phaidra) esiste o finché non decide di cambiarle come stabilito dai suoi diritti. In ogni caso ogni eventuale cambiamento viene reso noto tramite il sito internet. L'utente inoltre, nel momento in cui imposta il proprio account deve fornire i propri dati personali come nome, cognome, data di nascita, indirizzo, assicurando che siano corretti, aggiornati e completi. Accetta inoltre che questi possano venire trattati secondo le condizioni stabilite, avendo comunque il diritto di accesso, rettifica, cancellazione o blocco degli stessi.

Il produttore di dati, nel momento in cui decide di caricare una risorsa online, ne diventa automaticamente il *possessore*. Come tale, ha diritto d'accesso illimitato (lettura, manipolazione, modifica e autorizzazioni) sull'oggetto e ne ha piena responsabilità. Deve garantire però, di essere titolare dei diritti di sfruttamento degli oggetti multimediali caricati e ha il diritto di scegliere il tipo di licenza che più ritiene opportuno. Importante è la scelta della giusta repository in cui caricare l'oggetto. Su Phaidra, nell'istante in cui viene caricato un file, questo rimane per un tempo tendente all'infinito sulla piattaforma e su alcuni aspetti non può essere modificato. Se non c'è la certezza di questa necessità, una buona opzione può essere quella di utilizzare Phaidra Temp sfruttando l'archivio a medio termine. Infatti, quando un utente carica un oggetto digitale su Phaidra, non può decidere di cancellarlo come è stato stabilito dai principi e dalle linee guida della piattaforma. Esiste questa possibilità solo nel caso ci siano sufficienti ragioni per provare che il blocco del file non sia un provvedimento abbastanza efficace, come nel caso di una provata situazione di illegalità e/o di un caso che danneggi il database. È responsabile della qualità del contenuto; considerando che l'archiviazione avviene con l'idea di durare il più a lungo possibile, potrebbe risultare solo uno spreco di spazio caricare oggetti non adeguati. In questo bisogna tenere conto ad esempio dei formati, i quali vengono suggeriti a seconda del tipo di file, in modo da cercare di garantire una certa compatibilità e versatilità più a lungo possibile. Assicura di utilizzare il servizio in modo legalmente accettabile, senza perseguire scopi illegali e di non diffondere informazioni e oggetti discriminatori, estremisti o

---

<sup>35</sup> [https://phaidra.univie.ac.at/terms\\_of\\_use/show\\_terms\\_of\\_use](https://phaidra.univie.ac.at/terms_of_use/show_terms_of_use) [Ultimo accesso 27/08/2017]

pornografici. Questo comporta che, nel momento in cui una terza parte venga lesa intenzionalmente o meno e compori spese al fornitore del servizio, queste verranno indennizzate dall'utente in fallo. Sceglie la licenza, la quale successivamente non può più essere cambiata ed è suo compito assicurare che non ci siano infrazioni nel copyright. Normalmente gli utenti vengono incoraggiati ad usare la Creative Commons<sup>36</sup>. Con una licenza più chiusa si ha la possibilità di mostrare a tutti solo i metadati ed un'anteprima.

Il produttore di dati inoltre imposta i metadati, i quali, a differenza della licenza, possono venire cambiati. Questi sono fondamentali per rendere gli oggetti trovabili e collegabili ad altri oggetti affini e vanno considerati come parte integrante della pubblicazione scientifica. Gli oggetti affini tra l'altro possono venire anche inseriti in una collezione, la quale avrà un identificatore persistente proprio come ogni singolo oggetto.

La pubblicazione elettronica e l'archiviazione è totalmente gratuita per i produttori di dati, ma bisogna tenere conto che vanno rispettati determinati limiti di spazio. Quando l'utente ha completato l'operazione di caricamento, concede allo staff universitario il diritto di utilizzare l'oggetto multimediale, integralmente o in parte, per apportare modifiche allo stesso nel momento in cui siano necessarie per motivi tecnici o per quanto riguarda i requisiti di archiviazione a lungo termine.

Affinché tutti questo lavoro venga svolto nel migliore dei modi, è fondamentale conoscere e utilizzare lo strumento chiamato DMP (Data Management Plan), del quale si parlerà approfonditamente nel prossimo capitolo. Si tratta di un documento che, esplicitando formalmente tutti gli obiettivi e le finalità del progetto di ricerca, aumenta l'efficienza tracciando la via da utilizzare nel ciclo di vita dei dati.

Un'interessante possibilità per il produttore di dati, è quella di rendere le proprie risorse disponibili anche nel repository istituzionale U:Scholar. A questo scopo però, lo staff di U:Scholar deve mettere a punto gli oggetti e i metadati con le sue politiche di conformità. Ciò richiede il trasferimento dei diritti di proprietà dall'utente al repository istituzionale. Se l'oggetto risulta non conforme ai requisiti di U:Scholar, i diritti d'accesso tornano all'utente. In ogni caso il possessore originale ha sempre la possibilità di ritirare l'oggetto da U:Scholar e quindi riavere nuovamente i diritti d'accesso e cambiare i metadati.

---

<sup>36</sup> <https://creativecommons.org/> [Ultimo accesso 12/08/2017]

## 2.2 Le principali entità che offrono i servizi:

### A. La biblioteca

In qualità di istituzione alla base della conservazione e della diffusione della conoscenza, offre servizi e segue con cura quello che concerne le politiche interne di archiviazione e ha un ruolo di responsabilità sociale nei confronti dell'utenza.

*“...To achieve their ultimate goal of enhancing scholarly communication, libraries have served as crucial mediators for bridging the creators of information and knowledge to end users. Libraries have assumed responsibility for collecting, disseminating, and preserving scholarly information resources and providing other necessary services to scholars.”<sup>37</sup>*

Association of College and Research Libraries (ACRL), 2003

Questa citazione ritengo possa essere considerata un concentrato di ciò che la biblioteca in tutto questo dovrebbe rappresentare: la figura chiave per quanto riguarda il *coordinamento* del progetto.

Infatti è lei che si occupa di assicurare che gli studiosi abbiano l'opportunità di avere a disposizione infrastrutture e supporto per rendere i dati disponibili in base ad ogni loro tipologia; siano essi interni, esterni o nel repository istituzionale. È questo non riguarda solo la fornitura dello spazio per depositare i dati, ma anche la formazione e la guida per facilitare l'uso del Data Management e delle citazioni<sup>38</sup>. Il coordinamento infatti, gioca un ruolo fondamentale nel guidare i cambiamenti culturali all'interno dell'istituzione, articolando e rinforzando il suo orientamento attraverso l'accesso libero. Tramite incentivi, assicurando che il valore dell'Open Data sia riconosciuto.

Affinché ciò sia possibile è necessario che vengano promossi i benefici dell'RDM<sup>39</sup>, che gioca un ruolo fondamentale nel perseguire un determinato progresso condiviso nella gestione dei dati. Le biblioteche infatti, sono un soggetto dominante nel processo promozionale, grazie alla loro storia nell'accesso all'informazione scientifica, all'attuale esperienza nel campo dell'Open Access e al loro ruolo di giocatrici stabili all'interno dell'Università e di altri istituti di ricerca. Possono fornire un supporto RDM proattivo ai loro ricercatori e alle unità di supporto alla ricerca (ad esempio help desk, FAQ, portali tematici, formazione, infrastrutture, budget per l'acquisizione di DOIs – Digital Object Identifiers), promuovendo una cultura della *condivisione*. I servizi RDM dovrebbero inoltre affrontare le differenze disciplinari facendone tesoro e adattare le loro infrastrutture in modo da favorire un

---

<sup>37</sup> FERUS A., *Wissenschaftskommunikation, e-Infrastrukturen und wissenschaftliche Bibliotheken*, 2016. Disponibile a: <<http://phaidra.univie.ac.at/o:441295>> [Ultimo accesso 12/09/2017]

<sup>38</sup> LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 2013, pp 28-29. Disponibile a: <[http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf)> [Ultimo accesso 12/09/2017]

<sup>39</sup> LEARN PROJECT, *20 RDM Best-Practice Recommendations*, 2017. Disponibile a: <<http://learn-rdm.eu/wp-content/uploads/20-RDM-Policy-Recommendations.pdf>> [Ultimo accesso 12/09/2017]

orientamento che punti all'*interdisciplinarietà*. Il supporto a livello di gestione è un fattore chiave che consente alle biblioteche di adempiere al loro ruolo nel RDM<sup>40</sup>.

Si può dire dunque che la biblioteca, in quanto istituzione, abbia una certa *responsabilità sociale* in questo campo e per raggiungere il suo scopo deve seguire determinate politiche per offrire un *servizio di qualità* con uno sguardo ai *costi* complessivi e alla loro copertura secondo le tempistiche previste. I costi sono un campo piuttosto arduo da gestire, specialmente quando si tratta di progetti a lungo termine. Infatti vanno considerate varie spese che vanno dalle risorse umane alla formazione e dai servizi alla sicurezza. L'istituzione dovrebbe essere responsabile nell'occuparsi dei dati di ricerca creati all'interno della locale comunità scientifica attraverso l'intero ciclo di vita degli stessi. Per questo deve dotarsi di adeguate politiche specificando le procedure e le misure da prendere nel momento in cui un soggetto produttore decida di lasciare l'istituzione. Occorre mettere in atto servizi per supportare e accompagnare il ciclo di vita dei dati.

Tra i vari campi in cui la biblioteca offre il suo contributo a livello di coordinamento ne vediamo alcuni<sup>41</sup>, cercando di capire in cosa questo consiste. In più di un caso le tematiche verranno poi nuovamente trattate più approfonditamente quando si parlerà di centro di calcolo e servizi legali, in quanto hanno maggiore voce in capitolo.

- **Controllo degli accessi:** offre supporto nell'eventualità di problemi relativi agli accessi ai dati e al controllo degli stessi;
- **Citazione dei dati:** fornisce identificatori persistenti, inclusi gli Digital Objects Identifiers (DOIs) per ogni tipologia di oggetto e per le collezioni;
- **Documentazione dei dati:** aiuta i ricercatori a determinare come meglio documentare i propri dati sin dall'inizio di un progetto, seguendo gli standard disciplinari;
- **Pianificazione della gestione dei dati:** offre un approccio alla leadership istituzionale e alle unità di supporto alla ricerca per sviluppare processi di assistenza ai piani di gestione dei dati nel campus. Collega i ricercatori con le risorse locali e disciplinari per soddisfare i requisiti di eventuali finanziatori;
- **Hosting dei dati:** lavora con le unità di supporto istituzionale e di ricerca per fornire infrastrutture per l'hosting di dati (archivi di dati istituzionali) o per aiutare i ricercatori a mettere a disposizione infrastrutture (depositi disciplinari);
- **Proprietà intellettuale e copyright:** fornisce indicazioni sulla proprietà intellettuale e sui diritti d'autore riguardanti i dati di ricerca;

---

<sup>40</sup> Vedi anche: <https://www.clir.org/pubs/reports/pub167/RDM.pdf> [Ultimo accesso 25/09/2017]

<sup>41</sup> FERUS A., *Wissenschaftskommunikation, e-Infrastrukturen und wissenschaftliche Bibliotheken*, 2016. Disponibile a: <http://phaidra.univie.ac.at/o:441295> [Ultimo accesso 12/09/2017]

- **Conservazione:** consiglia i formati di dati più appropriati per la conservazione a lungo termine, senza precludere l'utilizzo degli altri;
- **Privacy e riservatezza:** consiglia i ricercatori e il personale dell'ufficio di ricerca in materia di privacy e riservatezza nella gestione dei dati;
- **Selezione dei repositories:** consente ai soggetti produttori di dati di selezionare archivi digitali attendibili per la conservazione dei loro oggetti digitali, siano essi repositories disciplinari o repositories gestiti in modo istituzionale;
- **Formazione RDM:** Comunica le raccomandazioni pratiche sviluppate dalla comunità RDM ai gruppi di ricercatori, in modo che questi siano sempre aggiornati;
- **Impatto scientifico:** promuove meccanismi per monitorare l'impatto della condivisione dei dati attraverso downloads, citazioni, ecc ...;
- **Output scientifico:** consente di collegare le risorse ad altre pubblicazioni scientifiche attraverso i metadati e i meccanismi di citazione.

Per quanto riguarda il futuro, molti studi hanno evidenziato come la figura del bibliotecario sia destinata a giocare un ruolo nuovo, che metterà in atto nuove competenze<sup>42</sup>. Queste dovranno venire poi impresse ai gruppi di lavoro che verranno formati. Si parla di metodi per la ricerca interdisciplinare dei dati digitali, di competenze IT, di questioni etiche e legali e di Technology watch. Con quest'ultimo campo si intende la capacità di osservare e cogliere vantaggi dai continui sviluppi tecnologici (supporti IT, sistemi, formati, metadati, DMPs, ...) <sup>43</sup>. Lo sviluppo delle competenze in Data Management e il riutilizzo dei dati è un fattore chiave per l'Open Data. Un'istruzione solida che si basa sulle migliori pratiche nella gestione dei dati garantirà che i dati di ricerca siano affidabili e riutilizzabili.

---

<sup>42</sup> LEARN PROJECT, *LEARN Toolkit of Best Practice for Research Data Management*, 2017, pp. 94-102. Disponibile a: <<https://doi.org/10.14324/000.learn.00>> [Ultimo accesso 12/09/2017]

<sup>43</sup> ROVIRA C., *Technology Watch and Competitive intelligence for SEM-SEO*, "Hipertext.net", num. 6, 2008. Disponibile a: <<https://www.upf.edu/hipertextnet/en/numero-6/vigilancia-tecnologica.html>> [Ultimo accesso 25/09/2017] e <<http://www.dpconline.org/knowledge-base/tech-watch-reports>> [Ultimo accesso 25/09/2017]



### B. Il centro di calcolo

Mentre come abbiamo appena visto, la biblioteca si occupa del coordinamento e della gestione dei progetti, la gestione e lo sviluppo a livello informatico vengono realizzati nel centro di calcolo universitario (ZID). È fondamentale un lavoro a stretto contatto e in piena collaborazione tra le parti per arrivare ad offrire un servizio adeguato sotto tutti i punti di vista necessari. Tra le mansioni principali del centro di calcolo possiamo trovare in primo luogo quella di *fornire un'infrastruttura* adeguata all'archiviazione degli oggetti digitali, in modo da salvaguardare ciò che viene salvato.

Quest'infrastruttura deve essere versatile in quanto deve potersi adattare agli specifici bisogni della ricerca e della conservazione, sempre considerando il valore dell'interdisciplinarietà e quindi la possibilità di avere al proprio interno reindirizzamenti piuttosto intrecciati tra loro. In certi casi, a seconda della natura dei dati e delle risorse disponibili, potrebbero essere potenzialmente le biblioteche ad assumersi la responsabilità per la cura e l'archiviazione dei dati. Ad esempio, se prendiamo in considerazione l'esempio delle scienze umanistiche, può capitare che le biblioteche si assumano la responsabilità di curare il patrimonio culturale digitale, come i dati utilizzati per la ricerca nelle scienze umanistiche digitali. Infatti, nelle discipline che producono quantità di dati minori e più eterogenei può essere la biblioteca a fornire l'infrastruttura per rendere disponibili questi tipi di dati. Nel caso di Phaidra e rimanendo sempre nell'ottica dell'EOSC, ci confrontiamo con dei dati di natura piuttosto diversa tra loro e che formano una mole imponente.

Per ovviare a questa difficoltà, la biblioteca dovrebbe fornire le linee guida per quanto riguarda i piani di gestione dei dati, le infrastrutture di accesso aperto, le basi della buona gestione dei dati, la proprietà intellettuale, la valutazione dei dati e la loro citabilità. Di conseguenza, gli esperti IT che vengono guidati su questo piano possono arrivare a sviluppare soluzioni flessibili ed efficaci, identificando le tecnologie necessarie ai ricercatori per massimizzare il lavoro della loro ricerca<sup>44</sup>.

Non bisogna mai dimenticare che il software viene spesso sviluppato con lo scopo di generare dati, ma successivamente anche per la loro visualizzazione e riutilizzo. È necessario quindi a posteriori continuare a gestire il software anche dopo la fine del progetto di ricerca e dei finanziamenti<sup>45</sup>.

Un'altra funzione del centro di calcolo consiste nel *fornire supporto* per quanto riguarda la creazione dei Data Management Plans, l'archiviazione dei dati, la loro ricerca e la loro citabilità. Può quindi dare consigli su come strutturare i dati affinché questi sfruttino al meglio la corrente tecnologia. Si tratta in questo caso di competenze tecniche che possono essere fornite agli altri soggetti sia nel momento in cui ci sia una necessità, sia tramite una formazione preventiva.

Si occupa inoltre di gestire le procedure di *autenticazione* al sistema e di *accesso* alle risorse<sup>46</sup>.

---

<sup>44</sup> LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 2013, pp. 28-29.

<sup>45</sup> FERUS A., *Wissenschaftskommunikation, e-Infrastrukturen und wissenschaftliche Bibliotheken*, 2016.

<sup>46</sup> LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 2013, pp. 28-29.

Garantisce quindi che l'utente, nel momento in cui voglia caricare, visualizzare o riutilizzare una risorsa possa farlo nel pieno delle potenzialità offerte dalla piattaforma. Anche il centro di calcolo, come la biblioteca, ha il dovere di tenere d'occhio il progresso tecnologico (*Technology watch*), in modo da poter sempre offrire la soluzione più adatta che rispecchi gli ideali dei cardini del progetto.

Il centro di calcolo non si occupa di visualizzare e/o controllare gli oggetti multimediali e le informazioni memorizzate dagli utenti da un punto di vista legale, sia prima che dopo la sua pubblicazione online. Non esiste qui alcun controllo in merito alla situazione della proprietà intellettuale, in particolare per quanto riguarda l'ammissibilità alla tutela e alla proprietà dei diritti. Però, nel momento in cui riceve informazioni su eventuali attività illegali legate ad una risorsa o se l'illegalità di un oggetto multimediale diventa evidente, gli amministratori hanno il diritto di rimuoverlo immediatamente senza creare ulteriori danni. Allo stesso modo, possono essere eliminati o rimossi oggetti o dati multimediali che potenzialmente potrebbero danneggiare il database del servizio<sup>47</sup>.

Volgendo lo sguardo alle prospettive future, la via tracciata è quella del “sistema di sistemi”, dove ogni singola piattaforma degli istituti di ricerca potrà fornire i propri contenuti multimediali dati alle altre piattaforme affini, come ora succede tra Phaidra e altri progetti come Europeana e OpenAIRE. Si parla a livello europeo di un virtual layer<sup>48</sup> (livello virtuale), il quale cambierà determinati concetti di gestione (software, errori, funzionamento, supporto, ecc..) e quindi anche parzialmente il ruolo dei singoli centri di calcolo.

---

<sup>47</sup> [https://phaidra.univie.ac.at/terms\\_of\\_use/show\\_terms\\_of\\_use](https://phaidra.univie.ac.at/terms_of_use/show_terms_of_use) [Ultimo accesso 25/09/2017]

<sup>48</sup> WEIGEL T., WITTENBURG P., *Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data*, 2017. Disponibile a:

<<https://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf>> [Ultimo accesso 12/09/2017]

### C. I servizi legali

L'attuazione di determinate politiche all'interno dell'istituzione per garantire dei risultati, è fondamentale per la creazione di servizi RDM. La loro formulazione non può prescindere dalla necessità di avere una rete di esperti legali da consultare onde evitare problemi di natura giuridica<sup>49</sup>. Un utente, trovandosi in una situazione spiacevole legata ai diritti su una determinata risorsa, può far ricorso ad un parere il quale viene fornito tramite la mediazione della biblioteca.

Le prime tematiche a cui viene naturale attingere nel momento in cui si parla di fattori legali sono normalmente la privacy e la gestione dei dati personali. Queste vengono fuori anche nel campo della ricerca, ad esempio, nel momento in cui questa si basa su sondaggi personali. Spesso le stesse istituzioni promotrici di tale ricerca adottano politiche piuttosto restrittive, facendo sì che i risultati non possano essere condivisi apertamente. Phaidra si basa idealmente sui concetti di Open Science, Open Access e Open Data, quest'ultimo così definito: "Open data and content can be freely used, modified, and shared by anyone for any purpose."<sup>50</sup> Bisogna considerare attentamente quindi come le norme sul diritto d'autore riguardino i dati digitali e che opzioni ci possano essere nel momento in cui si decida di condividere i dati liberamente. È un aspetto molto importante in quanto il riuso è uno dei principi FAIR da soddisfare e i produttori di dati devono avere questa possibilità. Come definito dai principi, sia dati che metadati devono essere creati con una chiara e accessibile licenza d'uso.

Per capire che tipo di licenze è possibile utilizzare, è fondamentale capire per prima cosa quali diritti sono in gioco. Gli oggetti digitali sono piuttosto complessi in relazione al copyright, in quanto possono essere in molti diversi formati e legati a diverse discipline; possono essere numeri, testi, immagini, ecc. In Phaidra, i diritti degli autori, dei fornitori e di terzi, sono protetti dalle *condizioni d'uso*<sup>51</sup>, redatte e concordate assieme agli esperti legali. Queste, accettate tramite l'accesso alla piattaforma, stabiliscono i rapporti giuridici presenti, in particolare quello con l'utente finale è regolato dall'utilizzo delle licenze. I produttori di dati sono normalmente incoraggiati ad usare le licenze Creative Commons, in quanto più adatte a perseguire gli scopi della ricerca. Poi sarà loro responsabilità assicurare l'osservanza del diritto d'autore e dello sfruttamento da parte di terzi, insieme ad un uso consapevole come indicato nel sottocapitolo 2.1.

Prima di pensare alla licenza da utilizzare per il riuso, è molto importante controllare che i dati siano legalmente utilizzabili e di conseguenza non ci siano implicazioni per la privacy, per la sicurezza o per la confidenzialità. Se vogliamo perseguire l'obiettivo di un ampio riuso, dobbiamo evitare licenze che possano restringere certi campi d'azione come ad esempio quella contro la creazione di materiali derivati. Quelle che richiedono solo il riconoscimento della paternità della risorsa da parte dell'autore

---

<sup>49</sup> LEARN PROJECT, *20 RDM Best-Practice Recommendations*, 2017. Disponibile a: <<http://learn-rdm.eu/wp-content/uploads/20-RDM-Policy-Recommendations.pdf>> [Ultimo accesso 12/09/2017]

<sup>50</sup> <http://opendefinition.org/> [Ultimo accesso 25/09/2017]

<sup>51</sup> [https://phaidra.univie.ac.at/terms\\_of\\_use/show\\_terms\\_of\\_use](https://phaidra.univie.ac.at/terms_of_use/show_terms_of_use) [Ultimo accesso 25/09/2017]

nulla tolgono all'obiettivo di un completo riutilizzo.

In questo breve elenco ci sarà una breve spiegazione delle tipologie di licenza<sup>52</sup>:

- ◆ **Creative commons**: si tratta di una tipologia di licenza nata 15 anni fa con lo scopo di trovare una buona soluzione relativa alla problematica della condivisione di contenuti mantenendo il diritto d'autore<sup>53</sup>. Si ispira al modello copyleft già esistente da prima in ambito informatico. È un modello equilibrato tra le due opzioni più estreme ossia, “all right reserved”, garante del copyright completo e “no right reserved”, che si riferisce al pubblico dominio. In questo caso vengono offerte 6 diverse licenze (più la CC0 che riguarda il pubblico dominio) che coprono alcuni diritti, a seconda del tipo di protezione che l'utente sceglie e per questo si può definire come “some right reserved”. Se il produttore di dati volesse condividere quanto creato, mantenendo con le dovute garanzie la paternità dell'oggetto o di determinate sue peculiarità, si tratta di una buona scelta. Recentemente è stata sviluppata la versione 4.0, che include un nuovo livello di protezione chiamato “sui generis” relativo ai database. Alla base di questa licenza, raccomandata da Phaidra grazie alle possibilità che offre per il riuso, stanno le seguenti libertà: di condividere, ossia di copiare, distribuire e trasmettere l'opera e di rielaborarla e/o riadattarla. Le clausole di utilizzo sono quattro e individuano una condizione particolare a cui il fruitore deve sottostare per usufruire dell'oggetto; queste possibilità intrecciate tra loro, arrivano a creare le sei licenze sopracitate. Le condizioni in questione sono: l'*attribuzione* (viene richiesto che l'autore sia sempre indicato secondo le modalità da lui richieste), la *non-commerciale* (permette che l'opera venga copiata, rielaborata e condivisa solo per scopi non commerciali), la *non opere derivate* (i contenuti possono essere copiati e condivisi, ma non rielaborati) e la *condividi allo stesso modo* (i derivati dell'opera possono essere condivisi solo con la stessa licenza o compatibile).
- ◆ **GNU licence**: licenza fortemente copyleft, ha l'obiettivo di garantire la libertà di condividere e modificare l'oggetto e di fare in modo che esso rimanga libero per tutti gli utenti.
- ◆ **No licence**: Se non viene scelta una licenza, la risorsa può essere utilizzata solo nei limiti del diritto d'autore. Inoltre, non è permesso alcun sfruttamento commerciale come la copia, la distribuzione, la trasmissione, la fornitura e l'elaborazione protezione.
- ◆ **Public Domain Trademark 1.0**: si utilizza per la marcatura di beni o servizi pubblici. Nel contesto di beni creativi, esiste un consenso nel caso in cui una creazione non sia soggetta a restrizioni di copyright. Questo è il caso nel quale un pezzo creativo non soddisfa i requisiti imposti dal diritto d'autore su un lavoro protetto, ad esempio perché non è sufficientemente

---

<sup>52</sup> <http://datamanagement.univie.ac.at/en/rdm/legal/licensing-models/> [Ultimo accesso 25/09/2017]

<sup>53</sup> ALIPRANDI S., *Creative commons: manuale operativo. Guida all'uso delle licenze e degli altri strumenti cc*, Ledizioni/Copyleft-Italia.it, 2008.

individuale o non può essere classificato in una categoria di copyright. Vi è anche la libertà di utilizzo quando il periodo di protezione di un lavoro è scaduto. Questo accade solitamente in materia di diritti d'autore 70 anni dopo la morte dell'autore. Il Public Domain Trademark è destinato a beni creativi per i quali non ci sono più diritti di proprietà intellettuale o che sono stati precedentemente pubblicati nel dominio pubblico globale da parte del proprietario del copyright. Non dovrebbe essere utilizzato se la creatività in questione appartiene al dominio pubblico solo in determinati ordini legali, vale a dire in pubblico, mentre in altri è ancora protetta.<sup>54</sup>

Vedremo poi le applicazioni più pratiche e il processo di scelta della licenza nel corso del capitolo 4, quando verranno descritte e visualizzate su interfaccia le fasi di lavoro .

---

<sup>54</sup> <http://datamanagement.univie.ac.at/en/rdm/legal/licensing-models/public-domain-trademark-10/> [Ultimo accesso 25/09/2017]



### Capitolo 3 - La struttura interna di Phaidra

Phaidra è una piattaforma che si occupa di dati della ricerca e di varie tipologie di documenti digitali, che coprono non solo tutta l'area di interesse di enti come una biblioteca e un archivio, ma anche tutti i campi di lavoro di istituzioni scientifiche, come per esempio un intero Ateneo.

La piattaforma Phaidra non va considerata come un mondo a sé stante, che ruota solo attorno alle entità descritte dell'università di Vienna. È inserita all'interno di un ecosistema digitale universitario, il quale rientra nei canoni previsti per la “European Open Science Cloud”.

Considerando questi presupposti, c'è tutto un percorso da fare per addentrarsi passo dopo passo nella sua struttura interna. Tra la gestione e l'architettura del sistema ci sono le chiavi che lo rendono efficiente, quindi l'obiettivo sarà vedere come è stato progettato strutturalmente e come funzionano i processi lavorativi che si attivano al suo interno. Nella figura 4 si può notare la posizione di Phaidra tra l'ecosistema digitale dell'Università di Vienna e l'European Open Science Cloud, ad indicare quello che diventerebbe un mezzo di connessione per la condivisione di tutti i dati contenuti nelle istituzioni di ricerca del continente. Inoltre si può vedere come il trapezio che indica la piattaforma sia diviso in tre livelli: governance, servizi e infrastrutture. La base, più grossa, è posizionata sull'infrastruttura ad indicare il punto di forza. Nel livello dei servizi invece c'è una casella ad indicare le immagini, le quali nel quinto capitolo saranno tema di studio specifico.

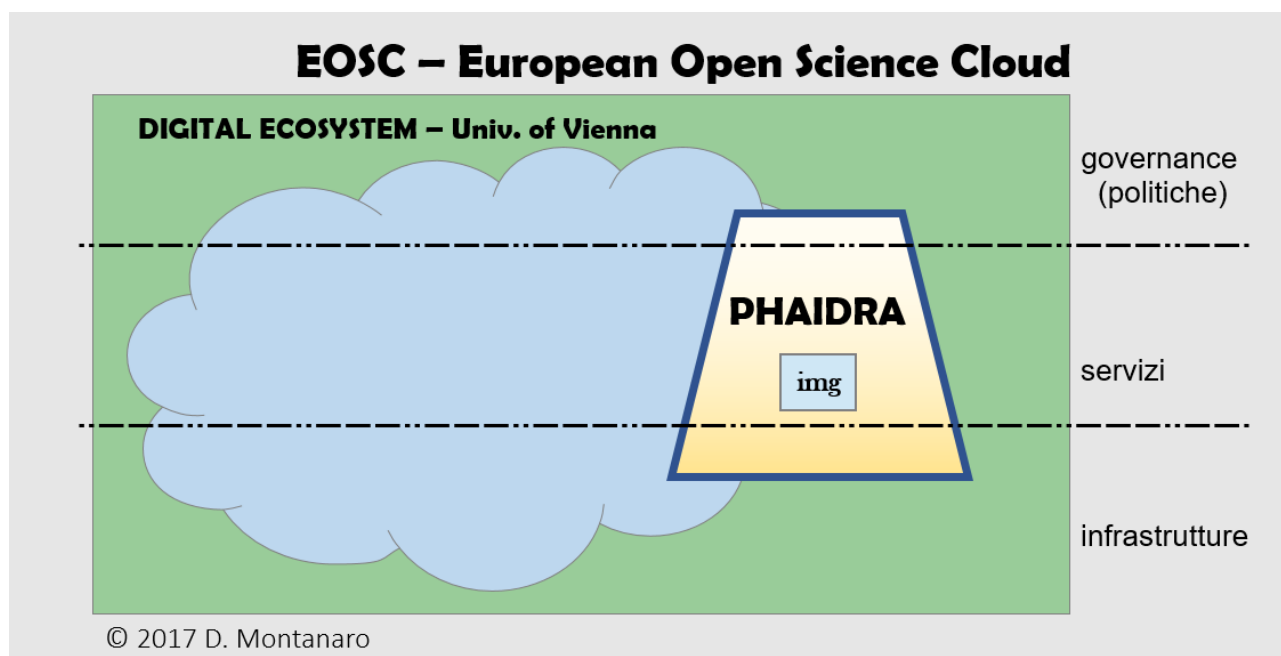
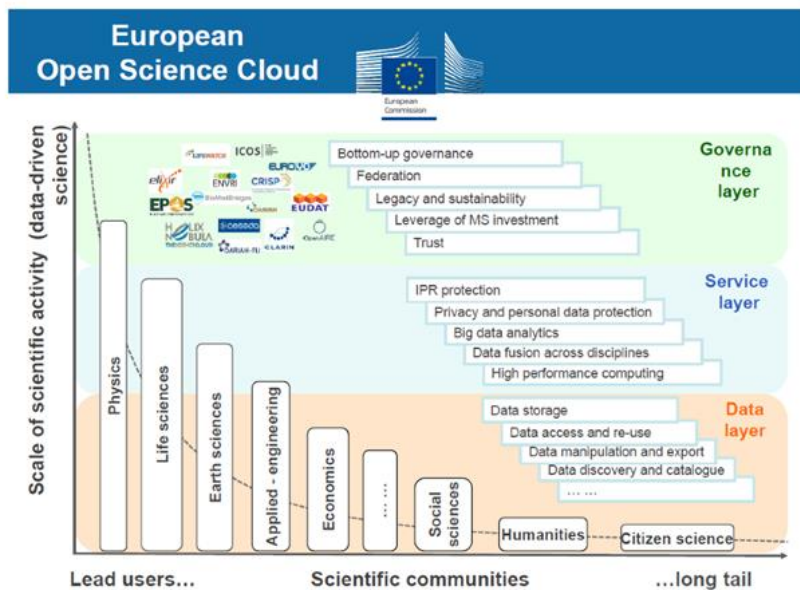


Figura 4 – Rappresentazione di Phaidra all'interno dell'EOSC

Il grafico di J.C. Burgelman, a seguire, illustra le prospettive in seguito alla futura creazione dell'European Open Science Cloud.

# European Open Science Cloud



Presentation “Open Science policy: Results of the consultation on ‘Science 2.0: Science in transition’ and possible follow up” by J.C. Burgelman, June 3 2015 at e-IRG workshop



Paolo Budroni, Raman Ganguly- Towards the European Open Science Cloud, January 2017

Figura 5 – Grafico J.C. Burgelman



### 3.1 La gestione del sistema:

Nell'istante in cui si pensa alla gestione di un sistema come Phaidra, risulta naturale che debba esserci dietro una programmazione ben fatta e uno sguardo che va oltre gli interessi della singola università viennese. A tal proposito, sono state sviluppate da una commissione europea delle linee guida sulla gestione dei dati<sup>55</sup>, le quali parlano della necessità del DMP (Data Management Plan): si tratta di un documento formale, che va redatto nel momento in cui si dà vita al progetto per fissare gli obiettivi, verificarli al termine e capire le proprie possibilità preventivamente. Si parla quindi, ad esempio, di saper programmare e di prevedere nel limite del possibile le risorse finanziarie a disposizione; ma anche di avere chiaramente pianificato la tipologia di dati su cui si va a lavorare, la loro accessibilità, i collegamenti, l'archiviazione e la preservazione. Vedremo dunque come funziona e cosa sta dietro alla gestione di un progetto come Phaidra e le sue attività.

*“... Data Management Plans (DMPs) detailing what data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.”*

[European Commission, *Guidelines on Data Management in Horizon 2020*]

*“Research Data should be open by default, so long as the RDM policy provides an opt-out clause. The policy needs to stipulate reasons for an opt-out to be considered (e. g. Privacy, confidentiality, legal or commercial reasons). These considerations should be assessed at the beginning of the research project as part of a Data Management Plan. The policy should be 'permissive', allowing any number of different approaches, including educational re-use.”*

[LEARN, *20 RDM Best-Practice Recommendations*]

La gestione dei dati di ricerca ci appare come un processo molto complicato a causa delle molte diverse sfaccettature che i contenuti possono assumere. Ogni progetto di ricerca però, ha una sua fase di studio preliminare la quale è volta a massimizzare la resa dello stesso. A questo fine, all'interno delle commissioni di vari progetti europei come EOSC, LEARN, E-Infrastructures, ..., si è molto discusso sulle modalità per uniformare e sfruttare al meglio le ultime evoluzioni nell'RDM. Un fattore di enorme rilevanza riguarda l'Open Access: rendere il più possibile le pubblicazioni scientifiche ad accesso aperto, ossia disponibili apertamente senza restrizioni e depositate all'interno di un repository di ricerca. Chiaramente qualche restrizione ci sarà sempre, ma più sono limitate, più si aprono ampie possibilità di riuso e progresso. In più, a partire dal 2017, è diventata obbligatoria la redazione di un Data Management Plan per ogni progetto di ricerca finanziato da Horizon2020, dove vanno messe

---

<sup>55</sup> EUROPEAN COMMISSION, *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*, 2016.

per iscritto motivazioni legate alle limitazioni dell'Open Access<sup>56</sup>.

Un Data Management Plan (DMP) è una linea guida strutturata che descrive l'intero ciclo di vita dei dati, durante la ricerca e dopo averla completata e che può essere aggiornato se necessario. Indica quali tipologie di dati verranno create e illustra i piani di condivisione e conservazione, anche tramite un confronto con la natura dei dati e le loro restrizioni. Può essere un documento oppure uno strumento online. I Data Management Plans devono assicurare che i dati della ricerca siano rintracciabili, disponibili, autentici, citabili, adeguatamente memorizzati e che rispettino parametri legali chiaramente definiti e misure di sicurezza appropriate per la gestione dell'uso. Idealmente, i DMPs dovrebbero essere consegnati e funzionare in un formato operativo su macchina. Assicurano che i dati siano al sicuro per il presente e per il futuro.

È importante non scordarsi come la condivisione dei dati possa portare ad enormi benefici nel mondo della ricerca; una sana consapevolezza, trasmessa attraverso uno strumento come i DMP, può rendere possibili molti passi avanti. Per questo è stato pensato, proposto e diventato praticamente obbligatorio per tutte le comunità di ricerca. È infatti diventato un requisito fondamentale per ricevere finanziamenti da parte di diverse istituzioni come la commissione europea per le proposte di Horizon2020 o la National Science Foundation negli Stati Uniti. Inoltre è richiesto da alcune università e relative biblioteche e da alcuni editori<sup>57</sup>.

All'interno dei regolamenti di Horizon2020 ad esempio, il DMP deve essere aggiornato nelle varie fasi del progetto (inizio, metà, fine) e la prima versione deve essere per forza consegnata entro i primi sei mesi dalla sua partenza. Non viene prescritta alcuna forma specifica dalla commissione però viene proposta DMPOne. Possono venire anche utilizzati altri modelli se vengono rispettati i requisiti. Nella maggior parte dei casi, un DMP si crea compilando un modello di documento oppure rispondendo alle domande di una checklist già impostata. C'è anche la possibilità di utilizzare alcuni strumenti software dedicati dove gli utenti scelgono il modello appropriato dei finanziatori. Successivamente vengono presentate domande relative solo all'orientamento di pertinenza e il risultato che viene creato può essere esportato o inviato direttamente. Due strumenti creati a questo scopo sono: DMPTool<sup>58</sup> e DMPOne<sup>59</sup>.

---

<sup>56</sup> EUROPEAN COMMISSION, *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*, 2016.

<sup>57</sup> MIKSA T., SCHRAUF C., SÁNCHEZ SOLÍS B., *Data Management Plans (DMP)*, 2016. Disponibile a: <<http://phaidra.univie.ac.at/o:441308>> [Ultimo accesso 12/09/2017]

<sup>58</sup> <https://dmp.Cdlib.org> [Ultimo accesso 25/09/2017]

<sup>59</sup> <https://dmponline.dcc.ac.uk/> [Ultimo accesso 25/09/2017]

Vediamo ora che soggetti andranno a beneficiare del DMP<sup>60</sup>, il quale non è considerabile come un ulteriore lavoro da imporre ai ricercatori. Anzi, si può affermare che assieme ai committenti dei progetti, agli enti finanziatori e ai valutatori di ricerca, agli archivi e ai soggetti che usufruiscono del ri-uso e alla società nel suo complesso abbiano una serie di vantaggi che specificheremo nelle prossime righe.

I *ricercatori* infatti, potranno risparmiare notevoli quantità di tempo dato che in questo modo i dati sono organizzati bene fin dall'inizio, è necessario uno sforzo minore per la pubblicazione degli stessi e c'è un minore rischio nella perdita dei contenuti. Viene aumentata l'efficienza in quanto risulta più facile localizzare i dati più appropriati per ogni situazione ed è più facile condividerli con gli altri consentendo maggiori contenuti. Oltre a questo la ricerca può avere più credito e visibilità grazie al fatto che i dati, così documentati, sono più facili da comprendere e quindi da riutilizzare da parte di terzi. Questo permette al ricercatore di guadagnare maggiore credito come fornitore di dati. I repositories guadagnano velocità nel processo di caricamento, perché i metadati sono già stati forniti e il formato dei dati è conosciuto in quanto già precedentemente deciso. Inoltre i problemi etici ed i diritti legali sono già chiariti in partenza, evitando questioni relative alla possibilità di divulgare o meno certi dati, a quali azioni di conservazione possono essere eseguite o alla durata della memorizzazione.

I *committenti*, i quali non sono sempre i soggetti che stanno conducendo la ricerca, possono stimare al meglio i costi; sia che si tratti di spese per l'archiviazione e il backup, sia che si tratti di licenze per i software. Nel momento in cui viene pianificata l'archiviazione a lungo termine, può assegnare le varie responsabilità e predisporre il denaro per coprire le spese.

Sempre i committenti, assieme ai *valutatori di ricerca* possono riporre maggiormente la loro fiducia nei risultati che sono reali e credibili. Con ciò la cooperazione tra progetti e scienziati può risultare rafforzata.

Gli *utenti* e la *società* nel suo complesso hanno la possibilità di avere maggiore fiducia nelle affermazioni degli scienziati; infatti avendo i dati è più facile fare una verifica, anche se non necessariamente alla portata di tutti. Oltre a questo, alcuni sforzi passati possono tornare utili, grazie al riutilizzo dei dati.

Come affermato nelle ultime pagine, il DMP è stato reso praticamente obbligatorio tramite l'evidenza della sua importanza messa a nudo da numerosi incontri affrontati dagli esperti del campo. A partire da queste esperienze, l'università di Vienna si è messa al lavoro per tenersi aggiornata con le nuove disposizioni, arrivando a creare nel corso del 2015 un proprio modello di DMP<sup>61</sup>. Questo è stato

---

<sup>60</sup> MIKSA T., SCHRAUF C., SÁNCHEZ SOLÍS B., *Data Management Plans (DMP)*, 2016.

<sup>61</sup> BUDRONI P., BLUMESBERGER S., PAUSZ R., PREZA J. L., SÁNCHEZ SOLÍS B., *Data Management Plan for*

messo a punto per i progetti all'interno dell'ateneo e del suo ecosistema, ovviamente non dimenticando di considerare Phaidra come repository consigliata. A seguire si può osservare le varie sezioni in cui è stato diviso e organizzato il documento redatto.

- **Administrative Data:** l'obiettivo di questa sezione è quello di fornire informazioni di base sul progetto di ricerca, permettendo l'identificazione del progetto stesso e dei suoi responsabili; questi ultimi devono poter venire contattati. Non è una sezione dove vengono ripetute informazioni riguardo al progetto che possono venire trovate in altri documenti.
- **Data Collection:** lo scopo è di identificare i set di dati che sono stati creati ed usati durante il progetto. La descrizione non è limitata solo ai dati che necessitano successivamente di essere archiviati e conservati, come poi sarà specificato nella sezione "Selection and Preservation". Attraverso l'identificazione dei dati utilizzati nel corso del progetto, i ricercatori potranno stimare più precisamente i requisiti di cui l'infrastruttura e il relativo software hanno bisogno per funzionare al meglio.
- **Documentation:** La documentazione si propone come sezione per descrivere le pratiche attive durante il processo di ricerca, al fine di facilitare la corretta interpretazione e la provenienza delle collezioni di dati creati durante la stessa. Alcuni di questi processi possono essere: il contesto e la sua interpretazione, la struttura dei dati, le relazioni con altre entità, che cambiamenti e procedure sono necessarie per preparare e analizzare le risorse. Una buona documentazione dovrebbe indirizzare verso le seguenti domande: Perché, chi, cosa, dove, quando e come.
- **Metadata:** Più i metadati sono descritti in maniera dettagliata, più sarà facile dopo trovare e riutilizzare gli oggetti digitali quando questi vengono depositati in un sistema di archiviazione. Vocabolari standardizzati e classificazioni (come ÖFOS, Eurovoc, ACM or Getty) aiutano a rendere i dati più visibili e riutilizzabili.
- **Ethics and Legal Compliance:** L'obiettivo di questa sezione relativa a etica e norme legali è quello di identificare come i dati sensibili possano venire processati, depositati e pubblicati.
- **Storage and Backup:** in questo caso il fine consiste nella descrizione della messa in sicurezza dei dati nel corso del progetto. Si focalizza sulle azioni che assicurano che non ci siano perdite di contenuti e che gli accessi avvengano solo da parte di soggetti autorizzati. In questa sezione ci si occupa solo di queste operazioni nella fase di gestione durante il progetto e non dopo.
- **Selection and Preservation:** Sezione che fornisce informazioni sui dati che hanno la

necessità di essere conservati a lungo termine, come la messa in sicurezza dopo la fine del progetto. Il concetto chiave di queste informazioni non è così diverso da quello descritto nella sezione “Data Collection”. Anche in questo caso i ricercatori dovrebbero stimare come queste operazioni possano venire finanziate e quanto potrebbe costare nel complesso. In questo dovrebbero essere aiutati dai repositories dove decidono di depositare gli oggetti digitali.

- **Data sharing:** La data sharing decide quali dati, come ed in che forma possono venire condivisi con altri stakeholders o sistemi (es. Europeana). Le decisioni che vengono prese in questa sezione sono influenzate da quelle presenti in Ethics and Legal Compliance.
- **Responsibilities and Resources:** L'obiettivo di quest'area è quello di identificare quali soggetti sono responsabili dall'implementazione del DMP. Oltre a questo, riassume le risorse extra richieste per portare a termine il piano, come ad esempio quelle necessarie per il caricamento dei dati all'interno di un repository particolare (risorse umane, infrastrutture, denaro, tempo).

### 3.2 L'architettura del sistema:

Phaidra si basa sul software open source Fedora, il quale è molto flessibile e garantisce l'interoperabilità. Basandosi su un'architettura Repository ha il vantaggio di offrire una gestione rapida ed efficiente di una mole di dati molto voluminosa. Questi dati, facendo parte di un sistema relazionale a tabelle sono collegati tra loro rimanendo parzialmente indipendenti, mentre la gestione di sicurezza e backup rimane completamente centralizzata.

La "European Science Agenda", individua tre diversi livelli da utilizzare per la scienza dei dati: dati (data), servizi (services) e governance<sup>62</sup>. Questi vengono coperti da infrastrutture digitali e costituiscono la base per la conservazione dei dati, in quanto i dati vengono gestiti e curati grazie all'ausilio di infrastrutture. Tenendo quindi l'infrastruttura come fondamento e seguendo la proposta della commissione europea di utilizzare i tre livelli, arriviamo alla costruzione di servizi per il caricamento ed il ri-uso. Questi servizi dovrebbero essere semplici da utilizzare e appropriati per ogni possibilità di utilizzo da parte del produttore di dati. Infine, la governance è il quadro attraverso il quale, con l'uso di politiche funzionali e pubblicamente esplicitate, si fornisce un formato istituzionale per la conservazione dei dati.

Vanno quindi messe in atto una serie di strategie con l'obiettivo di costruire un ecosistema digitale di una certa qualità. Vediamo ora nel dettaglio di cosa si occupano i livelli:

- **Il livello "Dati":** All'università di Vienna, si è iniziato con il creare infrastrutture per i servizi speciali e per i sistemi speciali di archiviazione. Il team ha cominciato con Phaidra, puntando quindi all'archiviazione a lungo termine di dati generici, i quali possono essere memorizzati anche se di diversa natura. In questa repository, venivano direttamente amministrati i metadati e i vocabolari controllati. Per definizione, Phaidra fornisce un identificativo persistente per tutti i dati, che non possono essere eliminati. Secondo certi requisiti, alcune tipologie di dato dovrebbero avere la possibilità di essere eliminati dopo un determinato periodo di tempo. A questo scopo è stata istituita una seconda repository dove i dati possono venire cancellati e, se ritenuto opportuno, trasferiti verso l'archivio a lungo termine. Attualmente non è possibile stabilire un tempo massimo entro il quale gli oggetti di questa repository a medio termine possano venire cancellati. Tale caratteristica richiede uno strumento di pianificazione di livello, legato ad una politica correlata. Esiste anche l'opportunità di usufruire di un repository per testare i dati, in modo che gli utenti possano eseguire dei test di qualità. Viene chiamata "Sandbox" ed è fondamentalmente un clone dell'archivio a lungo termine.

Esiste un ulteriore sistema di Repository, chiamato Unidam, il quale è stato creato da due

---

<sup>62</sup> GANGULY R., BUDRONI P., SÁNCHEZ SOLÍS B., "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud." In *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, ELPUB. Limassol, Cyprus, 2017, pp. 203-210.

facoltà viennesi ed è stato recentemente integrato nell'infrastruttura centrale di gestione dei dati. Questo fornisce agli utenti dei repositories la possibilità di ottenere maggiori funzionalità per i loro dati, in particolare nel campo delle “digital humanities”. Sulla base dell'indagine nazionale “Researchers and Their Data. Results of an Austrian Survey (2015)”<sup>63</sup>, la quale è stata rivolta a quasi tutti i ricercatori austriaci (36000 persone), è stato possibile individuare che quasi il 25% dei progetti di ricerca utilizza software sviluppati durante il processo. Guardando il software sviluppato e utilizzando ciò che sappiamo su repositories ben consolidate a questo scopo, è stato possibile implementare un repository Github Enterprise<sup>64</sup> per tale ricerca e integrarla nell'ecosistema. Ciò consente ai dati di essere collegati a una release software, che potrebbe anche essere identificata da un identificatore persistente.

- **Il livello “Servizi”**: A livello di servizi, il riutilizzo è il valore più rilevante. Per questo, nel momento in cui è stata riorganizzata l'architettura di Phaidra, è stata aggiunta una API (application programming interface) - interfaccia di programmazione di un'applicazione, per consentire ad altre applicazioni di collegarsi a Phaidra. È stata una modifica che ha inoltre aiutato nell'integrare un server che permette di presentare immagini di grandi dimensioni sul web e un servizio streaming memorizzato nel repository per i materiali audio e video.

Esiste un'ulteriore sezione del livello che fornisce strumenti per la gestione dei dati. È stato implementato un server terminologico (terminology server), basato sullo standard SKOS (Simple Knowledge Organisation System)<sup>65</sup> il quale, basandosi su RDF, è un'area di lavoro che sviluppa specifiche per supportare l'utilizzo di sistemi di organizzazione della conoscenza (KOS) quali thesauri, schemi di classificazione, sistemi di intestazione soggettiva e tassonomie nell'ambito del Semantic Web. In questo modo gli utenti hanno la possibilità di scegliere vocabolari controllati sfruttando una gamma di scelta più ampia. Un server di supporto inoltre (Handle server) crea gli identificatori persistenti in tutto l'ecosistema digitale, consentendo un riferimento tra oggetti coerente.

- **Il livello “Governance”**: Si tratta di un livello di rilievo nell'operazione di progettazione degli ecosistemi, ma non ne è l'attore principale.

Nel corso del capitolo 1.2 è stato accennato il processo di riprogettazione avvenuto durante l'evoluzione della piattaforma, il quale, alla luce delle problematiche uscite in seguito ad un uso sempre più massiccio, ha cambiato completamente tipo di approccio. Il passaggio principale dunque

---

<sup>63</sup> Disponibile a: [https://phaidra.univie.ac.at/detail\\_object/o:409473](https://phaidra.univie.ac.at/detail_object/o:409473) [Ultimo accesso 25/09/2017]

<sup>64</sup> <https://github.com/phaidra> [Ultimo accesso 25/09/2017]

<sup>65</sup> <https://www.w3.org/2004/02/skos/> [Ultimo accesso 25/09/2017]

è stato quello che ha portato da una singola, semplice repository, al concetto di Eco-sistema digitale basato sulle relazioni possibili all'interno dei repositories che lo compongono. Basandosi sulle tracce lasciate dal confronto con gli stakeholders, sono stati identificati tre diversi modelli che fanno da guida nei processi di ricerca per i dati. Il *primo modello (Data Lifecycle)* si focalizza principalmente sul ciclo di vita dei dati; il *secondo (Digital Workflow)* si occupa del caricamento del materiale all'interno del sistema di archiviazione, descrivendolo attraverso un workflow digitale, e della successiva disponibilità di questo materiale per il riutilizzo. Nel momento in cui la gestione dei dati viene implementata sin dall'inizio, il riutilizzo degli stessi è già incluso come passo successivo all'interno del ciclo di vita della risorsa. Il *terzo modello* parte dal principio che nessun sistema può essere perfetto per ogni diversa tipologia di dato. Da qui la parola chiave nel descriverlo, *Data Evaluation*, in quanto ha come scopo quello di guidare ad una corretta valutazione dell'oggetto in vista della sua acquisizione nel sistema di archiviazione più adatto. Ma vediamoli ora un po' più in dettaglio:

- ◆ **Data Lifecycle model:** Quando i dati vengono pubblicati il loro volume è generalmente piccolo ed esistono già formati di archiviazione appropriati. Nonostante ciò questa è solo la punta dell'iceberg, ossia la parte visibile, come risulta evidente quando si guardano i dati nel processo di ricerca. Infatti, il valore delle pubblicazioni dipende dalla loro corretta conservazione, come afferma il rapporto PARSE Insight<sup>66</sup>: “Digital preservation of research data here means the careful storage of all research output in such a way that it remains accessible, usable and understandable over the long term.”

Come mostra l'illustrazione a seguire però, c'è tutto un mondo che va oltre i dati pubblicati. Si tratta, in questo caso, di una lettura della questione che avviene dal punto di vista dei dati anziché da quello delle pubblicazioni. A partire dai dati grezzi c'è tutta una selezione che passa attraverso risultati inconclusivi e/o negativi, i quali venivano considerati inutili e scartati. All'interno di Phaidra è stato creato questo modello basato sulla “Data Publication Pyramid”, che permette di includere i risultati non inclusi nelle pubblicazioni e renderli in questo modo disponibili e utili agli stakeholders che ne potrebbero avere bisogno.

---

<sup>66</sup> [http://libereurope.eu/wp-content/uploads/PARSE-Insight\\_D3-5\\_InterimInsightReport\\_final.pdf](http://libereurope.eu/wp-content/uploads/PARSE-Insight_D3-5_InterimInsightReport_final.pdf) [Ultimo accesso 29/08/2017]



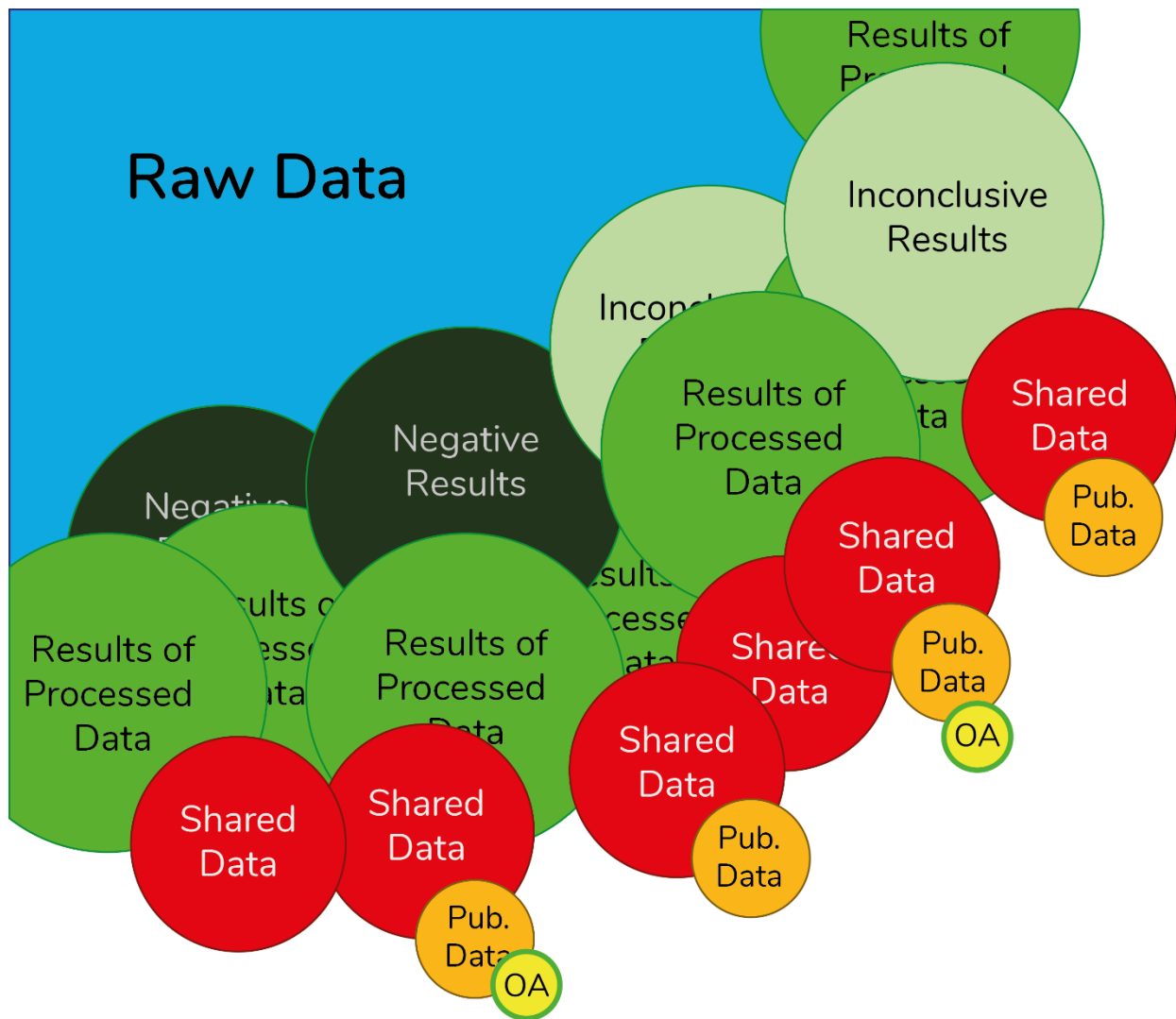


Figura 6 – Ciclo di vita dei dati

CC BY-NC-SA 4.0 International Raman Ganguly. <http://phaidra.univie.ac.at/o:387240>

- ◆ **Digital workflow model:** Si tratta del modello che più di tutti racchiude gli ideali dell'ecosistema digitale ed è basato sul modello di riferimento OAIS. I concetti espressi da questo modello sono molto semplici; la sua struttura prevede gli archivi messi al centro e circondati da produttori, consumatori e gestori. In questo caso però si tratta di una descrizione più accurata e specifica della gestione dei dati rispetto al modello OAIS, il quale come già accennato è molto generico. Avendo come parti in causa il produttore di dati, il gestore dell'archivio e il riutilizzatore delle risorse, il modello definisce come e quando i dati verranno trasferiti da un utilizzatore all'altro.

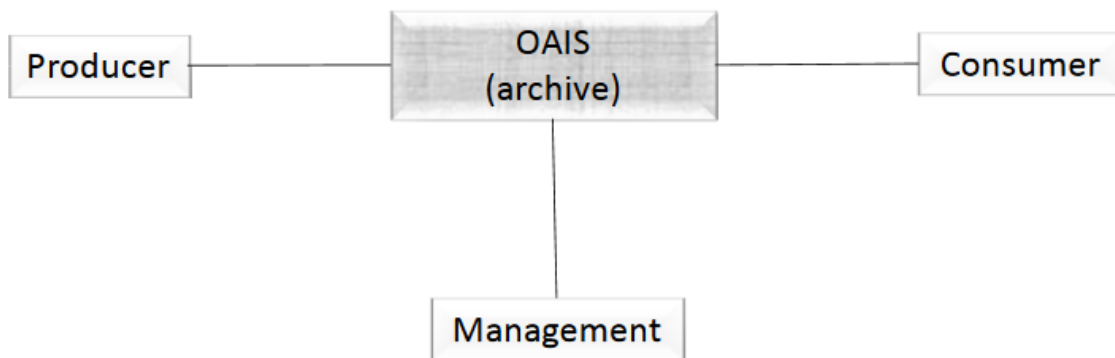
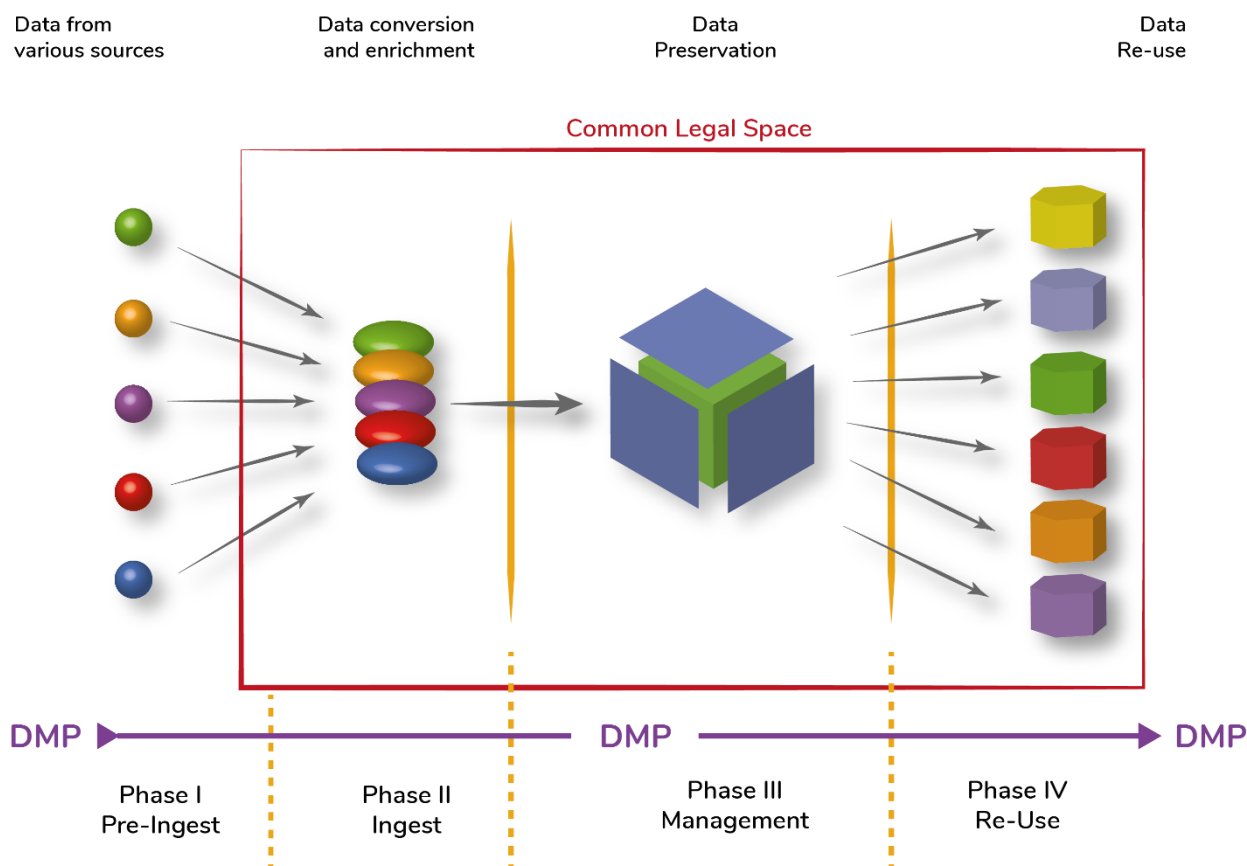


Figura 7 – Modello OAIS

Secondo il modello e in accordo con le condizioni di utilizzo dell'archivio digitale, le quali si riferiscono a tutte le fasi di vita del documento digitale, il produttore di dati è la parte in causa che, per l'appunto, crea e poi possiede l'oggetto. Fa parte del suo compito stabilire, nel corso della fase preparatoria, la qualità, la durata, il contesto e in che modo risorse e relativi metadati potranno venire nuovamente utilizzati. Per rispondere nel modo più esaustivo possibile ai dubbi a cui questa operazione può portare, è bene utilizzare e fare riferimento al *Data Management Plan*, strumento chiave per non perdere di vista le finalità dell'intero progetto. Infatti si tratta di un piano progettuale per i dati all'interno dell'ecosistema, in modo che ognuno di essi sia sempre aggiornato. L'unico requisito è un output leggibile dalla periferica a partire dallo strumento di gestione dei dati.

Le informazioni provenienti dal produttore di dati sono fondamentali ai fini della gestione; questa infatti si occupa di mantenere una certa qualità per un tempo specifico e assicura che solo altri utenti autorizzati possano accedervi. Durante il processo di caricamento c'è un trasferimento di contenuti e responsabilità dal produttore (producer) al gestore (management). Successivamente ci potrà essere un nuovo trasferimento di responsabilità, nel momento in cui c'è un riutilizzatore (consumer) che riceve questo onere. Questo però deve avere bene chiari i metodi consentiti di riutilizzo, per questo le licenze devono assolutamente essere fornite al consumer insieme all'oggetto digitale.




 Digital workflow model for data management  
Raman Ganguly, University of Vienna 2017

Figura 8 – Modello di Digital Workflow

<http://phaidra.univie.ac.at/o:527220>

I principi FAIR (Findable, Accessible, Interoperable, Re-usable)<sup>67</sup> dovrebbero fungere da guida per il riutilizzo dei dati e dovrebbero essere rispettati, lavorando bene, sin dalla fase di caricamento, mentre si verificano conversioni ed arricchimento dei dati. Il modello così rappresentato mette in mostra quanto appena descritto nel corso del paragrafo. Si può notare, all'interno del rettangolo rosso, il *common legal space*, ossia un'area all'interno della quale sono contenuti i dati della piattaforma. Si può considerare una specie di area Schengen per gli oggetti contenuti che vengono preservati e gestiti al suo interno. Al suo interno si possono muovere liberamente senza barriere legali da un sistema all'altro, grazie alle comuni condizioni di utilizzo che, come già sottolineato prima, devono essere chiare già dal processo di caricamento. Il tutto viene completato, a livello istituzionale, dalla chiarificazione di politiche, governance, regole e termini d'uso per servizi e dati.

- ◆ **Evaluation of data:** Questo terzo modello è volto ad esaminare più specificatamente la gestione dei dati stessa e la valutazione sul dove depositare le risorse. Considerando che i dati

<sup>67</sup> <https://www.force11.org/group/fairgroup/fairprinciples> [Ultimo accesso 28/09/2017]

della ricerca possono essere molto diversi tra loro, è assodato che un solo archivio non può adattarsi perfettamente a tutte le tipologie di dato. Per questo va deciso dove archiviare gli oggetti con lo scopo di mantenere intatta la qualità e rendere disponibile il riutilizzo. Il modello quindi vuole assolvere al compito, tramite i suoi attributi, di aiutare nel valutare la complessità dei formati, la durata ideale di archiviazione, la quantità di dati. Questi attributi dovrebbe essere scritti e dichiarati nel Data Management Plan, tramite il quale si possono comparare con le caratteristiche del sistema.

È facile determinare la quantità di files e vedere quanto spazio occupano; più importante è però in questo caso capire se ci sono tanti piccoli files o solo pochi grandi. Questo fattore è fondamentale nel momento in cui va scelto un sistema di archiviazione. Bisogna infatti ricordare che lo spazio di archiviazione è costoso e non tutti i tipi di dati hanno la necessità di essere conservati a lungo termine. Per alcuni di questi, come magari alcune risorse per l'insegnamento, potrebbe essere sufficiente un tempo di salvataggio compreso tra i tre e i dieci anni. Anche quest'operazione però va correttamente pianificata ed eseguita. La complessità del formato dovrebbe essere esaminata dalla prospettiva di conservazione e riutilizzo dei dati. I file audio e video, ad esempio, sono più complessi dei documenti. Le banche dati e il software (oltre ai relativi metadati di contesto e provenienza) hanno esigenze particolari nella fase di riutilizzo. Come illustra il grafico sottostante, la valutazione dei dati e dei repositories possono essere inseriti su una griglia e confrontati. Si tratta di un metodo per trovare un sistema di archiviazione adatto per i dati.

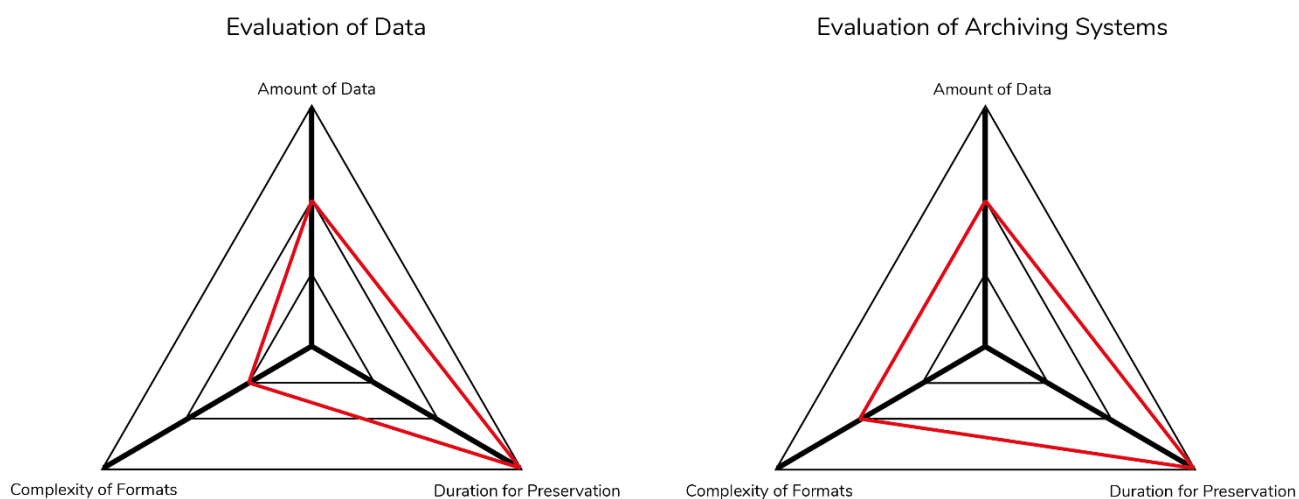


Figura 9 – Valutazione di Dati e Repositories

CC BY 4.0 International Ganguly Raman. <http://phaidra.univie.ac.at/o:527223>

### 3.3 La gestione elettronica dei processi lavorativi:

In questo sotto-capitolo, le attenzioni saranno puntate sulle dinamiche di interazione tra i tutti i soggetti coinvolti nello scambio di informazioni: persone, applicazioni, servizi. Cosa succede all'interno del software quindi, quando ad esempio qualcuno decide di intervenire caricando dei dati? L'obiettivo è di arrivare a comprendere queste dinamiche; le validazioni con i loro circuiti, le tempistiche, i compiti dei singoli soggetti.

Per comprendere al meglio i meccanismi che regolano i processi lavorativi, è necessario fare una piccola introduzione all'oggetto digitale o digital object (DO), in quanto si tratta dell'unità di base fondamentale all'interno del sistema; infatti essa non è composta solo dalla risorsa che poi verrà visualizzata e riutilizzata dall'utente, bensì da una serie di altri contenuti che permettono di garantire tra l'altro una facile identificazione e reperibilità dell'oggetto nel tempo. In precedenza, è già stato accennato l'uso di un *identificatore permanente* (PID), il quale permette di rendere il file sempre reperibile tramite un indirizzo che non cambia e l'utilizzo di *metadati*, i quali forniscono a tutti i soggetti che ruotano attorno al sistema un'importante via per descrivere nel migliore dei modi la risorsa, rendendola trovabile e ben collegata. Oltre a questo, come ben visibile nelle immagini a seguire, l'oggetto è composto anche da altri contenuti che lo rendono completo ai fini di un suo completo utilizzo in Phaidra.

#### Digital Object – Model

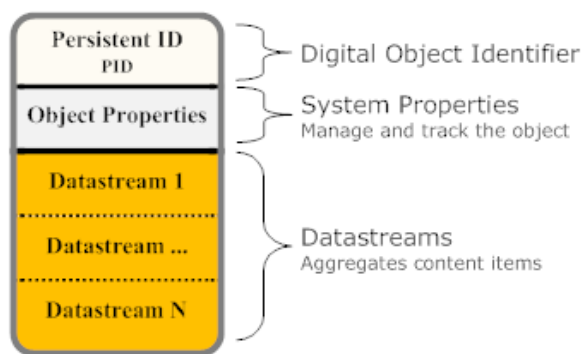


Figura 10 – Modello di DO

<https://github.com/phaidra/>

#### Digital Object - Example

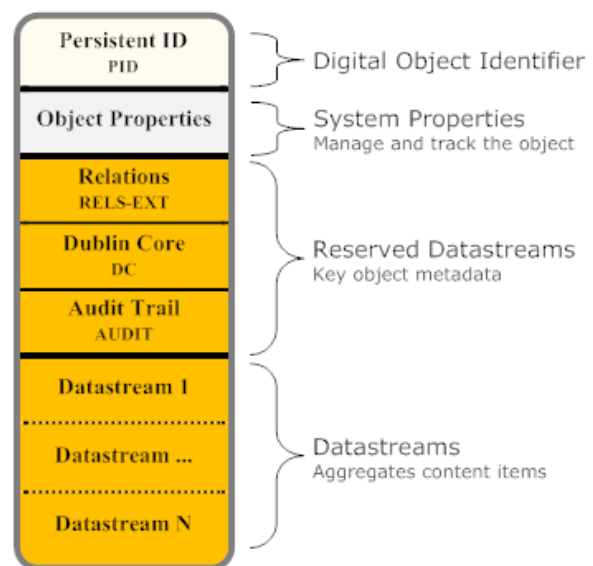


Figura 11 – Esempio di DO

<https://github.com/phaidra/>

Prima di chiarire in ogni suo aspetto la composizione degli oggetti digitali, bisogna fare un passo

indietro analizzando cos'è *Fedora*<sup>68</sup> e perché è stato scelto come base per costruire la piattaforma. Il nome è un acronimo che sta per “Flexible Extensible Digital Object Repository Architecture” ed è stato pensato come sistema di repositories open source robusto e modulare; questa sua caratteristica lo rende particolarmente adatto per le biblioteche e gli archivi digitali, sia per quanto riguarda l'accesso che per quanto riguarda la conservazione. È inoltre utilizzato per fornire un accesso specializzato all'enorme quantità di dati digitali che costituiscono collezioni di materiali storici e culturali, nonché pubblicazioni scientifiche. Ha una base di utenza molto sviluppata in tutto il mondo, all'interno della quale si possono notare varie organizzazioni del patrimonio accademico e culturale, università, istituti di ricerca, biblioteche universitarie, biblioteche nazionali e agenzie governative. Fedora definisce un modello generico per gli oggetti digitali, il quale può essere utilizzato per fornire le caratteristiche essenziali per molti tipi di contenuti digitali, inclusi documenti, immagini, libri elettronici, oggetti di apprendimento multimediale, collezioni di dati, metadati e molti altri. Questo modello di oggetti digitali è un elemento fondamentale dell'architettura Content Model e di tutte le altre funzionalità fornite da Fedora. Si potrebbe definire il modello come un “oggetto digitale composto”, dato che aggrega più elementi di contenuto all'interno dello stesso. Questi elementi di contenuto, possono essere di qualsiasi formato e possono venire archiviati in due modi: localmente nel repository oppure memorizzati esternamente e collegati tramite un riferimento all'oggetto digitale. Il modello di oggetti digitali creato da Fedora è piuttosto semplice e flessibile per dare la possibilità di creare diversi tipi di oggetti digitali; tuttavia, la natura generica dell'oggetto digitale in Fedora, consente di gestire in modo coerente tutti gli oggetti all'interno del repository. Il modello di oggetti digitali Fedora è definito tramite l'uso del linguaggio XML, vediamo ora quali sono i suoi componenti di base:

- **PID:** l'identificativo persistente e univoco dell'oggetto.
- **Object Properties:** si tratta dell'insieme di proprietà descrittive definite dal sistema, per gestire e monitorare l'oggetto nel repository.
- **Datastream:** è un componente dell'oggetto che rappresenta un elemento contenuto; può essere solo uno, come ce ne possono essere diversi. Ogni datastream registra attributi utili sul contenuto che rappresenta come, ad esempio, il MIME-type per la compatibilità Web o l'URI (Uniform Resource Identifier) che identifica il formato del contenuto. Un datastream viene trattato come un flusso di bit opaco, implicando che sia l'utente a determinare come interpretare il contenuto, sia esso dati o metadati ad esempio. Il suo contenuto, può essere archiviato integralmente nel repository oppure memorizzato in remoto; in questo caso Fedora utilizza un puntatore sotto forma di URL che direziona al contenuto. Ogni datastream è dotato

---

<sup>68</sup> <http://fedorarepository.org/> [Ultimo accesso 13/09/2017]

di un suo identificatore univoco nell'ambito dell'oggetto digitale. Fedora inoltre, riserva a datastream quattro diversi identificatori per il suo utilizzo:

- DC (Dublin Core), impostato di default e creato automaticamente se non fornito, viene utilizzato per contenere i metadati relativi all'oggetto;
- AUDIT, che registra una traccia di controllo relativa a tutte le modifiche apportate all'oggetto e non può essere modificata in quanto controllata dal sistema;
- RELS-EXT, utilizzato principalmente per descrivere le relazioni con gli altri oggetti digitali;
- RELS-INT, funzionale nel descrivere le relazioni interne all'oggetto.

Oltre a questi quattro identificatori, un oggetto digitale Fedora può contenere numerosi altri datastream personalizzati per rappresentare contenuti definiti dall'utente.

- **Datastream identifier:** l'identificatore di datastream deve essere univoco all'interno dell'oggetto digitale, mentre non necessariamente a livello globale.
- **State:** lo stato del datastream che può essere attivo, inattivo o cancellato.
- **Created date:** il giorno e l'ora in cui il datastream è stato creato. Viene assegnato dal repository in automatico.
- **Modified date:** il giorno e l'ora in cui il datastream è stato modificato. Viene assegnato dal repository in automatico.
- **Versionable:** un indicatore booleano (vero o falso) che stabilisce se il servizio di Repository deve o meno dare una versione al datastream. Di default quest'impostazione è attiva.
- **Label:** un'etichetta descrittiva per il datastream.
- **MIME Type:** obbligatorio, si riferisce al tipo di datastream affinché sia trasmissibile in rete.
- **Format identifier:** un identificatore di formato opzionale per i datastream simile all'emergente PRONOM<sup>69</sup>, sviluppato dagli archivi nazionali del Regno Unito.
- **Alternate identifiers:** uno o più identificatori alternativi per il datastream, i quali possono essere locali o globali come il DOI (Digital Object Identifier).
- **Checksum:** utile a controllare che il contenuto del datastream sia integro. Può essere utilizzato tramite un calcolo effettuabile mediante vari algoritmi standard come MD5 o SHA-1.
- **Bytestream Content:** il contenuto, visto come una risorsa di flusso, viene rappresentato o incapsulato dal datastream a seconda della sua tipologia (documento, immagine, video, ... )
- **Control Group:** L'approccio utilizzato dal Datastream per rappresentare o incapsulare il contenuto come uno dei quattro tipi o gruppi di controllo elencati a seguire:
  - Internal XML Content: il contenuto viene memorizzato come XML all'interno del file

---

<sup>69</sup> <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx> [Ultimo accesso 05/10/2017]

XML dell'oggetto digitale.

- Managed Content: il contenuto è memorizzato nel repository e l'oggetto digitale XML mantiene un identificatore interno che può essere utilizzato per recuperare il contenuto dall'archivio.
- Externally Referenced Content: il contenuto viene memorizzato al di fuori del repository, ma l'oggetto digitale XML mantiene un URL che punta allo stesso in una località remota, rendendolo sempre recuperabile. Questo URL viene utilizzato nel momento in cui c'è una richiesta di accesso per questo datastream, la quale viene mediata dal repository. Ciò significa che Fedora prende esternamente il contenuto e lo serve al soggetto richiedente come se fosse servito direttamente da Fedora.
- Redirect Referenced Content: anche in questo caso il contenuto viene memorizzato al di fuori del repository e c'è un URL in grado di puntarlo. Il contenuto però, non viene trasmesso attraverso il repository. Ciò è vantaggioso quando si desidera che il repository venga messa da parte quando bisogna fornire il contenuto. Un esempio potrebbe essere quando si vuole far uso di uno streaming media server il quale, avendo a disposizione un determinato video da offrire, prende il controllo per lanciare il contenuto al posto di Fedora.

Nonostante ogni oggetto digitale di Fedora si conformi al modello appena descritto, ci sono quattro differenti tipi di oggetto che possono essere depositati all'interno del repository. La distinzione tra queste quattro tipologie è fondamentale per capire come funziona il sistema di archiviazione. In Fedora ci sono oggetti che contengono contenuti digitali (Data Object), oggetti che contengono descrizioni relative al servizio (Service Definition Object), oggetti utilizzati per dislocare i servizi (Service Deployment Object) e oggetti usati per l'organizzazione di altri oggetti (Content Model Object).

Un *Data Object* è un tipo di oggetto utilizzato per rappresentare un'entità di contenuto digitale. È quello a cui normalmente pensiamo quando ci immaginiamo un repository che memorizza le collezioni digitali. I Data Objects possono rappresentare soggetti molto diversi tra loro come immagini, libri, testi elettronici, learning objects, pubblicazioni, collezioni e molte altre entità. Uno o più Datastream sono usati per rappresentare le parti del contenuto digitale.

Un *Service Definition Object* è un tipo speciale di oggetto utilizzabile per depositare il modello di un servizio. Contiene quindi un set integrato di operazioni che il Data Object supporta. In termini di programmazione orientata agli oggetti, definisce un'interfaccia che elenca le operazioni supportate, ma non definisce esattamente come viene eseguita ciascuna operazione.

Un *Service Deployment Object* è un tipo speciale di oggetto di controllo, che descrive come un



repository specifico consegnerà le operazioni di servizio descritte in un Service Definition Object per una classe di oggetti dati descritti in un Content Model Object. Non si tratta di codice eseguibile ma contiene invece informazioni che indicano al repository Fedora come e dove eseguire la funzione rappresentata.

Un *Content Model Object* è un nuovo oggetto di controllo specializzato che funziona come contenitore per il suo documento, il quale è un modello formale che caratterizza una classe di oggetti digitali. Può anche fornire un modello di relazioni permesse, escluse o richieste tra gruppi di oggetti digitali. Tutti gli oggetti digitali in Fedora appena esposti sono organizzati in classi dal Content Model Object.

Già a partire da una descrizione approfondita degli oggetti digitali in Fedora, si riescono a capire parecchie cose sul funzionamento della piattaforma Phaidra. Possiamo passare quindi a vedere, facilitati dal supporto grafico, come funzionano i processi lavorativi all'interno del sistema compresi gli attori e i servizi che entrano in gioco quando un oggetto digitale si muove.

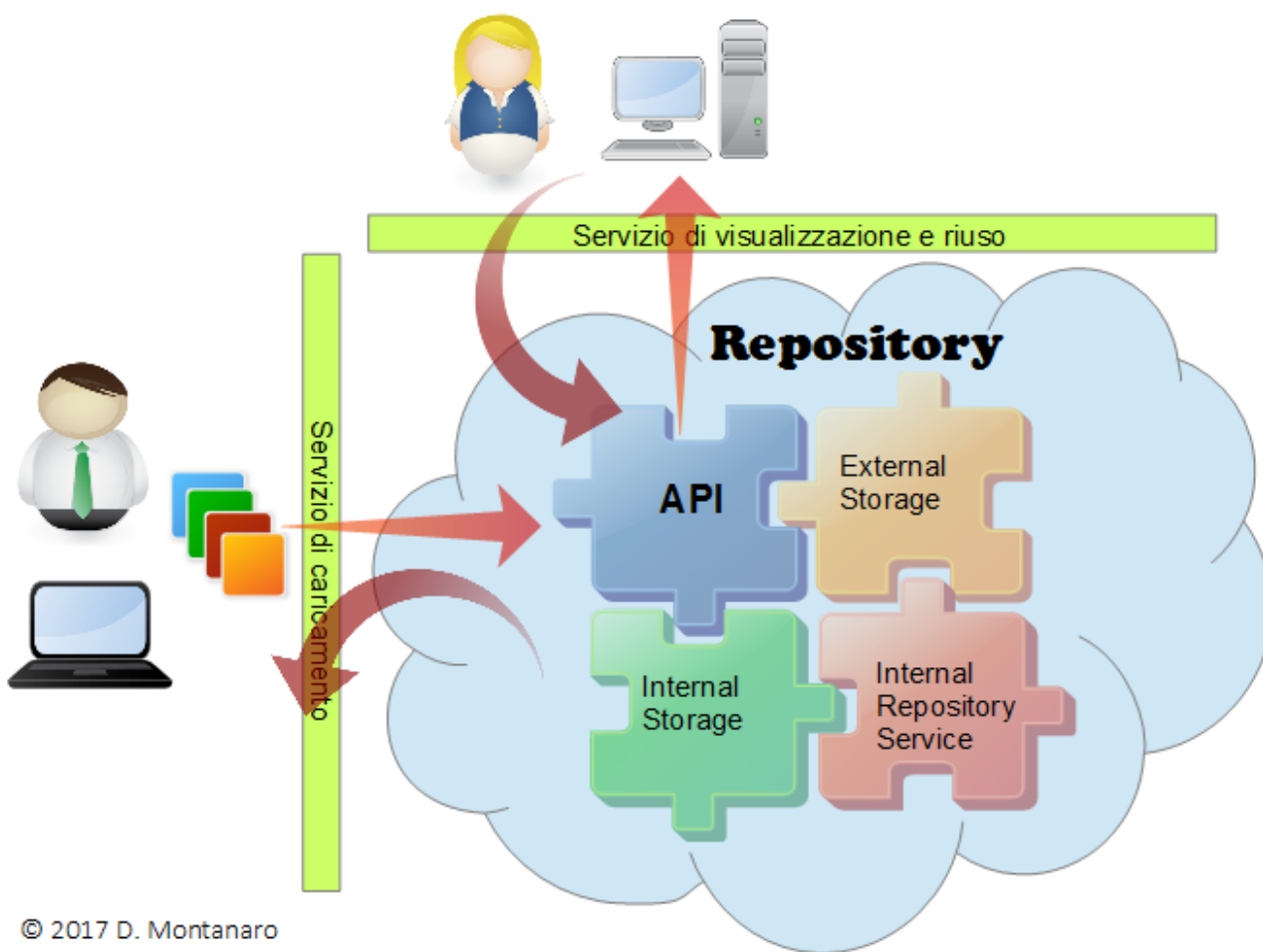


Figura 11 – La gestione dei processi

Partiamo dal presupposto che ci sono diversi tipi di file che possono venire caricati su Phaidra,

differenziati in altrettante categorie (immagini, video, libri, documenti, ...). Questi file, provengono da due diverse possibili tipologie di utenza: *umana*, ossia una persona in carne e ossa che prepara la risorsa e la processa tramite l'interfaccia oppure *tecnica*, ossia un altro sistema impostato per fare questo lavoro e che importa un oggetto digitale. In ogni caso, la prima mossa sarà quella di accettare le condizioni di utilizzo trattate nello scorso capitolo, con tutto ciò che ne comporta. Il compito di caricamento, operato da parte dell'utente umano, consisterà nell'affiancare alla risorsa i relativi metadati nel modo più accurato possibile, in taluni casi sfruttando anche dei modelli preimpostati; questi riguardano anche tutta una serie di condizioni relative ai permessi di accesso. Per quanto riguarda l'utente tecnico, i metadati possono venire importati mediante dei modelli RDF (Resource Description Framework) o XML. Al termine di quest'operazione, l'utente avrà delle risposte da parte del sistema; saprà quindi, se la risorsa è stata processata con successo o meno e riceverà ogni informazione necessaria su quanto avvenuto. Tra queste, in caso di esito positivo, il PID che consente il raggiungimento permanente del file tramite un identificatore univoco e che non è destinato a cambiare.

Ci spostiamo ora all'interno della nuvoletta che delimita il repository, dove avviene tutta la fase di gestione dei dati. Questa, si appoggia sull'uso delle *API*; in Phaidra sono realizzate in Perl e Java e forniscono le procedure in grado di far interagire il repository con il mondo esterno, garantendo il funzionamento dei servizi sia in entrata che in uscita. Nell'*Internal Storage* sono depositati in locale gli oggetti con i loro metadati, ossia quelli sopra definiti Data Object. Gli ultimi due pezzi a completare il puzzle si chiamano *Internal Repository Services* ed *External Storage*. Il primo riguarda i servizi interni, che sono rappresentati invece da gli altri tipi di oggetti: Content Model Object, Service Definition Object e Service Deployment Object. Infatti è qui che vengono gestiti tutti i servizi interni nelle loro diverse funzionalità. Si parla quindi di *conservazione*, tramite servizi quali il Backup, l'archiviazione e la conservazione a lungo termine. Oppure di tutto ciò che riguarda il *processo* che segue l'oggetto dal momento in cui si avvicina alla piattaforma, come la codifica, la conversione, i flussi di dati. Infine di servizi di *gestione interna*, come la definizione dei permessi o l'indicizzazione. Per quanto riguarda l'External Storage, si occupa prevalentemente di far funzionare i collegamenti tra oggetti di sistemi diversi, tramite l'uso dei metadati. Ricordiamo il caso degli oggetti che vengono visualizzati dentro Phaidra oppure di quelli che vengono visualizzati esternamente tramite un servizio di streaming. Anche in questo caso però c'è un dato che permette di puntare alla risorsa desiderata. Esiste ovviamente l'opzione inversa dove il contenuto interno di Phaidra viene richiesto da un'altra applicazione affine e fornito. Inoltre c'è una Cache dedicata che permette il caricamento temporaneo della risorsa esterna in modo da velocizzare le operazioni desiderate.

Per concludere abbiamo il *servizio di visualizzazione e riuso*. Può venire anche in questo caso utilizzato sia dall'utente umano tramite interfaccia, che dall'utente tecnico tramite codice. È possibile

trovare, grazie al server di ricerca, tramite le relazioni o addirittura casualmente, gli oggetti digitali contenuti nel repository (dati e metadati) e se permesso visualizzarli e/o scaricarli. Non è invece possibile modificarli bensì riutilizzarli, quindi creare eventualmente una nuova versione con i suoi metadati e caricarla.



## Capitolo 4 – Il digital workflow: il documento e le fasi di lavoro

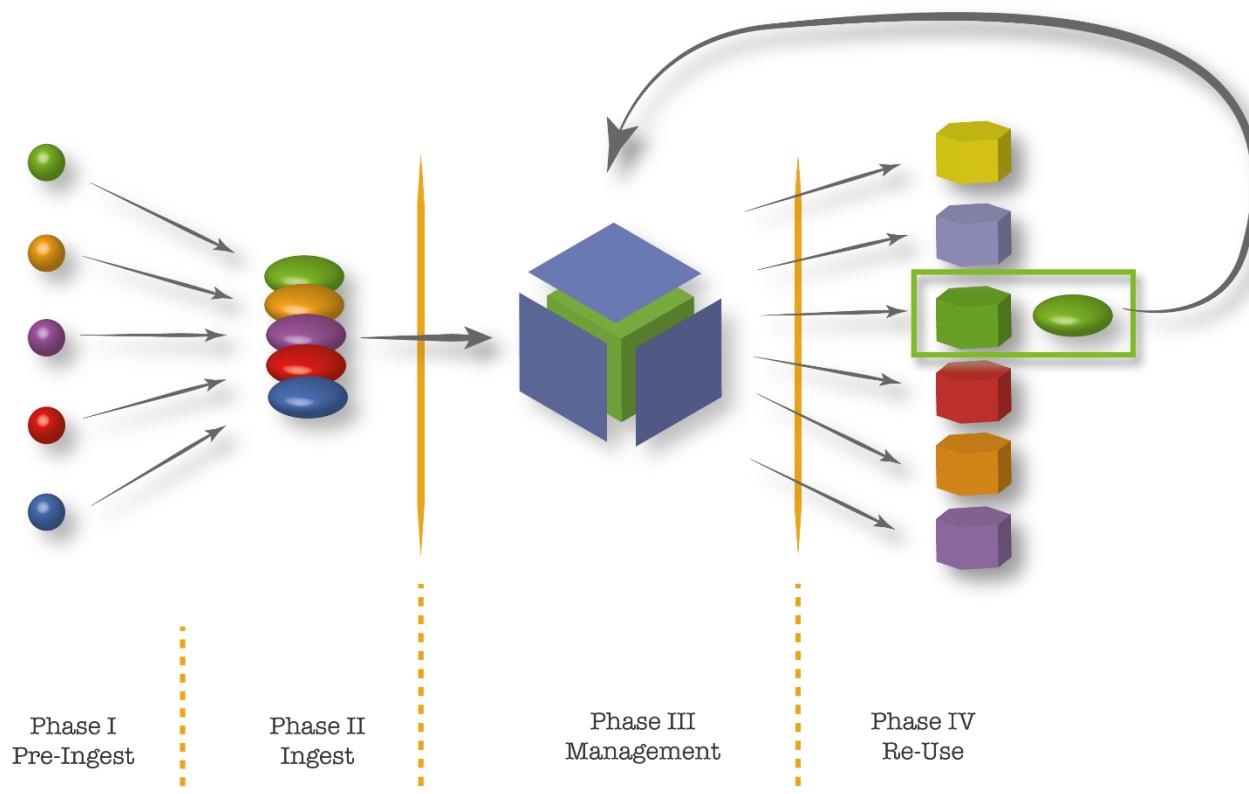


Figura 12 – Workflow ciclico

CC BY 4.0 International Raman Ganguly. <http://phaidra.univie.ac.at/o:424614>

Dopo questo primo sguardo ad ampia visuale sul mondo di Phaidra, è giunto il momento di analizzare le singole fasi che il documento affronta nel corso della sua esistenza. Onde evitare studi troppo generalisti e superficiali, la decisione è stata quella di concentrarmi per questo elaborato solo su un gruppo di risorse: le immagini. Da questo momento, pur non trascurando completamente l'esistenza di altre tipologie di dati, sarà il cardine sul quale verterà la tesi.

Il documento, nel momento in cui si decide che verrà processato, deve seguire tutta una determinata procedura a carico dei soggetti, la quale segue delle politiche ben precise<sup>70</sup>. Quindi, a partire dal documento grezzo ci saranno le varie fasi di lavoro, dalla sua preparazione fino al management e gli eventuali riutilizzi. Sempre che il processo vada a buon fine ovviamente. È altamente consigliabile, ai fini di un utilizzo efficiente del software, di considerare i singoli passaggi di archiviazione a lungo termine e di conservarli per iscritto. Nel corso del capitolo verranno spiegate dettagliatamente le quattro fasi portanti (preparazione; caricamento; archiviazione e gestione; estrazione e visualizzazione), con un occhio di riguardo nei confronti delle responsabilità che queste implicano e di come esse vadano affrontate. Il tutto a partire da alcune legittime domande che il soggetto chiave

<sup>70</sup> Phaidra, The Ten Commandments for Policy, <http://phaidraservice.univie.ac.at/en/phaidra/policy/> [Ultimo accesso 22/09/2017]

deve porsi per affrontare al meglio il suo compito. Infine, tramite l'utilizzo di Phaidra Sandbox, verrà offerto un esempio pratico di come vengono processati gli oggetti a partire dal loro caricamento. Saranno mostrati degli screenshots per far sì che il risultato sia più chiaro possibile.

#### 4.1 La preparazione (pre-ingest):

Nel momento in cui un soggetto produttore ha delle risorse che ritiene opportuno caricare sulla piattaforma, deve preoccuparsi di prepararle nel modo più accurato possibile. Dal momento che si tratta di un'archiviazione a medio-lungo termine, sarebbe un peccato sprecare del tempo per caricare del materiale che non si rivela ottimale. Per evitare questa situazione, sono stati scelti accuratamente dei formati consigliati; questi mirano a garantire qualità, un lungo periodo di conservazione e un alto livello di compatibilità.

Questa sezione si pone l'obiettivo di chiarire, attraverso una serie di strumenti, cosa si intende e in cosa consiste il pre-ingest nel processo che porta al caricamento dei dati in Phaidra<sup>71</sup>. Per cominciare, è lecito farsi qualche domanda per comprendere a pieno le sue finalità: ad esempio, quali attività devono essere eseguite prima del caricamento effettivo nel repository? Chi ne è responsabile? Di quali informazioni ha bisogno? È fondamentale come prima cosa definire che, in questa fase, la responsabilità ricade sul *produttore di dati* il quale deve decidere quali dati andranno archiviati sulla piattaforma. Deve inoltre valutare e verificare la loro qualità, oltre a scegliere consapevolmente che tipo di licenza andrà a garantire il loro riutilizzo. Affinché il produttore di dati possa adempiere ai propri compiti efficacemente, deve sapere utilizzare e conoscere uno strumento come il **Data Management Plan** (Cap. 3.1). Questo lo guida passo dopo passo dalla concezione del progetto fino al suo termine. Attraverso il DMP, è infatti possibile seguire tutte le fasi con metodologia e pianificazione grazie all'ausilio di questo supporto che aiuta tra il resto nella valutazione dei costi, la quale deve tenere conto della durata dell'archiviazione, della complessità dei formati di dati archiviati e della dimensione dei dati.

Tra le attività che caratterizzano questa fase, si inizia con una procedura più teorica che comprende l'ideazione del progetto, la sua applicazione, la sua descrizione e la sua approvazione. Successivamente il lavoro diventa un po' più pratico, dato che vengono prodotti, ordinati e raggruppati i primi dati. Questi poi vanno definiti tramite la scelta del formato ideale, possibilmente tra quelli consigliati, l'allocazione dei metadati (descrittivi, tecnici, contestuali, di provenienza, ...), l'eventuale uso di classificazioni e traduzioni, la chiarificazione delle questioni legali e il loro monitoraggio. Il tutto considerando che, questa prima fase di lavoro, non viene direttamente supportata da Phaidra. Per far sì che queste operazioni siano svolte opportunamente, è necessario che il produttore di dati abbia determinate competenze. Tra queste c'è bisogno di esperienza nei seguenti campi:

- i. Research Data Management (RDM);

---

<sup>71</sup> BUDRONI P., PREZA J. L., PAUSZ R., SÁNCHEZ B., BLUMESBERGER S., *Data Management Plan - Eine Anleitung zur Erstellung von Data Management Plänen. Projekt e-Infrastructures Austria (Deutsch): Eine Anleitung zur dauerhaften Sicherung von digitalen Beständen*, 2015. Disponibile a: <<http://phaidra.univie.ac.at/o:367863>> [Ultimo accesso 12/09/2017]

- ii. utilizzo di archivi e conoscenza delle procedure di archiviazione;
- iii. attività di ricerca e analisi di dati;
- iv. gestione dei diritti e delle questioni legali relative ai dati e alla loro condivisione.

Vediamo ora come, a partire dalla definizione del progetto, si svolgono nel dettaglio le operazioni elencate sopra. Bisogna considerare che non tutte le questioni che verranno tirate in ballo sono necessarie in ogni contesto; sono strettamente legate al tipo di lavoro svolto, il quale implica la possibilità di non preoccuparsi di determinati fattori.

Come primo passo è importante capire di che *tipologia di dati* si occuperà il progetto, quanti potrebbero essere e quanto spazio potrebbero occupare. Riguardo allo spazio va fatta una stima, cercando di capire inoltre quanto potrebbero essere pesanti i singoli files, specialmente i più grossi. Ogni tipologia dispone di diversi formati, tramite i quali il produttore può creare i propri oggetti. Phaidra suggerisce i seguenti<sup>72</sup>; per le ragioni che vedremo in seguito:

## supported Formats

Formate	Recommended	Possible	not recommended
Image	.tiff	.jpg	.psd, .bmp, .gif, .png, ...
Video	.avi, mpeg2	.mp4	.wmv, .flv, .mov, ...
Audio	.wav	.mp3	.wma, .aac, .ogg, ...
Documents	.pdf	.tex, .html	.doc, .ppt, .indd, .qxd, ...

Figura 13 – Formati supportati

I fattori chiave che influenzano la scelta del formato sono la qualità che esso offre, la sua compatibilità e le garanzie che offre per una conservazione a lungo termine. Se necessario, nonostante un lavoro sia stato portato avanti tramite un determinato formato, è possibile convertirlo successivamente in un nuovo formato più adeguato allo scopo di migliorarne l'archiviazione. In ogni caso, Phaidra offre l'opportunità di caricare dati in qualsiasi formato. Osservando l'immagine tratta dal sito ufficiale, è evidente come ci siano parecchi formati non raccomandati, i quali hanno una caratteristica comune: sono formati proprietari. Ciò implica che non siano liberamente fruibili le specifiche tecniche complete, negando quindi l'utilizzo dei files a chi non è in possesso del software autorizzato, che detiene una sorta di monopolio per l'accesso. Questo li rende assolutamente sconsigliabili, in quanto vanno contro ai principi del libero accesso e della conservazione a lungo termine. Può capitare, ad esempio, che un utente possa perdere tutte le informazioni contenute in un oggetto con formato

<sup>72</sup> Phaidra – Supported Formats, <https://backend.univie.ac.at/index.php?id=23004&L=2&id=23004> [Ultimo accesso 26/09/2017]



proprietario se l'unico software proprietario in grado di operare correttamente con tali dati non è più reso disponibile. Nel caso invece di quanto elencato nelle colonne “Recommended” e “Possible”, parliamo di formati aperti le cui specifiche tecniche sono di pubblico dominio; ovviando al problema appena descritto. La differenza che qui rende un formato migliore di un altro, sta nella qualità che essi offrono agli utenti. Ovviamente una migliore qualità influisce anche sullo spazio fisico occupato dal file, che aumenta; spesso però è la scelta migliore. Va comunque fatto un ragionamento completo che parte col chiedersi come i dati vengono creati e di conseguenza salvati e raccolti. Includendo tutto il processo e considerando eventuali usi di software o hardware speciali. Nel momento in cui un altro utente voglia riutilizzare un oggetto, la scelta della tecnologia usata, del formato e dei metadati sarà determinante.

È possibile che agli oggetti sia necessario allegare una *documentazione*, che descriva il lavoro svolto. Anche questa merita attenzione sulla scelta del formato, che va deciso a partire dal modo in cui i dati vengono documentati: ad esempio appunti di laboratorio, note a margine, file audio, ...

Affinché tutto questo lavoro abbia un senso, è necessario che vengano impostati adeguati *metadati*. Come abbiamo visto nel capitolo 3.3, fanno parte dell'oggetto stesso e sono indispensabili ai fini della ricerca e del riuso. Se sappiamo che esiste un file, ma non sappiamo come trovarlo, diventa perfettamente inutile. Qui entrano in gioco i metadati che si possono definire come informazioni strutturate che descrivono, spiegano e individuano o altrimenti facilitano il recupero, l'utilizzo o la gestione di una risorsa. Per gestirli al meglio va compreso come vadano strutturati caso per caso. È necessario in questa fase avere tutte le informazioni necessarie per archiviare gli oggetti, come il titolo, la descrizione, le persone legate all'oggetto e il loro ruolo, la licenza. Sapere di che diritti si è in possesso, ad esempio nel caso in cui si vogliono usare metadati creati da altri ricercatori. Nel caso di un gruppo di ricerca può essere inoltre interessante capire che soggetti sono responsabili per il loro uso. In Phaidra vengono utilizzati due standard di metadati: LOM e DC. Il LOM (Learning Object Metadata) è lo standard più completo ed è quello che viene usato principalmente. È stato creato appositamente per i Learning Object e fissa le proprietà minime necessarie per una loro corretta gestione, allocazione e valutazione. È composto da 9 aree descrittive (General, Lifecycle, Meta-Metadata, Technical, Educational, Rights, Relation, Annotation, Classification), le quali contengono gruppi di attributi strutturati ad albero; in totale conta 70 elementi descrittivi che vedremo nel dettaglio seguendo la fase di caricamento. Il DC (Dublin Core) è invece uno standard molto semplice e generale, che non permette una descrizione adeguata e completa delle risorse caricate. Tuttavia, proprio queste sue caratteristiche, gli permettono di essere molto usato e di diventare utile quando serve garantire l'interoperabilità. Questo vale nel caso delle esportazioni/importazioni di oggetti tra Phaidra e sistemi come OpenAire e Europeana. È in corso attualmente un ragionamento che punta a

capire quale potrebbe essere lo standard più adeguato per il futuro. È interessante il caso di MODS<sup>73</sup> (Metadata Object Description Schema), sviluppato dalla Library of Congress degli Stati Uniti.

Altro aspetto importante che va considerato in questa fase è quello legato all'*etica* e alla *legalità*. È necessario infatti approfondire gli aspetti etici che stanno dietro al progetto sviluppato e capire che diritti di proprietà intellettuale rientrano in gioco quando si ha a che fare con le risorse utilizzate. Il discorso etico può essere affrontato sia a partire dai contenuti, che a partire dai dati sensibili i quali non possono venire divulgati automaticamente. Potrebbe essere il caso di renderli anonimi, anche se questo anonimato va documentato affinché sia credibile. Vanno poi considerate tutte una serie di limitazioni: potrebbe essere necessario ad esempio limitare l'accesso a determinati soggetti e/o gruppi. Oppure bisogna assicurarsi di avere un permesso scritto nel momento in cui si utilizzano materiali creati da terzi nel proprio oggetto come una colonna sonora. Fondamentale è che siano chiari tutti gli aspetti relativi al copyright e ai diritti individuali. Il primo passo, quando si carica un oggetto in Phaidra, è quello concernente l'accettazione dei termini d'uso del servizio con tutto ciò che ne consegue. Il responsabile quindi deve garantire di avere tutti i permessi necessari per caricare i dati, in modo che questi, assieme ai metadati, possano venire visualizzati su internet. Si può considerare in caso di necessità periodi in embargo, durante i quali i dati sono caricati ma non visibili per un tempo prestabilito. Vanno infine decise le licenze da impostare, tenendo presente che un copyright completo (tutti i diritti riservati) sarà poi modificabile con una licenza creativa (alcuni diritti riservati), mentre non è possibile il contrario.

Nel corso della preparazione non vanno dimenticate le problematiche inerenti il *deposito* e *backup* dei dati, compresa la gestione dell'accesso e la sicurezza. Durante la ricerca quindi, oltre ad avere il dovere di controllare di avere spazio sufficiente a disposizione, il produttore di dati deve considerare la possibile necessità di servizi aggiuntivi come consigli tecnici, l'implementazione di un sito web del progetto, l'utilizzo di un Content Management System, ecc. Tematiche importanti inoltre sono quelle legate all'uso del backup: come viene fatto, chi ne è responsabile, gli eventuali piani di emergenza nel caso di un incidente. Riguardo all'accesso e sicurezza va valutato il grado di sensibilità dei dati e se c'è un reale rischio che qualcuno possa accedervi illegalmente e manipolarli. D'altro canto però va anche assicurato che i collaboratori possano al contrario accedervi in sicurezza. Per questo vanno chiariti i ruoli di responsabilità, come quello relativo alla sicurezza, all'interno del progetto ed eventualmente fornire una password adeguata. Tra le valutazioni da fare c'è anche quella di distinguere i dati che andranno conservati a lungo termine e quelli che non ne hanno bisogno. Al termine del progetto poi, il processo di pianificazione deve far sì che i dati archiviati a lungo abbiano delle garanzie.

---

<sup>73</sup> <http://www.loc.gov/standards/mods/> [Ultimo accesso 27/09/2017]

Una volta che i dati saranno effettivamente pronti, c'è uno stadio di *selezione e conservazione* dove va deciso quali oggetti andranno conservati e/o condivisi. È possibile che non tutto vada conservato a lungo termine e, in questo caso, Phaidra non è il repository adatto. La cancellazione infatti è una procedura esclusivamente d'emergenza. Una volta quindi selezionati i file che andranno caricati, va considerato l'utilizzo dell'identificatore permanente. Phaidra ne assegna uno di default che è sempre disponibile globalmente e può essere citato o condiviso. Il suo link fornisce l'accesso all'oggetto (se con accesso limitato solo ad anteprima e metadati), che può venire visualizzato nel browser e scaricato se necessario. Altre tipologie importanti di identificatore permanente conosciute possono essere DOI, Handle o URN. Se per gli scopi del progetto in questione, è necessario utilizzare più repositories, la cosa migliore è avere ben chiara la suddivisione dei dati e i costi che l'archiviazione comporta. Non va mai dimenticato che i progetti devono tenere presente il futuro dei dati anche dopo il loro termine. La fase di pre-ingest comporta anche la pianificazione della successiva *condivisione* dei dati. Come questi quindi potranno venire trovati; decidere se esiste o meno la necessità di usare eventuali diritti d'accesso differenziati, o decidere se la pubblicazione debba avere eventualmente libero accesso e se mai con quale licenza o se usare periodi di embargo. Inoltre va deciso se i dati avranno la possibilità di essere citati. Ciò comporta anche la discussione sul possibile uso degli oggetti a fine progetto. In certi casi, ad esempio, i dati potranno essere spostati da un repository ad un'altra (come nel caso di Phaidra Temp) o potranno essere visualizzati attraverso piattaforme esterne.

Per concludere è necessario assegnare la *responsabilità* riguardo al Data Management Plan<sup>74</sup>, il quale deve essere implementato prima e revisionato e aggiornato poi. Oltre a ciò va considerata qualsiasi risorsa aggiuntiva (personale, software, consulenze legali, ...) che può risultare necessaria per preparare efficacemente i dati per l'archiviazione a lungo termine.

La fase di preparazione non è considerata all'interno del modello OAIS, il quale presuppone che il processo che porta all'archiviazione dei documenti digitali inizi con il caricamento. Ciò risulta sconveniente nel momento in cui si pensa ad un livello globale, visto che questo modello sta alla base di parecchi archivi digitali. La sua mancanza è problematica, in quanto viene meno tutta una fase di valutazione delle risorse che permette di garantire degli standard elevati. Sia nel caso si tratti di risorse create da utenti, che nel caso di materiali ricevuti da donatori che hanno bisogno di essere revisionati. Sicuramente è un processo che ha dei costi, i quali poi però danno il loro frutto nel lungo periodo. Ci sono attualmente delle discussioni in corso su un suo possibile inserimento nel modello OAIS, affinché questo porti ad aumentare la visibilità del problema.

---

<sup>74</sup> BUDRONI P., BLUMESBERGER S., PAUSZ R., PREZA J. L., SÁNCHEZ SOLÍS B., *Data Management Plan for projects at the University of Vienna. Recommended Repository: Phaidra (V2.0/customised for UNIVIE) (English): Version 2.0 / customised for UNIVIE*, 2015.

#### 4.2 Il caricamento (ingest):

È in questa fase che l'oggetto entra effettivamente a far parte della piattaforma. Infatti è qui che al documento viene assegnato un identificatore, necessariamente univoco, da parte del sistema. Al produttore invece, spetta il delicato compito di descrivere nella maniera più precisa possibile il file, attraverso i metadati. Questi ultimi seguono i principi FAIR di cui parleremo più avanti e riguardano sia il contenuto (lingua, autore, luoghi di riferimento, licenze, etc.), sia le caratteristiche tecniche. È molto importante che siano fatti bene per garantire la reperibilità e l'interconnessione dei dati stessi.

Dopo aver preparato tutti i file per essere caricati e continuando a seguire quanto stabilito nella redazione del Data Management Plan, si può passare alla seconda fase. Ed è in questa fase di caricamento che il gestore del repository si prende carico della responsabilità. Affinché esso sia in grado di archiviare i dati è necessario, prima di tutto, un processo di conversione degli stessi<sup>75</sup>. Le conversioni sono fondamentali per il trasferimento, al fine di garantire l'archiviazione permanente. Queste possono essere di natura puramente *tecnica*, come la conversione dei formati. Da un punto di vista *non tecnico* invece, si parla di diritti di archiviazione e riutilizzo nonché di metadati. Questo processo deve garantire che la qualità della risorsa non venga intaccata.

Partendo dalla prospettiva più tecnica bisogna tornare ad affrontare il tema dei formati supportati dal software iniziato nello scorso sotto-capitolo. Va considerato che fondamentalmente tutti i formati possono essere caricati, ma tra questi vengono raccomandati quelli più adatti per l'archiviazione a lungo termine. Questa scelta è basilare ai fini di una prospettiva a lungo termine, infatti i formati utilizzati hanno un impatto diretto sulla capacità di aprire i file e accedervi. I formati multimediali descrivono come vengono memorizzati i dati e come questi devono essere interpretati. Senza queste descrizioni, i dati sarebbero una sequenza incomprensibile di 0 e 1. Oltre al formato multimediale, esiste anche il tipo di supporto: testo, immagine, audio, video, ecc. Esistono tipologie di supporti che possono avere più formati; ad esempio i testi possono essere memorizzati in .pdf, in .doc o in .odt per dirne alcuni. Tuttavia, esistono anche formati multimediali che contengono tipi di supporto diversi. Questi cosiddetti formati "contenitori" possono in certi casi contenere contemporaneamente testo, audio e video. Col procedere del tempo la quantità dei dati continua ad aumentare e questo può implicare due diversi aspetti: da un lato migliora la qualità di registrazione, che porta all'aumento della densità dei dati. Dall'altro ci sono possibilità di produzione degli stessi più semplici, che consentono l'accesso ad un maggior numero di persone. La quantità di spazio su disco consumato in un sistema di archiviazione, ha un peso economico che viene spesso utilizzato come misura del costo totale in quanto è facile da quantificare. Per consumare meno memoria, è possibile utilizzare formati

---

<sup>75</sup> BUDRONI P., PREZA J. L., PAUSZ R., SÁNCHEZ B., BLUMESBERGER S., *Data Management Plan - Eine Anleitung zur Erstellung von Data Management Plänen. Projekt e-Infrastructures Austria (Deutsch): Eine Anleitung zur dauerhaften Sicherung von digitalen Beständen*, 2015.

multimediali che comprimono i dati. Nei metodi di compressione, i dati vengono ridotti escludendo alcune sequenze di bit non necessarie alla loro rappresentazione. Ciò avviene attraverso appositi algoritmi. A seconda dei casi poi queste informazioni contenute nei bit potrebbero essere ricostruite con metodi matematici o meno, comportando una categorizzazione in due tipi di procedure: “lossy”, ossia con perdita e “lossless”, senza perdita.

Per quanto riguarda invece il punto di vista non tecnico, va discusso in precedenza con il team come deve essere fatto il caricamento degli oggetti a livello di metadati e licenze. È necessario quindi avere a disposizione uno scenario complessivo del processo di trasferimento. Viene quindi lecito domandarsi se i metadati vengano raccolti automaticamente dopo il caricamento o se questi vengano inseriti manualmente. Phaidra offre entrambe queste possibilità, infatti i metadati possono venire eseguiti tramite l'API come anche mediante l'interfaccia utente.

Vediamo ora cosa succede effettivamente quando l'ingest viene utilizzato per trasferire i dati all'interno della gestione del repository<sup>76</sup>. Potenzialmente qualsiasi utente può caricare i media direttamente nel software, agendo tramite l'interfaccia. Questa opzione lo porta ad utilizzare esclusivamente la modalità di caricamento singolo, significando che entra nel sistema un oggetto digitale alla volta. A seguire verrà inserita la descrizione dello stesso tramite le 9 schermate dei metadati (una per area descrittiva). Per poter sfruttare la modalità di caricamento multiplo, come nel caso dei libri o delle collezioni, è necessario usare delle applicazioni “docked”: queste, come ad esempio Phaidra Importer, permettono di caricare attraverso un altro software connesso con la piattaforma le risorse. Anche i metadati vengono inseriti esternamente. Una terza opzione consiste nell'importazione di oggetti direttamente da altri sistemi.

Una volta caricati, gli oggetti entrano ufficialmente a far parte dello spazio legale (Schengen for data) e subiscono l'assegnazione di un identificatore permanente (PID) che li renderà sempre trovabili. Viene logico domandarsi, per concludere, se la piattaforma offre la funzione di Checksum, per controllare e monitorare l'integrità del file. In Phaidra esiste, ma la sequenza di bit non viene resa pubblica in quanto viene trasportata in un inventario separato.

A seguire vengono visualizzate, con supporto di Phaidra – Sandbox, delle catture derivate dal processo di caricamento di un'immagine. In questo caso è stato utilizzato l'editor di Metadati Light, che permette l'inserimento di meno informazioni però senza tralasciare quelle fondamentali. Per

---

<sup>76</sup> BERGER C. et al., *Digitale Archivierung und Bereitstellung von AV-Medien – Erfahrungen und Praxisbeispiele aus dem tertiären Bildungssektor*, BoD – Books on Demand, 2017.

cominciare ci sarà la schermata di caricamento del contenuto, la quale può avvenire dal proprio computer o tramite un URL. Successivamente verranno mostrati area per area i metadati da inserire in questa tipologia di editor. Una volta terminata questa procedura, il caricamento può dirsi concluso.

The screenshot shows the Phaidra user interface. At the top left is the University of Vienna logo. The user is logged in as 'Davide Montanaro' in the 'IT' language. A navigation menu includes 'Search', 'New object', 'My objects', 'Bookmarks', 'My groups', 'Templates', and 'Logout'. The main heading is 'Crea un oggetto IMMAGINE'. Under 'Scegli un file...', a file named 'DSCF3825.JPG' is selected, and an 'Inizia il caricamento' button is visible. Below this, there is a section for 'o digita un URL...' with a text input field and a login form. The login form shows 'Nome utente: davidem82' and a masked password. A message states: 'Nome utente e password sono necessari se l'URL richiede di autenticarsi.' Below the login form is another 'Inizia il caricamento' button. A progress bar at the bottom of the upload section shows '100%' completion. Below the progress bar, the status is 'Status: Transfer complete (2.11M)', the time taken is 'Time: 0:02 (~ 0:00 to go)', and the rate is 'Rate: 1.39M (avg 1.18M)'.

Figura 15 – Caricamento su Phaidra

The screenshot shows the 'Dati generali' (General Data) section of the Phaidra metadata editor. It contains the following fields:

- Titolo\*:** 'Vista di Castel Telvana' in the text box, and 'nella lingua: Italiano' in a dropdown menu with a '+' icon.
- Lingua\*:** 'Italiano' in a dropdown menu with a '+' icon.
- Descrizione\*:** 'Foto scattata dal monte Armentera, nel comune di Borgo Valsugana (TN). E' possibile vedere dall'alto il castello medievale.' in a text box, and 'nella lingua: Italiano' in a dropdown menu with a '+' icon.

Below these fields is a link 'Aggiungi uno o più link'. At the bottom, there is a section for 'Identificatori' with a '+' icon, containing a 'Standard:' dropdown set to 'DOI' and an empty 'Identificatore:' text box.

Figura 16 – Inserimento tramite Editor Metadati Light (Fase 1)

**Ciclo di vita**

Stato\*:

Sottoposto a revisione paritaria (Peer Reviewed):

---

**Contributo**

Ruolo\*:

---

**Ruolo / Dati personali\***

Nome:  Cognome:  Titolo:  Titolo:

Figura 17 – Inserimento tramite Editor Metadati Light (Fase 2)

**Diritti & Licenze**

Costi\*:

Copyright e altre limitazioni\*:

Versione breve della licenza

Attribution 4.0 International (CC BY 4.0)

**Tu sei libero di:**

Condividere — riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato.

Modificare — remixare, trasformare il materiale e basarti su di esso per le tue opere per qualsiasi fine, anche commerciale.

Il licenziante non può revocare questi diritti fintanto che tu rispetti i termini della licenza.

Licenza scelta\*: **Alle seguenti condizioni:**

Attribuzione — Devi riconoscere una menzione di paternità adeguata, fornire un link alla licenza e indicare se sono state effettuate delle modifiche. Puoi fare ciò in qualsiasi maniera ragionevole possibile, ma non con modalità tali da suggerire che il licenziante avalli te o il tuo utilizzo del materiale.

Divieto di restrizioni aggiuntive — Non puoi applicare termini legali o misure tecnologiche che impongano ad altri soggetti dei vincoli giuridici su quanto la licenza consente loro di fare.

Questo non è il testo della licenza, ma un riassunto non vincolante.  
Qui si trova il testo vincolante della licenza scelta:  
[Versione lunga](#)

**ESCLUSIONE DI RESPONSABILITÀ (DISCLAIMER)**

Figura 18 – Inserimento tramite Editor Metadati Light (Fase 3)

## Classificazione

### Classificazioni (Classi, Sottoclassi) +

[Aiuto per la ricerca](#)

Tipo di classificazione:

Soggetto\*:

Figura 19 – Inserimento tramite Editor Metadati Light (Fase 4)



### 4.3 L'archiviazione e la gestione (management):

Una volta che i documenti con relativi metadati sono stati caricati, questi sono parte dell'archivio a tutti gli effetti. In questo momento entra in gioco la gestione vera e propria delle risorse: il produttore rimane sempre proprietario del file con esclusività e all'occorrenza può aggiornarlo. Da notare che gli aggiornamenti non cancellano i precedenti metadati, ma ne creano una nuova versione. Qualsiasi utente, tramite l'interfaccia web, può accedere e visualizzare il documento integralmente o nei limiti imposti dalla licenza. A questo scopo non è necessario essere membri dell'ateneo.

L'obiettivo chiave dell'archiviazione e gestione dei dati, è quello di assicurare la loro sicurezza e la possibilità di accedervi. Il tutto nella qualità richiesta per garantire la riusabilità ad un certo livello. Questa è la fase centrale dal punto di vista della gestione dei dati, il che significa che hanno luogo i processi legati all'archiviazione. Viene definito ad esempio come e quando inizia il riutilizzo o l'utilizzo multiplo, oppure come sono collegati tra loro i dati digitali. Sempre in questa fase avviene anche la formazione delle collezioni<sup>77</sup>. I dati sono qui forniti ad un determinato gruppo di utenti, nella qualità in cui sono stati caricati dal produttore. La durata è indeterminata, in quanto stiamo parlando di archiviazione a lungo termine. Nel caso di oggetti impostati con libero accesso, il gruppo di utenti include l'intero pubblico; altrimenti può essere limitato se precedentemente richiesto.

Al fine di garantire l'utilizzo dei principi FAIR (make data findable, accessible, interoperable and reusable) in un contesto di Open Data, esistono tutte una serie di funzioni<sup>78</sup> sviluppate per rendere l'organizzazione dei dati efficiente. Partiamo da quelle che permettono di rendere i files trovabili. Vengono offerte delle *funzioni di ricerca*, la quale può essere semplice, avanzata o a scorrimento. Nel primo caso va attraverso tutti i campi dei metadati, ma nel secondo questi campi possono essere limitati attraverso dei filtri, per restringere la ricerca nel caso si abbiano le idee più chiare. La funzione di scorrimento invece permette di vedere tutti gli oggetti, "scorrendoli" in base al loro tipo. Si possono in questo caso facilmente ordinare in maniera crescente o decrescente per data o seguendo l'alfabeto. È presente la possibilità di usufruire della *categorizzazione*, nonché dell'*assegnazione di parole chiave*. Si possono creare diverse *versioni* dell'oggetto, le quali poi non possono essere successivamente modificate. Nel caso in cui alcuni dati o alcuni metadati debbano subire un aggiornamento è necessario creare una nuova versione, eccetto nel caso della versione DC dei metadati che viene sovrascritta; dal momento che siamo all'interno di un sistema di archiviazione a lungo termine, le risorse non possono venire modificate.

In questa fase c'è anche la *gestione dei diritti dell'utente*: esso può accedere in qualsiasi momento ai

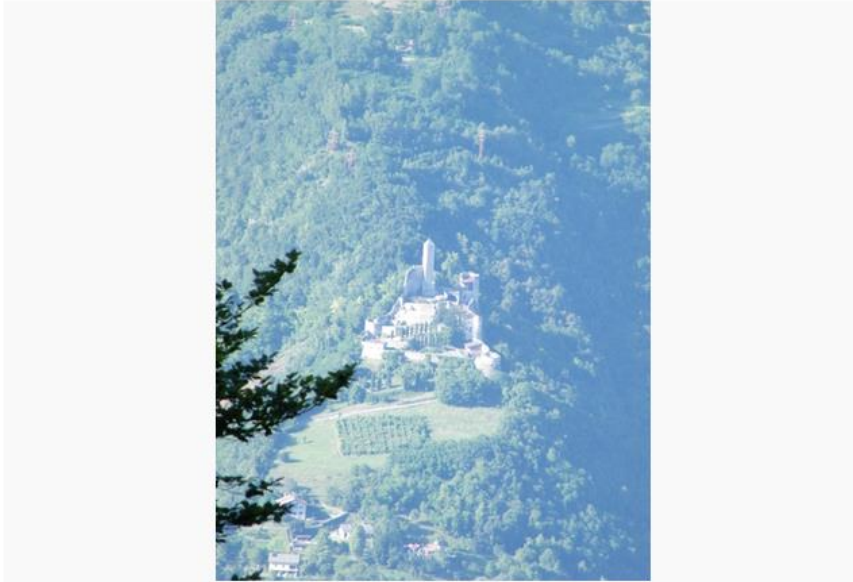
---

<sup>77</sup> BUDRONI P., PREZA J. L., PAUSZ R., SÁNCHEZ B., BLUMESBERGER S., *Data Management Plan - Eine Anleitung zur Erstellung von Data Management Plänen. Projekt e-Infrastructures Austria (Deutsch): Eine Anleitung zur dauerhaften Sicherung von digitalen Beständen*, 2015.

<sup>78</sup> BERGER C. et al., *Digitale Archivierung und Bereitstellung von AV-Medien – Erfahrungen und Praxisbeispiele aus dem tertiären Bildungssektor*, 2017, pp 108-112.

propri oggetti, creare blocchi e aggiornare i metadati. Per quanto riguarda le *licenze*, queste sono obbligatorie in Phaidra ed è disponibile una certa scelta che spazia a seconda del tipo di diritto che il produttore si vuole riservare. Vengono sempre in ogni caso, come già spiegato, consigliate le Creative Commons. Un'altra funzione riguarda i *rapporti con gli altri oggetti*, in quanto Phaidra supporta la creazione di relazioni: gli oggetti possono infatti essere raggruppati in collezioni. Inoltre, è possibile stabilire una relazione diretta per le immagini (fronte e retro), mentre l'oggetto "Libro" è una raccolta speciale di pagine singole. La *memorizzazione di annotazioni* non è attualmente disponibile tramite interfaccia, però è possibile attraverso alcune Docked Applications. Il software è supportato dallo *standard di metadati UW*, basati sullo schema LOM (Learning Object Metadata); non c'è un modello predefinito, ma è stato scelto questo dato che è valido ed è stato creato appositamente per la ricerca e l'insegnamento. I metadati sono inoltre esportabili in DC (Dublin Core), che è ottimo per quanto riguarda l'interoperabilità, ma sfortunatamente ha a disposizione solo pochi campi molto generici. Phaidra supporta anche METS (Metadata Encoding and Transmission Standard), sviluppato da Digital Library Federation, che permette di codificare i metadati in XML. METS e LOM sono più completi e lasciano la possibilità di esportare in DC. Per concludere questa fase va aggiunto che i dati rientrano in una procedura di Backup e che viene impostato un URL persistente.

In questa sezione gli screenshots mostreranno come l'immagine viene visualizzata all'interno di Phaidra – Sandbox, con le relative opzioni di gestione. Si può notare come nella fascia destra è possibile accedere a tutta una serie di opzioni tipiche di questa fase: come creare relazioni o utilizzare quelle esistenti, visualizzare i metadati e crearne una nuova versione, visualizzare diritti e categorizzazioni, ecc. Ci sarà la schermata in cui l'utente produttore può visualizzare gli oggetti che ha creato, con la barra laterale che permette di selezionarli a seconda delle loro caratteristiche. Ci sarà anche un esempio di piccola collezione, all'interno della quale gli oggetti sono collegati tra loro per via di una relazione. Inoltre ci sarà una striscia contenente le 9 aree descrittive dei metadati LOM più il campo "libro digitale", esattamente come viene visualizzato sulla piattaforma al momento del caricamento o della eventuale creazione di nuova versione di metadati.

[« Back to search results](#)

## Identifiers

<http://phaidra-sandbox.univie.ac.at/o:202617>

## Proprietario

Davide Montanaro

## Tipo di oggetto

IMMAGINE

## Versione

Versione 1

## Detail page views

1

## Link dell'oggetto

Visualizza  
Scarica (2.11 MB)  
Dublin Core  
Metadati di Phaidra  
Visualizzatore EXIF

## Questo è il recto di

Modifica

## Funzioni di Phaidra

Editor dei Metadati  
Diritti di accesso  
Modifica le relazioni  
Crea nuova versione

[Mostra i metadati completi](#)

## Titolo (ita)

Vista di Castel Telvana

## Autor

Montanaro, D. (Davide)

## Descrizione (ita)

Foto scattata dal monte Armentera, nel comune di Borgo Valsugana (TN). E' possibile vedere dall'alto il castello medievale.

## Object languages

Italiano

## Date

2017-09-05T12:10:18.604Z

## Diritti



Quest'opera è distribuita con Licenza  
Creative Commons Attribuzione 4.0 Internazionale.

<http://creativecommons.org/licenses/by/4.0/legalcode>

## Classificazione

EuroVoc 4.2, provincia autonoma di Trento

Figura 20 – Esempio di immagine su Phaidra SandBox

SEARCH\_PLACEHOLDER...

6 objects

↓ ↑ ↺ ↻



### Vista di Castel Telvana

Davide Montanaro

Foto scattata dal monte Armentera, nel comune di Borgo Valsugana (TN). E' possibile vedere dall'alto il castello medievale.

o:202617

05.09.2017



### Audio\_Test

Davide Montanaro

Si tratta di un test per il caricamento di un file audio su Phaidra Sandbox

o:202616

05.09.2017



### Facciata principale dell'eremo di San Lorenzo in Armentera

Davide Montanaro

Foto della facciata dell'eremo di San Lorenzo in Armentera, situato in Val di Sella nel comune di Borgo Valsugana. Risale al XIII secolo e al suo interno sono presenti numerosi affreschi di buona fattura.

o:159164

12.10.2016



### Affreschi nelle chiese medievali della Valsugana

Davide Montanaro

La presente collezione ha l'obiettivo di radunare tutto il materiale possibile riguardo agli affreschi medievali presenti nella varie chiese della Valsugana.

o:159163

[\[Browse\]](#)

12.10.2016



### Foto di un affresco dell'oratorio di San Rocco a Borgo Valsugana

Davide Montanaro

Si tratta dell'affresco principale, visibile non appena si entra dal portone superiore. Sono presenti anche parecchi interessanti graffiti.

o:159162

12.10.2016



### Abside dell'eremo di San Lorenzo in Armentera: Foto dell'altare, dalla quale si possono vedere piuttosto bene gli affreschi

Davide Montanaro

Per ulteriori informazioni visitare il sito: [http://www.visitvalsugana.it/it/cosa-scoprire/chiese-antiche/eremo-di-san-lorenzo\\_2184\\_ids/](http://www.visitvalsugana.it/it/cosa-scoprire/chiese-antiche/eremo-di-san-lorenzo_2184_ids/)

o:159161

## Filters

- Access
- Type
  - Image (4)
  - Book
  - Article
  - Text
  - Collection (1)
  - Video
  - Data
  - Container
  - Map
  - Resource
  - Sound (1)
- Size
- License
- Date
- Owner

davidem82

x

- Authors
- Roles

Figura 21 – Pagina per visualizzare i propri oggetti

« Back to search results

Browse

## Identifiers

http://phaidra-sandbox.univie.ac.at  
/o:159163

## Titolo (ita)

Affreschi nelle chiese medievali della Valsugana

## Autor

Montanaro, Davide

## Descrizione (ita)

La presente collezione ha l'obiettivo di radunare tutto il materiale possibile riguardo agli affreschi medievali presenti nella varie chiese della Valsugana.

## Object languages

Italiano

## Classificazione

EuroVoc 4.2, storia

## Oggetti della collezione (2)



o:159162 Foto di un affresco dell'oratorio di San Rocco a Borgo Valsugana



o:159161 Abside dell'eremo di San Lorenzo in Armentera: Foto dell'altare, dalla quale si possono vedere piuttosto bene gli affreschi

## Proprietario

Davide Montanaro

## Tipo di oggetto

COLLEZIONE

## Versione

Versione 1

## Detail page views

5

## Link dell'oggetto

Dublin Core  
Metadati di Phaidra

## Funzioni di Phaidra

Editor dei Metadati  
Diritti di accesso  
Modifica le relazioni  
Editor della collezione

Mostra i metadati completi

Figura 22 – Pagina che visualizza una collezione

Dati generali Ciclo di vita Dati tecnici Ambito didattico Diritti & Licenze Classificazione Relazioni Dati dell'oggetto originale  
Provenienza Libro digitale

Figura 23 - Aree descrittive di LOM

#### 4.4 L'estrazione e la visualizzazione (visualization and re-use):

In quest'ultima fase si parla di accesso permanente alla risorsa. Questa, essendo dotata di un link permanente, può essere citata senza rischio che successivamente ci si possa trovare in un vicolo cieco, ovvero senza la possibilità di conoscere la fonte. Ciò crea inoltre relazioni tra documenti, che sfociano in una realtà che esce dalla singola piattaforma Phaidra. Attraverso la rete e i suoi collegamenti diventa attuabile un'esportazione di file verso altri progetti affini, come Europeana e Open AIRE.

Questa fase ha il compito di assicurare l'accesso permanente ai dati, in accordo con le impostazioni offerte dal gestore del servizio e poi scelte dal produttore di dati. In questo modo si stabilisce, ad esempio, come rendere possibile l'invio degli oggetti ad applicazioni esterne specializzate nel riuso dei dati forniti. Questi dati possono inoltre essere utilizzati in sistemi di E-Learning, presentazioni o svariate altre tipologie di processi di ricerca. La responsabilità qui finisce a carico del fornitore di contenuti facente parte dell'applicazione esterna. Nella fase di riutilizzo quindi, i dati vengono trasferiti agli utenti autorizzati. Questa è la più importante delle quattro fasi dato che dà significato all'intera gestione dei dati di ricerca. Il trasferimento è simile a quello che avviene verso la gestione del repository, dove è necessaria una stretta collaborazione con gli utenti. Tuttavia, c'è la differenza che l'utente non ha alcun contatto diretto con la gestione del repository; pertanto, sono necessarie applicazioni che sono adattate alle esigenze di un successivo utilizzo<sup>79</sup>.

Tramite l'interfaccia utente si può facilmente accedere alla ricerca, riprendendo le sue funzioni spiegate nel sottocapitolo precedente. Questa funziona tramite parole chiave, le quali permettono al sistema di selezionare, tramite una query SQL, tutti i risultati affini all'interno del repository. Queste parole possono essere inserite in quanto parte della descrizione vera e propria dei contenuti intellettuali dell'oggetto, oppure possono essere legati ai permessi di accesso, al loro formato, etc. Praticamente riguardo a tutto quello che è stato inserito nei metadati. A partire dal funzionamento delle interfacce, avviene la trasmissione e la condivisione dei dati (Data Sharing). Questo punto è di grande importanza, in quanto la stessa idea di Open Science richiede che i risultati dei processi di ricerca e i relativi dati siano messi a disposizione degli istituti di ricerca e delle comunità scientifiche. La ricerca di un oggetto digitale può essere svolta ai fini di una semplice *visualizzazione* o di un eventuale *riuso*. Nel primo caso, dopo aver trovato la risorsa desiderata, è possibile ottenere la visualizzazione della stessa senza uscire dalla pagina web. Il software infatti, dispone di un proprio lettore per la riproduzione di alcune tipologie di supporto. Si parla qui di formati audio e video che vengono riprodotti (se di formato riconosciuto) all'interno del player preimpostato. Per quanto

---

<sup>79</sup> BUDRONI P., PREZA J. L., PAUSZ R., SÁNCHEZ B., BLUMESBERGER S., *Data Management Plan - Eine Anleitung zur Erstellung von Data Management Plänen. Projekt e-Infrastructures Austria (Deutsch): Eine Anleitung zur dauerhaften Sicherung von digitalen Beständen*, 2015.

riguarda le risorse esterne inoltre è possibile che intervenga il servizio di streaming. In altri casi, come quello dei documenti pdf, la risorsa viene scaricata in locale per essere visualizzata tramite il proprio software. In ogni caso qualsiasi oggetto, previo permesso da parte del produttore, è scaricabile gratuitamente. Caso particolare invece quello dei libri, in quanto vengono visualizzati nel Book Viewer, servizio offerto da Phaidra. Si tratta di un lettore che funziona online e dà la possibilità di sfogliare le pagine come se fosse aperto davanti a noi.

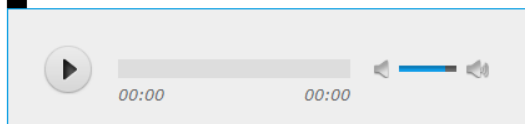
Nel caso del riuso invece, si passa ad un livello successivo. Infatti i dati escono dalla piattaforma, ma poi rientrano tramite un nuovo oggetto. La visualizzazione in questo può essere il primo passo e servire come una presentazione orientata ad aprire nuovi spazi di ricerca, sia che rientrino nell'ambito del singolo ricercatore che nell'ambito di un intero istituto di ricerca. In certi casi, con opportuni permessi, il riuso può avvenire tramite la creazione di una nuova versione del precedente oggetto. Altrimenti anche solo attraverso il riutilizzo di informazioni contenute nello stesso, le quali verranno conseguentemente citate. È necessario, prima di riutilizzare un oggetto, verificare la tipologia di licenza utilizzata onde evitare problemi legali. Talvolta è possibile ad esempio prendere un oggetto e rielaborarlo, in altri casi può essere solo preso e inserito così come è stato pubblicato. Nel momento in cui si desidera scaricare un oggetto, bisogna considerare che Phaidra non offre un sistema di conversione sul lato server. Non è possibile infatti operare una conversione diretta, ma si può integrare questa scelta tramite un plug-in.

Mentre i metadati sono sempre accessibili a livello pubblico, l'accesso ai contenuti dell'oggetto può essere limitato anche solo per un lasso determinato di tempo. Se il periodo di blocco scade, l'oggetto viene automaticamente messo a disposizione di tutti gli utenti. Phaidra offre una varietà di modi per collegare la directory dell'utente e quella dell'organizzazione in modo che le funzionalità interne elencate possano essere realizzate. Per quanto riguarda l'accesso esterno, Phaidra fornisce alcune API, denominate applicazioni “docked”, che consentono ad altre applicazioni di accedere a metadati e oggetti. Ne esistono già molteplici, alcune delle quali sono sotto licenza open source. Una parte di queste si trovano nello spazio phaidra di Github, tra cui l'OAI provider, che permette di operare con i metadati DC<sup>80</sup>. Si può mettere in comunicazione ad esempio con Europeana o OpenAIRE, indirizzando i file presenti e rendendoli disponibili dalla loro interfaccia.

In questa sezione vedremo, col supporto delle immagini, come viene visualizzato il lettore audio sulla piattaforma, come viene visualizzato il Book viewer ed un esempio di come i metadati vengano esportati in DC all'interno di un file XML. Da notare in quest'ultimo l'utilizzo del protocollo OAI che va ad aprire e chiudere il codice.

---

<sup>80</sup> BERGER C. et al., *Digitale Archivierung und Bereitstellung von AV-Medien – Erfahrungen und Praxisbeispiele aus dem tertiären Bildungssektor*, 2017, pp 108-112.

[« Back to search results](#)

### Titolo (ita)

Audio\_Test

### Autor

Montanaro, D. (Davide)

### Descrizione (ita)

Si tratta di un test per il caricamento di un file audio su Phaidra Sandbox

### Object languages

Italiano

### Date

2017-09-05T11:50:42.123Z

### Diritti

© Tutti i diritti riservati

### Classificazione

ÖFOS 2012, Komposition

### Identifiers

<http://phaidra-sandbox.univie.ac.at/o:202616>

#### Proprietario

Davide Montanaro

#### Tipo di oggetto

AUDIO (MP3)

#### Versione

Versione 1

#### Detail page views

1

#### Link dell'oggetto

[Visualizza](#)  
[Scarica \(2.21 MB\)](#)  
[Dublin Core](#)  
[Metadati di Phaidra](#)  
[Visualizzatore EXIF](#)

#### Funzioni di Phaidra

[Editor dei Metadati](#)  
[Diritti di accesso](#)  
[Modifica le relazioni](#)  
[Crea nuova versione](#)[Mostra i metadati completi](#)

Figura 24 – Visualizzazione lettore audio



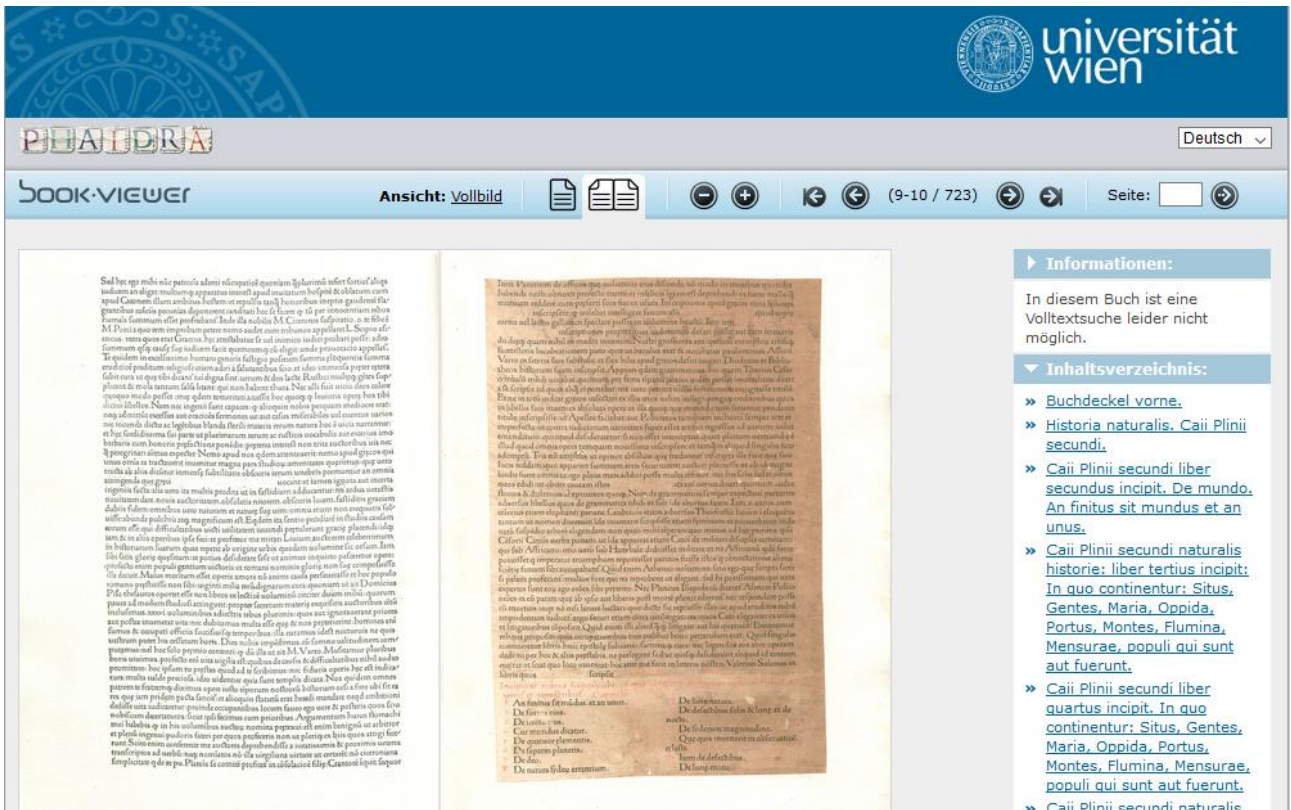


Figura 25 – Book Viewer

```

- <oi:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oi_dc http://www.openarchives.org/OAI/2.0/oi_dc.xsd">
  <dc:title xml:lang="ita">Affreschi nelle chiese medievali della Valsugana</dc:title>
  <dc:creator>Montanaro, Davide</dc:creator>
  <dc:subject xml:lang="eng">EuroVoc 4.2, history</dc:subject>
  <dc:subject xml:lang="ita">EuroVoc 4.2, storia</dc:subject>
  <dc:subject xml:lang="deu">EuroVoc 4.2, Geschichtswissenschaft</dc:subject>
- <dc:description xml:lang="ita">
  La presente collezione ha l'obiettivo di radunare tutto il materiale possibile riguardo agli affreschi medievali presenti nella varie chiese della Valsugana.
  </dc:description>
  <dc:type xml:lang="eng">Collection</dc:type>
  <dc:identifier>o:159163</dc:identifier>
  <dc:identifier>http://phaidra-sandbox.univie.ac.at/o:159163</dc:identifier>
  <dc:language>ita</dc:language>
</oi:dc:dc>

```

Figura 26 – Esportazione metadati in DC



## **Capitolo 5 - La gestione delle immagini: presupposti e formati presi in esame**

Dal momento che la mia scelta è quella di concentrarmi sui file di tipo immagine, il capitolo sarà incentrato su questo, attraverso un'analisi che percorre tutte le fasi di lavorazione dell'oggetto. Usualmente si tende a considerare le immagini ristrettamente al campo della fotografia; in un ambiente digitale risulta necessario ampliare questa visione anche a categorie che di fatto non corrispondono all'uso più canonico del termine. Parliamo quindi anche di mappe geografiche, di libri e di tutti quei documenti che dopo esser stati trattati da uno scanner diventano "immagine" (nel mondo anglosassone per esempio un libro elettronico viene anche chiamato "image of a book"). Queste risorse, nel corso della loro digitalizzazione vengono trattate in maniera affine seppur con le loro necessarie differenze. Una breve sarà poi dedicata alle "immagini che si muovono", ossia i filmati; che appunto, altro non sono, che una lunghissima sequenza di immagini collegate tra loro.

Va considerato che, nella fase di visualizzazione, può esserci un plug-in diverso a seconda del tipo di contenuto specifico in Phaidra. Ad esempio, il tipo "Libro" viene visualizzato con il "Visualizzatore di libri", così come ci sono possibilità simili per mappe geografiche e video. Nei sotto-capitoli si entrerà nel dettaglio di ogni tipologia di immagine, a partire dai presupposti FAIR. Questi, come verrà spiegato nel corso della sezione 5.1, fungono da linee guida per tutto quello che riguarda le pratiche per un'ampia gestione e interoperabilità dei dati. Nonostante poi si parlerà esclusivamente di immagini, questi sono validi per tutte le tipologie di dati.

### 5.1 I presupposti (FAIR Principles):

Considero molto interessante quanto riportato nell'ambito dei principi FAIR, progettati per funzionare come linee guida nel campo della ricerca, riutilizzo, accessibilità ed interoperabilità dei dati. Questo è un punto che deve urgentemente essere sviluppato nel mondo accademico, in quanto con la messa a punto di meta-dati liberi e universalmente accettati, la descrizione adeguata del patrimonio documentario diventa possibile e apre infiniti orizzonti di scoperta. Questi principi, che verranno illustrati nel corso del sotto-capitolo, sono elencati nel sito di FORCE11<sup>81</sup>, comunità di professionisti che hanno a cuore la ricerca e la condivisione del sapere. Interessante anche comprendere una lettura di questa concezione di metadati da un punto di vista ideale, come suggerito dalla rivista online nature.com<sup>82</sup>.

*There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders - representing academia, industry, funding agencies, and scholarly publishers – have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals.*

[NATURE.COM, *The FAIR Guiding Principles for scientific data management and stewardship*]

I risultati della ricerca ottenuti tramite un finanziamento pubblico, dovrebbero essere accessibili ai fini del riuso. Durante i processi di ricerca scientifici vengono generati diversi tipi di output a seconda della disciplina, delle fonti e delle tipologie di analisi effettuate dagli studiosi. Per condividere e riutilizzare i dati all'interno dell'ambiente Open Science, è importante fornire certezze riguardo la qualità dei dati offerti e avere accordi effettivi e funzionali per un efficiente processo di riuso. Nel momento in cui le fonti sono archiviate e rese disponibili, deve essere chiaro come queste siano strutturate e che informazioni esse contengano. C'è un accordo piuttosto ampio nel sostenere che i volumi di dati e la loro complessità stanno aumentando esponenzialmente e allo stesso tempo sta emergendo la necessità del riuso in praticamente tutti i campi di ricerca, sfruttando le possibilità offerte dalla interdisciplinarietà e dalla collaborazione all'interno di entità politiche differenti. Come risultato di questi sviluppi sono emerse, come messo in rilievo dalla commissione europea, diverse

---

<sup>81</sup> The FAIR Data Principles, <https://www.force11.org/group/fairgroup/fairprinciples> [Ultimo accesso 30/09/2017]

<sup>82</sup> The FAIR Guiding Principles for scientific data management and stewardship, <http://www.nature.com/articles/sdata201618> [Ultimo accesso 30/09/2017]

carenze<sup>83 84</sup>:

- Vari tipi di inefficienza, come l'80% dei dati che diventa inaccessibile dopo un breve periodo o l'80% di tempo sprecato da parte degli specialisti per compiti di ordinaria gestione dei dati.
- La mancanza di una struttura concettuale in grado di unificare le infrastrutture e l'uso di pratiche adeguate che hanno portato ad un'enorme frammentazione, nonché la mancanza di interoperabilità.
- La carenza di esperti che possono aiutare in una fase di rapida transizione verso pratiche più efficienti.

La ragione per cui spesso abbiamo bisogno di parecchie settimane o mesi di sforzi per raccogliere i dati necessari a rispondere ai quesiti posti dalle varie aree di ricerca, non è la mancanza di tecnologia. Il problema è che spesso non viene data agli oggetti digitali l'attenzione e la cura che essi meritano al momento della loro creazione e successiva archiviazione. Esistono alcune repository che, ad esempio, accettano solo alcuni tipi di dati. Se da un lato questo aspetto diventa positivo per fattori come l'interoperabilità, dall'altro rischiano di venire escluse altre informazioni di supporto importanti quanto i risultati stessi. Per ovviare a questo problema sono emerse nuove repository più generali (siano esse di una singola istituzione o globali), che accettano tanti differenti tipi di dati e relativi formati; mancando però nella loro armonizzazione. In questo modo l'ecosistema dei dati si allontana dalla centralizzazione, diventa più diversificato e di conseguenza meno integrato, portando all'aggravarsi del problema della ricerca e del riuso da parte di soggetti umani e computazionali<sup>85</sup>. Affinché ciò venga evitato, i singoli oggetti digitali devono arrivare ad avere un ruolo di punta all'interno dell'ecosistema scientifico, dove la qualità e l'impatto della pubblicazione possano essere funzionali nel garantire un riuso accurato ed appropriato, oltre ad offrire una permanente possibilità di citazione da parte di tutti gli Stakeholders, siano essi umani o macchine.

A partire dalla consapevolezza raggiunta riguardo a queste problematiche, sono state sviluppate a livello europeo diverse raccomandazioni sulle quali è stato trovato un ampio consenso:

- ◆ I progetti di ricerca devono preparare e di conseguenza seguire un Data Management Plan, il quale porta i ricercatori a prendere la gestione dei dati più seriamente.
- ◆ È necessario attenersi al principio dell'Open Data.
- ◆ Una determinata parte dei finanziamenti ricevuti per il progetto di ricerca deve essere dedicata

---

<sup>83</sup> AMSTERDAM CALL FOR ACTION ON OPEN SCIENCE, basato sulla conferenza *Open Science – From Vision to Action*, tenutasi ad Amsterdam il 4 - 5 aprile 2016. Disponibile a: <http://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf> [Ultimo accesso 12/09/2017]

<sup>84</sup> WEIGEL T., WITTENBURG P., *Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data*, 2017. Disponibile a: <https://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf> [Ultimo accesso 12/09/2017]

<sup>85</sup> WILKINSON M. D. et al., *The FAIR Guiding Principles for scientific data management and stewardship*, 2016. Disponibile tramite DOI: 10.1038/sdata.2016.18

ad una corretta gestione e amministrazione dei dati, anche dopo il termine dello stesso.

- ◆ Vanno create delle condizioni ottimali per condividere i risultati della ricerca tramite l'introduzione dei principi FAIR, richiesti come base per un'appropriata gestione dei dati.

I principi FAIR in particolare, risultato di lunghe e ampie discussioni, hanno il potenziale per condurre verso notevoli miglioramenti. Si tratta di una serie di dichiarazioni relative agli oggetti digitali che vengono depositati in repositories affidabili, dove sono associati ai loro metadati e dove ricevono un identificatore PID (Persistent Identifier). Tutte queste operazioni devono essere svolte per tutti i tipi di oggetti, in modo da migliorare il riuso e l'interoperabilità. Si tratta di regole, semplici e ben definite, di cui va tenuto conto a partire dal momento in cui va redatto il Data Management Plan. A seguire vediamo in cosa consistono, seguendo la divisione per i quattro obiettivi principali (To be findable, accessible, interoperable, re-usable):

**1. To be Findable:**

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

**2. To be Accessible:**

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2 metadata are accessible, even when the data are no longer available.

**3. To be Interoperable:**

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

**4. To be Re-usable:**

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

Affinché si possa giungere agli scopi previsti, sono in corso delle azioni concrete che mirano ad aumentare la consapevolezza sul tema. La commissione europea ad esempio, tramite il programma H2020, ha redatto delle linee guida chiamate: “Guidelines on FAIR Data Management in Horizon 2020”<sup>86</sup>. L'obiettivo è quello di promuovere la pubblicazione, compatibile con i principi FAIR, dei

---

<sup>86</sup> EUROPEAN COMMISSION, *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*, 2016. Disponibile a: <[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-)

risultati della ricerca effettuati con fondi pubblici, possibilmente con la collaborazione delle autorità nazionali. Gli enti di ricerca, assieme ai loro finanziatori, dovrebbero introdurre degli incentivi per l'utilizzo dei principi FAIR, magari ricompensando chi se ne occupa. Il tutto richiedendo che i dati siano resi citabili in accordo con gli standard internazionali. Potrebbe essere anche incoraggiata la condivisione di esperti della disciplina affinché diffondano la sua conoscenza. In ogni caso va creato un ponte tra IT e ricerca, tramite la formazione di esperti di gestione dei dati.

Tra gli effetti positivi che questi miglioramenti possono apportare troviamo il miglioramento della qualità della ricerca, una migliore aderenza ai principi della buona ricerca scientifica e il favorirne l'integrità, migliorare l'impatto dei fondi pubblici spesi per la ricerca e assicurare la condivisione e il riuso dei dati, favorendo la scienza e l'innovazione<sup>87</sup>.

---

mgt\_en.pdf> [Ultimo accesso 12/09/2017]

<sup>87</sup> AMSTERDAM CALL FOR ACTION ON OPEN SCIENCE, basato sulla conferenza *Open Science – From Vision to Action*, 2016, pp. 16-17. Disponibile a:

<<http://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf>> [Ultimo accesso 12/09/2017]

## 5.2 Le immagini:

Come accennato precedentemente, le immagini sono uno dei punti chiave su cui è mio interesse focalizzarmi. Nell'immagine sottostante si può osservare la fase in cui si inizia la procedura per la creazione di un nuovo oggetto.

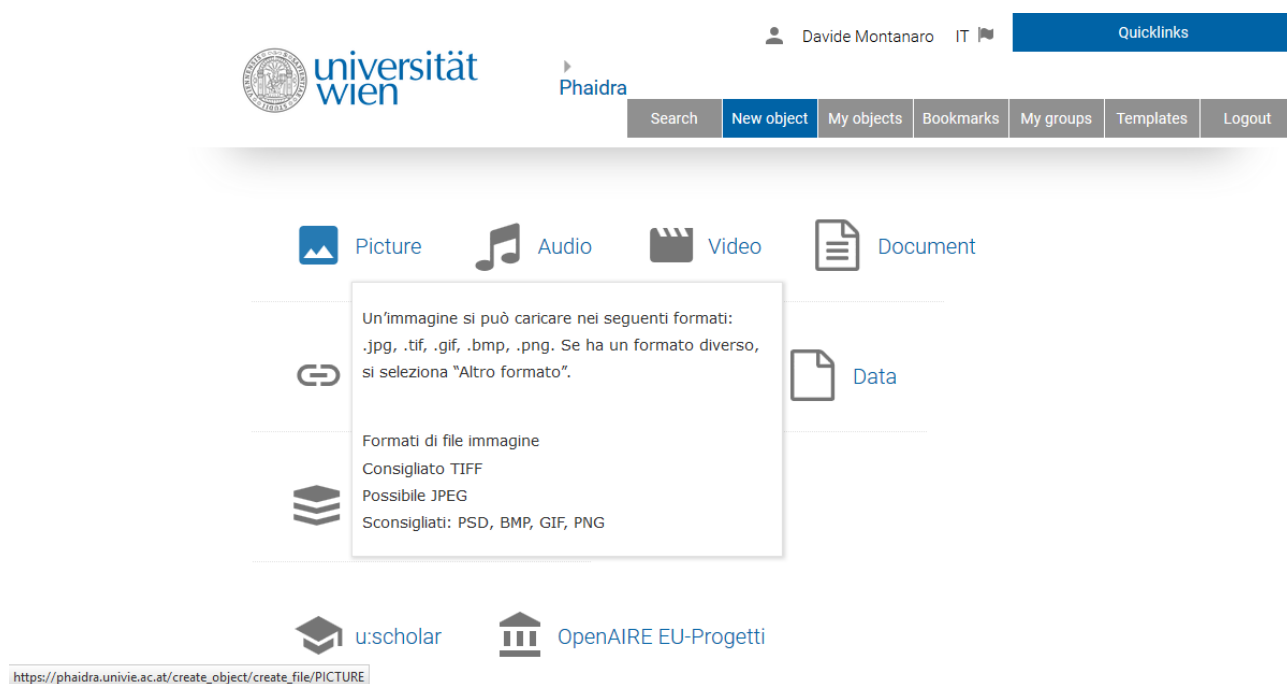


Figura 27 – Screenshot Phaidra

[https://phaidra.univie.ac.at/create\\_object](https://phaidra.univie.ac.at/create_object)

Con questa sezione, è mia intenzione entrare nell'ampio mondo delle immagini digitali in Phaidra, con lo scopo di offrire un quadro completo di questa dimensione. In queste prime righe vorrei introdurre i temi chiave su cui verterà questo sotto-capitolo, spiegando cosa verrà trattato tramite una sequenza che punta ad essere coerente con le fasi di lavoro esplicitate nel capitolo precedente. Per cominciare dunque, ci sarà una panoramica di ciò che accade nel corso della preparazione, comprendendo che tipologie di immagini possono venire trattate in Phaidra e i loro supporti, in relazione al fatto che queste entrano in un repository di archiviazione a lungo termine. Verrà tracciato un profilo dei formati utilizzabili per il caricamento e della possibile necessità di convertire i file, analizzando le motivazioni che inducono a prendere una determinata decisione. Ci sarà qualche accenno legato alla pratica della digitalizzazione di risorse fisiche in vista della loro conservazione e riuso online, con una traccia delle caratteristiche che andranno a delineare l'oggetto digitale. Successivamente verrà mostrato un esempio di caricamento su Phaidra Sandbox, dove sarà seguito passo per passo l'inserimento dei metadati divisi a seconda della loro area descrittiva; oltre a due parole sull'uso della tecnologia OCR per le immagini contenenti del testo e sulla possibilità di collegare tramite relazione gli oggetti. Infine, ci sarà spazio per parlare di visualizzazione e riuso tramite il successivo sfruttamento dei metadati, sia quelli precedentemente descritti, sia quelli relativi



allo standard EXIF.

Vediamo dunque brevemente che tipologie di immagini trattiamo in questa sezione, creando un quadro complessivo dell'ambiente che ci ruota attorno. Si parla di oggetti che possono essere fotografie, scansioni, analisi da microscopio, presentazioni create al computer, ecc. Quando questi file vengono caricati sulla piattaforma, possono venire da diversi supporti: il disco rigido del proprio pc o un disco esterno, una chiavetta USB, un cloud, un sito internet. Possono inoltre avere due diverse origini. Essere concepiti come oggetti fisici, come un classico negativo, da sottoporre successivamente a digitalizzazione oppure nascere direttamente come oggetto digitale. In entrambi i casi, devono passare attraverso degli strumenti di acquisizione che creano la sequenza di bit che caratterizza la risorsa, permettendoci di visualizzarla come desideriamo. Tra questi possiamo menzionare la macchina fotografica digitale, gli scanner a letto piano, gli scanner per pellicole, negativi e diapositive, gli scanner planetari, i sistemi di acquisizione per la microscopia elettronica. Tutti questi strumenti devono garantire la qualità dell'immagine come richiesto dal progetto, senza andare a danneggiare eventuali oggetti originali. Questa eventualità può essere causata ad esempio da luci utilizzate per la messa a fuoco. Interessante è capire il diverso utilizzo che può essere fatto dagli scanner. Nel caso del modello a letto piano, molto più comune da trovare, si processano normalmente documenti a fogli singoli o comunque rilegature facili da aprire; sempre di una dimensione inferiore all'A3. Alcuni esempi di materiale da digitalizzare potrebbero essere fotografie, volantini, spartiti musicali, lettere, piccole mappe, disegni. Lo scanner planetario invece, non presupponendo un contatto diretto con l'oggetto interessato, può venire utilizzato per risorse di dimensioni più grandi dell'A3 e di risorse che potrebbero venire danneggiate dal processo, come pergamene, libri antichi, disegni a pastello, quadri ad olio ed altro ancora. In determinati casi, a questo scopo può essere usata anche la macchina fotografica digitale.

Ma quali sono gli scopi per cui uno studioso dovrebbe mettere delle immagini su Phaidra? Se ne è parlato anche negli scorsi capitoli, esplorando le motivazioni per cui queste piattaforme stanno prendendo piede. Rivediamo brevemente questi concetti immaginandoli nell'ottica di un oggetto come le immagini, le quali possono valorizzare il patrimonio documentario dell'istituto universitario e soddisfare le esigenze di accesso rapido e facile degli utenti. Tutto ciò considerato in un'ottica multidisciplinare, promuovendo la collaborazione tra dipartimenti e diversi istituti di ricerca. Inoltre, possono promuovere la conoscenza di collezioni uniche su larga scala, creare collezioni virtuali tramite l'integrazione di vari formati o di materiali distribuiti in luoghi diversi, evitare un ulteriore consumo di documenti originali in condizioni precarie e, non per ultimo, garantire che tutto il materiale caricato sia a disposizione delle future generazioni.

Abbiamo parlato nel capitolo 4.1 di formati, i quali possono essere o meno proprietari. Vediamo ora nel dettaglio quali sono i principali formati utilizzati per le immagini, attraverso le loro caratteristiche e le motivazioni che li portano ad essere consigliati o sconsigliati per la conservazione a lungo termine. In Phaidra viene raccomandato l'uso del formato TIFF, ma viene supportato anche l'uso del JPEG. È bene partire dal presupposto che ci sono due tipologie di formati: Master e derivati. Il primo rappresenta la miglior resa possibile della digitalizzazione, di solito il risultato "grezzo" prima del processo di rielaborazione, che si presta a svariate possibilità di adattamento in vista dell'obiettivo finale. I derivati di solito rientrano nella seconda fase, escludendo determinata attrezzatura amatoriale che salta il passaggio elaborativo creando un prodotto già finito. Nel momento in cui si opera una scelta, è necessario fare un compromesso tra qualità, costi e opportunità di riuso. I costi vanno considerati a partire dall'acquisizione, come attrezzatura e personale qualificato, fino alla successiva gestione e archiviazione dell'oggetto. Va valutato quindi ogni singolo caso in base all'importanza del documento, alla sua dimensione e alle risorse a disposizione. Tendenzialmente è sempre possibile fare delle conversioni in vista di un caricamento sulla piattaforma, senza dimenticare di considerare che, in alcuni casi, certe informazioni vanno perse per sempre. Di seguito le caratteristiche con qualche breve considerazione sui più usati formati di immagini:

- **RAW:** Si tratta di un formato di cui si parla molto ultimamente, in quanto propone un'ottima qualità, registrando l'immagine esattamente come proviene dal sensore della fotocamera. È una specie di negativo digitale, in quanto non subisce elaborazioni ed è l'immagine nella sua forma più pura. Come vantaggi ha un'alta profondità di colori ed è adatto per un successivo riutilizzo, lasciando aperta qualsiasi possibilità di rielaborazione. Si tratta però di un formato proprietario, relativo solo alle fotografie e non adatto per una conservazione a lungo termine. Non usa alcun tipo di compressione, per questo i file diventano molto pesanti. Inoltre non molti applicativi garantiscono la sua apertura.
- **JPEG (Joint Photographic Experts Group):** Formato aperto e nato con l'idea di utilizzare metodi di compressione efficaci. Sfortunatamente la compressione è sempre con perdita, quindi più un'immagine viene compressa, più viene irrimediabilmente deteriorata. Ha limitate possibilità di cambiare i colori. È adatto però per la conservazione a lungo termine, in quanto si tratta di un formato universalmente accettato, è agile e pronto all'uso. Inoltre non occupa molto spazio ed è facilmente trasferibile. Questi fattori lo rendono supportato da Phaidra.
- **TIFF (Tagged Image File Format):** si tratta di un formato che supporta una compressione lossless, senza degrado dell'immagine. Questa opportunità è interessante perché, al contrario del JPG, se modifico molte volte un'immagine questa non perde mai di qualità. Inoltre viene supportato da quasi tutti gli applicativi che si occupano di immagini, seppur con qualche

piccola eccezione. È il formato consigliato per il salvataggio delle immagini su Phaidra. Queste occupano molto spazio, nonostante i vari tipi di compressione utilizzabili che in certi casi possono portare a delle perdite.

- **PSD** (PhotoShop Document): Formato proprietario di Adobe Photoshop. È diventato una sorta di standard nel campo della rielaborazione delle immagini, ma si tratta di un formato proprietario non adatto alla conservazione. Supporta il salvataggio di diversi livelli nel corso dei processi lavorativi dell'immagini, ma non è utilizzabile dalle altre applicazioni (con qualche eccezione).
- **BMP** (Windows Bitmap): Formato proprietario di Windows di tipo bitmap, non compresso.
- **PNG** (Portable Network Graphics): Formato aperto con metodo di compressione lossless; non possiede però un metodo per verificare l'integrità dei file.
- **GIF** (Graphics Interchange Format): Formato aperto con una strana caratteristica riguardo alle compressioni; infatti comprime senza perdita solo nel caso di immagini con un numero di colori inferiore o uguale a 256. In caso contrario la compressione diventa lossy.

Riprendendo quanto detto finora dal punto di vista dei processi lavorativi, è chiaro come le motivazioni, i formati, i supporti e il resto rientrano nella fase di preparazione (pre-ingest). Come già precedentemente affrontato però, non sono queste le uniche cose importanti. Va prestata attenzione anche agli aspetti legali connessi, come tutto ciò che riguarda la tutela della privacy, le leggi a riguardo e le licenze. Tutti questi fattori concorrono nelle valutazioni che portano alla selezione dei materiali da caricare sulla piattaforma. Questi infatti possono soddisfare diversi criteri per i quali valga la pena effettuare tutte le operazioni di conservazione; si parla di valore storico e culturale, di unicità e rarità, di materiale con permessi legali di digitalizzazione o senza vincoli a riguardo oppure di oggetti con un accesso ristretto dovuto ad uno stato di conservazione precario o alla difficile localizzazione. Potrebbe essere anche un oggetto che, da solo risulta poco interessante, ma aggiunto ad una collezione ne accresce il valore della stessa. Tutte queste informazioni poi possono essere inserite nella descrizione, diventando utili per alcune attività di visualizzazione e riuso.

Una volta terminata la fase di preparazione, avendo a disposizione l'oggetto e tutte le informazioni necessarie, siamo pronti per la fase di caricamento. È qui che, dopo un'attenta valutazione, va inserito il nome e tutti i tag che andranno a identificare per sempre l'oggetto. Risulta quindi di fondamentale importanza assegnare un nome trovabile al Digital Object in creazione. Infatti nel momento in cui un utente va a cercare delle risorse, non ha a disposizione alcune sottocategorie di immagini e ciò potrebbe rendergli la vita difficile. Se la ricerca non è bene indirizzata tramite i metadati descrittivi, il file rischia di diventare introvabile e quindi inutile.



Esiste la possibilità di creare delle relazioni tra immagini: si possono radunare in collezioni, anche con oggetti di altre tipologie, rendendole sempre correlate tra loro per vari motivi. Un altro tipo di

relazione invece legato solo al tipo immagine, è quello di fronte – retro. Ciò implica che la relazione sia esclusivamente tra due oggetti. Potrebbe essere, ad esempio, il caso di un volantino che viene scansionato su entrambi i lati, i quali non avrebbero motivo di stare separati e sarebbe uno spreco inutile creare una intera collezione per loro. Un’ulteriore opportunità a disposizione nella conservazione delle immagini è quella di sottoporre, tramite un software di riconoscimento del testo (OCR), il file ad un processo di lettura dei contenuti automatizzato. In questo caso risulta possibile creare dei file legati all’immagine: un PDF di testo, che diventa di conseguenza copiabile e ricercabile oppure un XML, nel momento in cui si voglia caricare più immagini a formare un libro. Questo secondo caso lo vedremo più dettagliatamente nel sotto-capitolo 4.4. Seguiamo ora, tramite il supporto di Phaidra Sandbox, il caricamento completo di un file immagine.

The screenshot shows the 'Dati generali' (General Data) form in Phaidra Sandbox. The form is organized into several sections, each with a label and a corresponding input field or dropdown menu. The language for all text fields is set to 'Italiano'.

- Identificatore\*:** o:211196
- Titolo\*:** Vista di Castel Telvana
- Sottotitolo:** Scatto da uno spiazzo del monte Ar
- Titolo alternativo:** Castel Telvana dal monte Armenter
- Externalview:** UniWien
- Lingua\*:** Italiano
- Descrizione\*:** Foto scattata dal monte Armentera, nel comune di Borgo Valsugana (TN). Sotto il castello è visibile anche una parte del paese.
- Parole chiave:** Castello, XIII secolo, Telvana, Borgo
- Copertura (temporale – geografica – amministrativa):** XIII secolo; Borgo Valsugana; Trenti

Figura 28 – Dati Generali

Identificatori  

Standard:

Identificatore:

- (nessuna selezione) ▾
- (nessuna selezione)
- DOI**
- eISSN
- HTTP/WWW
- ISBN
- ISMN
- ISSN
- Numero-AC
- PI
- PrintISSN
- URN
- VLB-Systemnummer
- VLB-Systemnummer

**Continua**

Information for ...

Figura 29 – Identificatori


Dati generali **Ciclo di vita** Dati tecnici Ambito didattico Diritti & Licenze Classificazione Relazioni Dati dell'oggetto originale Provenienza Libro digitale

Data di caricamento di Phaidra\*: 22.09.2017 - 19:48:24


Versione:  nella lingua: Italiano ▾

Stato\*: Completo ▾


Sottoposto a revisione paritaria (Peer Reviewed): No ▾

Contributo 

Ruolo\*: Autore dell'oggetto digitale ▾

**Ruolo / Dati personali\*** 

Nome:  Cognome:  Titolo:  Titolo:

Data: (DD.MM.YYYY): . nessuna specifica ▾.  

Si prega di utilizzare i seguenti campi solo nel caso di progetti OpenAIRE (FP7 Projects)

OpenAIRE Version Type: (nessuna selezione) ▾

Figura 30 – Ciclo di vita

Dati generali Ciclo di vita **Dati tecnici** Ambito didattico Diritti & Licenze Classificazione Relazioni Dati dell'oggetto originale Provenienza Libro digitale

Formato:  +

Dimensione: 2.11 MB

Link permanente: <http://phaidra-sandbox.univie.ac.at/o:211196>

Requisiti ⇌ +

Guida all'installazione:  nella lingua: Italiano ▾  
 Aggiungi uno o più link

Requisiti per l'utilizzo dell'oggetto:  nella lingua: Italiano ▾  
 Aggiungi uno o più link

Durata: (HH:MM:SS):  :  :

Figura 31 – Dati tecnici

Dati generali Ciclo di vita Dati tecnici **Ambito didattico** Diritti & Licenze Classificazione Relazioni Dati dell'oggetto originale Provenienza Libro digitale

Attenzione: devi compilare l'elemento "Tipo di risorsa didattica e pedagogica" se inserisci i metadati di questa scheda. È un'informazione necessaria agli utenti.

Ambito didattico ⇌ +

Figura 32 – Ambito didattico

Dati generali Ciclo di vita Dati tecnici Ambito didattico **Diritti & Licenze** Classificazione Relazioni Dati dell'oggetto originale Provenienza Libro digitale

In qualità di autore o di titolare dei diritti, è possibile autorizzare l'uso, in particolare la riproduzione, la distribuzione, la trasmissione, la fornitura e la modifica del tuo oggetto mediante una licenza Creative Commons. È possibile specificare altre condizioni per utilizzare il tuo oggetto scegliendo la licenza più appropriata.

Costi\*:

Copyright e altre limitazioni\*:

Licenza scelta\*: *Attenzione: quando scegli la formula "Tutti i diritti riservati", hai la possibilità di cambiarla con un'altra licenza. Quando scegli una licenza CC, non è più possibile cambiarla.!*

Tutti i diritti riservati  
 Licenza GNU  
 Public Domain Mark 1.0  
 Attribuzione 4.0 Internazionale  
 Attribuzione - Condividi allo stesso modo 4.0 Internazionale  
 Attribuzione - Non commerciale 4.0 Internazionale  
 Attribuzione - Non opere derivate 4.0 Internazionale  
 Attribuzione - Non commerciale - Non opere derivate 4.0 Internazionale  
 Attribuzione - Non commerciale - Condividi allo stesso modo 4.0 Internazionale  
 Other licenses

Versione breve della licenza

Se scegli Tutti i diritti riservati, il tuo oggetto può essere utilizzato solo nei limiti del diritto d'autore. Inoltre, non autorizzi né un uso commerciale né un uso non commerciale, specialmente per quanto riguarda riprodurre, distribuire, comunicare al pubblico, esporre in pubblico, rappresentare, eseguire e recitare quest'opera.

Scegliere una licenza Creative Commons promuove la diffusione della tua opera. Per ulteriori informazioni <http://www.creativecommons.it>

**ESCLUSIONE DI RESPONSABILITÀ (DISCLAIMER)**

Descrizione:  nella lingua:

[Aggiungi uno o più link](#)

Si prega di utilizzare i seguenti campi solo nel caso di progetti OpenAIRE (FP7 Projects)

OpenAIRE Access Rights:

OpenAIRE Embargo End: (DD.MM.YYYY):

Figura 33 – Diritti e Licenze

Dati generali Ciclo di vita Dati tecnici Ambito didattico Diritti & Licenze **Classificazione** Relazioni Dati dell'oggetto originale Provenienza Libro digitale

Classificazione

Classificazioni (Classi, Sottoclassi)

[Aiuto per la ricerca](#)

Tipo di classificazione:

Descrizione:

Parole chiave:

Figura 34 – Classificazione

Dati generali Ciclo di vita Dati tecnici Ambito didattico Diritti & Licenze Classificazione **Relazioni** Dati dell'oggetto originale Provenienza Libro digitale

Tipologia pubblicazione: (nessuna selezione) ▾ +

Data di approvazione: (DD.MM.YYYY):  . nessuna specifica ▾  📅 +

Struttura 🗑️ +  
 ▾

Corso di laurea ↗️ +

ulteriore assegnazione:

Figura 35 – Relazioni

Dati generali Ciclo di vita Dati tecnici Ambito didattico Diritti & Licenze Classificazione Relazioni **Dati dell'oggetto originale** Provenienza Libro digitale

Descrizione:  nella lingua: Italiano ▾ +

Dimensioni ↗️ +

Numero di riferimento ↗️ +



Timbro: (nessuna selezione) ▾ +

Note:  nella lingua: Italiano ▾  
 Aggiungi uno o più link

Coordinate geogr.: Longitudine:  Latitudine:

Figura 36 – Dati dell'oggetto originale



Informazioni sulla fonte  

Tipo di materiale:

Note:

nella lingua:  

[Aggiungi uno o più link](#)

Ruolo:

Dati personali o istituzionali 

Nome:  Cognome:  Titolo:  Titolo:  

Data a partire da: (DD.MM.YYYY):    

Data fino a: (DD.MM.YYYY):    

Periodo:  nella lingua:  

Collocazione:  nella lingua:  

Figura 37 – Provenienza

Dati generali   Ciclo di vita   Dati tecnici   Ambito didattico   Diritti & Licenze   Classificazione   Relazioni   Dati dell'oggetto originale   Provenienza   **Libro digitale**

Titolo della monografia:  nella lingua: Italiano

Pagine o volume della monografia:

Titolo della collana o della rivista:

Volume della collana o della rivista:

Numero di fascicolo della rivista:

da pagina:

a pagina:

Titolo della collezione:  nella lingua: Italiano

Luogo di pubblicazione:

Editore:

Data di pubblicazione:(DD.MM.YYYY):  . nessuna specifica

Edizione/ Tiratura:

Note:  nella lingua: Italiano

[Aggiungi uno o più link](#)

Supporto: (nessuna selezione)

URL Catalogo:

Figura 38 – Libro digitale

Una volta terminato il caricamento dell'immagine, c'è un'altra opzione, relativa solo alle immagini, che merita di essere notata: si tratta dei metadati EXIF. EXIF sta per Exchangeable Image File Format, e sintetizza tutte le informazioni relative all'acquisizione digitale dell'oggetto. Nel caso di una fotografia ad esempio, indica automaticamente il tipo di fotocamera e l'obiettivo utilizzati, i valori di esposizione e molti altri parametri. Nell'immagine a seguire ho estrapolato tutti i valori EXIF contenuti all'interno dell'ultima immagine da me caricata su Sandbox.

**MakerNotes**

Version	0130
Internal Serial Number	4C3230302020 2008:10:28 B2958WA05150
Quality	NORMAL
Sharpness	Normal
White Balance	Auto
Saturation	Normal
Fuji Flash Mode	Off
Flash Exposure Comp	0
Macro	Off
Focus Mode	Auto
AF Mode	Single Point
Focus Pixel	1824 1368
Slow Sync	Off
Picture Mode	Auto
Exposure Count	1
Auto Bracketing	Off
Sequence Number	0
Color Mode	Standard
Image Stabilization	Sensor-shift: On (mode 2, shooting only); 0
Faces Detected	0

**ExifTool**

ExifTool Version Number	10.20
-------------------------	-------

**PrintIM**

PrintIM Version	0250
-----------------	------

**EXIF**

Make	FUJIFILM
Camera Model Name	FinePix S2000HD
Orientation	Horizontal (normal)
X Resolution	72
Y Resolution	72
Resolution Unit	inches
Software	Microsoft Windows Photo Viewer 6.3.9600.17415
Modify Date	2015:07:12 23:50:45
Y Cb Cr Positioning	Co-sited
Exposure Time	1/500
F Number	5.4
Exposure Program	Program AE
ISO	400
Exif Version	0220
Date/Time Original	2015:07:11 06:27:06
Create Date	2015:07:11 06:27:06
Components Configuration	Y, Cb, Cr, -
Compressed Bits Per Pixel	2
Shutter Speed Value	1/512
Aperture Value	5.5
Brightness Value	5.6
Exposure Compensation	0
Max Aperture Value	3.5
Metering Mode	Multi-segment
Light Source	Unknown
Flash	Off, Did not fire
Focal Length	75.0 mm
Flashpix Version	0100
Color Space	sRGB
Exif Image Width	2736
Exif Image Height	3648
Interoperability Index	R98 - DCF basic file (sRGB)
Interoperability Version	0100
Focal Plane X Resolution	6129
Focal Plane Y Resolution	6129
Focal Plane Resolution Unit	cm
Sensing Method	One-chip color area

File Source	Digital Camera
Scene Type	Directly photographed
Custom Rendered	Normal
Exposure Mode	Auto
White Balance	Auto
Scene Capture Type	Standard
Sharpness	Normal
Subject Distance Range	Unknown
Padding	(Binary data 2.01 kilobytes)
Offset Schema	3970
Padding	(Binary data 2.01 kilobytes)
Compression	JPEG (old-style)
Orientation	Horizontal (normal)
X Resolution	72
Y Resolution	72
Resolution Unit	inches
Thumbnail Offset	8164
Thumbnail Length	5707
Y Cb Cr Positioning	Co-sited

**JFIF**

JFIF Version	1.01
Resolution Unit	inches
X Resolution	72
Y Resolution	72

**FlashPix**

Preview Image Width	320
Preview Image Height	240
Preview Image	(Binary data 24.05 kilobytes)

**File**

File Name	o_2111196+OCTETS+OCTETS.0
File Size	2.1 MB
File Modification Date/Time	2017:09:22 21:48:25+02:00
File Access Date/Time	2017:09:22 21:48:25+02:00
File Inode Change Date/Time	2017:09:22 21:48:25+02:00
File Permissions	rw-r--r--
File Type	JPEG
File Type Extension	.jpg
MIME Type	image/jpeg
Exif Byte Order	Big-endian (Motorola, MM)
Image Width	2736
Image Height	3648
Encoding Process	Baseline DCT, Huffman coding
Bits Per Sample	8
Color Components	3
Y Cb Cr Sub Sampling	YCbCr4:4:0 (1 2)

**Composite**

Aperture	5.4
Megapixels	10.0
Scale Factor To 35 mm Equivalent	5.8
Shutter Speed	1/500
Thumbnail Image	(Binary data 5.57 kilobytes)
Circle Of Confusion	0.005 mm
Field Of View	4.7 deg
Focal Length	75.0 mm (35 mm equivalent: 436.2 mm)
Hyperfocal Distance	201.61 m
Light Value	11.8

**XMP**

About	uuid:faf5bdd5-ba3d-11da-ad31-d33d75182f1b
Creator Tool	Microsoft Windows Photo Viewer 6.3.9600.17415

Figura 39 – Metadati EXIF

### 5.3 Le mappe geografiche:

Questa tipologia di documento, nonostante sia a pieno diritto all'interno della categoria immagini, ha delle caratteristiche da approfondire che le fanno meritare una sezione a parte. Attualmente sono caricabili come un'immagine qualunque, anche se possiedono una propria parziale autonomia. Trattasi sicuramente di file di dimensione maggiore, che presuppone anche la possibilità di un'interfaccia per la visualizzazione e lo studio diversa. È interessante fare una rapida statistica delle mappe presenti attualmente su Phaidra, per venire a conoscenza della loro dimensione media anche in rapporto a quella delle immagini comuni. In questo momento, sono presenti 185 mappe che il sistema dà la possibilità di dividere per spazio occupato: minore di 10 MB; tra i 10 e i 50 MB; tra i 50 e i 100 MB; tra i 100 e 200 MB; tra i 200 e 500 MB; tra 500 MB e 1 GB; maggiore di 1 GB. Il confronto tra le due categorie risulta dunque evidente. Le mappe geografiche occupanti più di 100 MB sono il 97,84% del totale mentre, al contrario, in questa fascia c'è l'11,08% delle immagini. È evidente come si tratti di due sottocategorie che meritano uno studio differenziato, nonostante ci sia una certa affinità tra loro. Già nella fase di acquisizione digitale, le mappe devono seguire un determinato procedimento a causa della loro maggiore grandezza. Ne risulta che questa fase è possibile tramite la digitalizzazione di una parte dell'oggetto alla volta, con conseguente necessità di comporre il tutto alla fine tramite un software professionale; oppure tramite strumenti in grado di acquisire grandi superfici senza perdere alcun dettaglio. Certo è che, più sofisticata e valida la strumentazione, più sono alti i costi del processo. Ai responsabili spetta la decisione su quale sia la via più opportuna.

Una volta terminata la fase di preparazione dell'oggetto digitale mappa, si arriva all'effettivo caricamento, il quale viste le dimensioni medie piuttosto alte, potrebbe risultare un po' più lento. Considerando che, tendenzialmente i file sono mappe piuttosto datate, è facile trovare tra loro elementi disponibili per il pubblico dominio. Al produttore di dati in ogni caso spetta l'ultima parola. Terminato il caricamento, l'oggetto diventa disponibile per l'utenza ma possiamo notare subito una prima differenza con le immagini: mentre queste ultime, vengono visualizzate in anteprima ed eventualmente possono poi essere scaricate, le mappe hanno a disposizione uno speciale lettore che può essere ingrandito anche a tutto schermo. Lo vediamo a seguire:

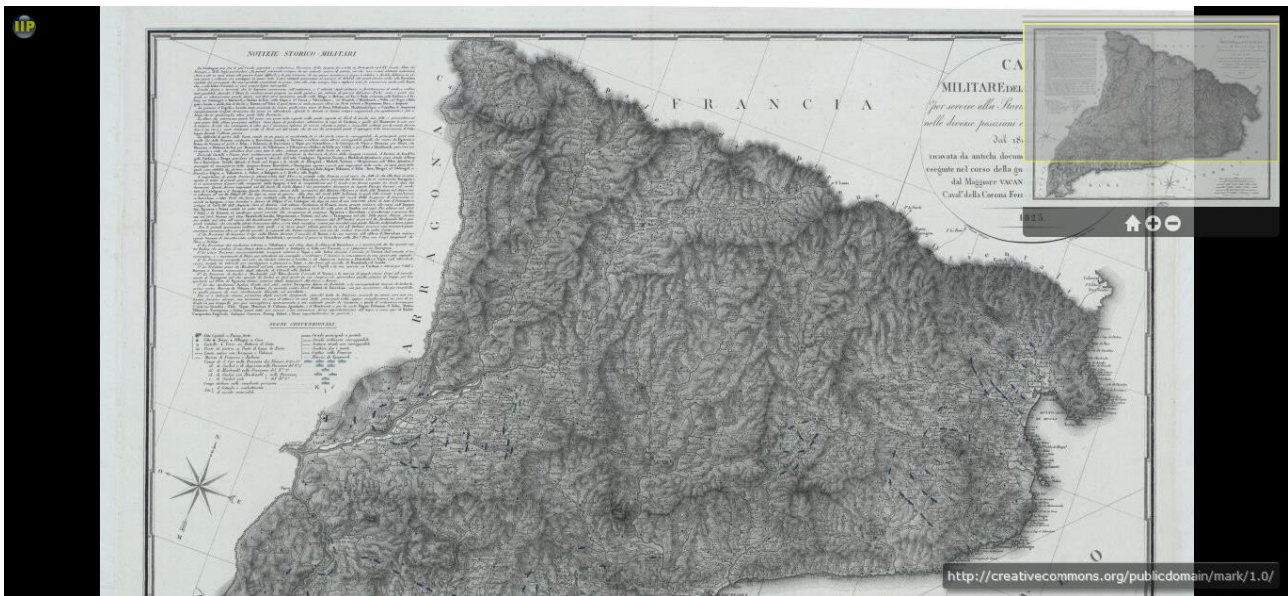


Figura 40 – Esempio di mappa geografica visualizzata sul lettore

<https://phaidra.univie.ac.at/imageserver/o:440168>

Risulta piuttosto semplice ed intuitivo muoversi all'interno della mappa, tramite i comandi in alto a destra. Aumentando l'ingrandimento si arriva a leggere comodamente i nomi delle località e i commenti scritti lateralmente, mentre per spostarsi basta tenere schiacciato il tasto sinistro del mouse e muoversi. L'oggetto è caricato sull'imageserver e il lettore è disponibile grazie ad un plug-in. È una possibilità questa al momento non disponibile presso l'Università di Padova.

Sono interessanti anche le diverse opportunità offerte nel caso un utente voglia scaricare l'oggetto: a seconda infatti dell'uso è possibile scegliere un diverso livello di definizione dello stesso. Per fare un esempio pratico, l'esempio mostrato nella figura 40 è una mappa che occupa 707.76 MB. Non necessariamente però ad una persona può servire un'immagine così ingrandibile senza perdita di qualche contenuto. Phaidra offre la possibilità di download a 72 dpi (dots per inch) e a 300 dpi, con l'obiettivo di risparmiare spazio pur con qualche costo. Chi ha bisogno può scegliere autonomamente; certo è che, se l'obiettivo è mostrare la mappa senza troppi dettagli su una pagina web, caricare tutto quel peso può essere controproducente per chi inserisce e per chi visualizza. Tornando all'esempio sopra citato, va detto che l'immagine a 72 dpi in questo caso occupa 1.9 MB, mentre a 300 dpi occupa 19.7 MB. Una differenza importante.

Link dell'oggetto

- [Visualizza](#)
- [Scarica 72 dpi](#)
- [Scarica 300 dpi](#)
- [Scarica \(707.76 MB\)](#)
- [Dublin Core](#)
- [MODS](#)
- [Visualizzatore EXIF](#)

Figura 41 – Link dell'oggetto

Un'altra opzione che spicca nella figura 41 e che non è presente nelle altre categorie di file, è la possibilità di visualizzare i metadati MODS. Come accennato nel capitolo 4.1, si tratta di uno standard creato dalla Library of Congress, il quale potrebbe essere preso in considerazione per gli sviluppi

futuri. A seguire il codice relativo all'oggetto visualizzato nella figura 40, dove si può notare la sua struttura divisa per aree descrittive.

```

- <mods:mods>
  <mods:identifier type="ac-number">AC04534351</mods:identifier>
- <mods:language>
  <mods:languageTerm authority="iso639-2b" type="code">ita</mods:languageTerm>
</mods:language>
- <mods:subject>
  <mods:cartographics>
  <mods:scale>462963</mods:scale>
</mods:cartographics>
</mods:subject>
- <mods:name authority="gnd" authorityURI="http://d-nb.info/gnd/" type="personal" valueURI="http://d-nb.info/gnd/103758100">
  <mods:namePart type="given">Giovanni B.</mods:namePart>
  <mods:namePart type="family">Bordiga</mods:namePart>
- <mods:role>
  <mods:roleTerm authority="marcrelator" type="code">ctb</mods:roleTerm>
</mods:role>
</mods:name>
- <mods:name type="personal">
  <mods:namePart type="given">Camillo</mods:namePart>
  <mods:namePart type="family">Vacani</mods:namePart>
- <mods:role>
  <mods:roleTerm authority="marcrelator" type="code">ctb</mods:roleTerm>
</mods:role>
</mods:name>
- <mods:titleInfo>
- <mods:title>
  Carta Militare Della Catalogna : per servire alla Storia delle Truppe Italiane nelle diverse posizioni e spedizioni di piu Eserciti dal 1808 al 1813
</mods:title>
- <mods:subTitle>
  ricavata da antichi documenti e da nuove ricognizioni eseguite nel corso della guerra agli Stati Generali italiano
</mods:subTitle>
</mods:titleInfo>
- <mods:note>
  Maßstab auch in graph. Form (Chilometri, Miglia Inglese, Miglia di Germania). - NE orient. - Nullmeridian: Paris. - Mit Schraffen. - Mit Erl. - Inselkt.
</mods:note>
<mods:classification authority="bkl" authorityURI="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/classification/" valueURI="http://phaidra.univie.ac.at/XML/metadata/lom/V1.0/classification/cls_10/1063851">Spanien, Portugal <Geschichte></mods:classification>
- <mods:subject>
  <mods:geographic authority="gnd" authorityURI="http://d-nb.info/gnd/" valueURI="http://d-nb.info/gnd/4029916-8">Katalonien</mods:geographic>
  <mods:topic authority="gnd" authorityURI="http://d-nb.info/gnd/" valueURI="http://d-nb.info/gnd/4324072-0">Militärkarte</mods:topic>
  <mods:topic authority="gnd">Altkarte</mods:topic>
</mods:subject>
- <mods:subject>
  <mods:topic authority="gnd" authorityURI="http://d-nb.info/gnd/" valueURI="http://d-nb.info/gnd/4182070-8">Spanischer Unabhängigkeitskrieg</mods:topic>
  <mods:temporal authority="gnd">Geschichte 1808-1813</mods:temporal>
  <mods:topic authority="gnd">Altkarte</mods:topic>
</mods:subject>
- <mods:originInfo>
  <mods:dateIssued encoding="w3cdtf" keyDate="yes">1823</mods:dateIssued>
- <mods:place>
  <mods:placeTerm type="text">[Milano]</mods:placeTerm>
</mods:place>
</mods:originInfo>
- <mods:physicalDescription>
  <mods:extent>1 Kt.</mods:extent>
  <mods:form>Kupferst., zerschn. u. auf Leinw. aufgez.</mods:form>
  <mods:extent>72 x 52 cm</mods:extent>
</mods:physicalDescription>
- <mods:recordInfo>
  <mods:recordContentSource>Universitätsbibliothek Wien</mods:recordContentSource>
  <mods:recordOrigin>Maschinell erzeugt</mods:recordOrigin>
  <mods:languageOfCataloging>
  <mods:descriptionStandard>rakwb</mods:descriptionStandard>
</mods:recordInfo>
- <mods:note type="statement of responsibility">
  Bestand der Kartensammlung der Fachbereichsbibliothek Geographie und Regionalforschung, Universität Wien
</mods:note>
<mods:accessCondition type="use and reproduction">http://creativecommons.org/publicdomain/mark/1.0/</mods:accessCondition>
</mods:mods>

```

Figura 42 – Metadati MODS

#### 5.4 I libri:

Nonostante si possa pensare ai libri come documenti strettamente legati a formati diversi, è da tenere fermamente in considerazione l'uso delle immagini. Queste, risultano così di veloce acquisizione e successiva consultazione aprendo vari orizzonti; tramite il *Book Viewer* è possibile sfogliare il libro in qualsiasi parte del mondo simulando un'esperienza reale. I produttori possono, attraverso *Phaidra Importer*, occuparsi di pubblicare il proprio libro e renderlo visualizzabile a tutti scegliendo la licenza che più gli aggrada. Il tutto diventa comodamente utilizzabile sia nel momento in cui si parla di risorse native digitali, sia per quanto riguarda la carta scansionata. Attualmente non è possibile caricare dei libri tramite l'interfaccia web, quindi l'unico metodo disponibile è l'applicazione che risulta essere semplice ed intuitiva. Ovviamente, così come per tutti gli altri tipi di risorse, è necessario preparare bene tutti i singoli oggetti che andranno poi a comporre l'entità libro. Gli unici requisiti per poter utilizzare il software sono i seguenti: possedere un account dell'università di Vienna; avere Java sul proprio computer; accettare le condizioni di utilizzo. L'eseguibile è presente sulla piattaforma come oggetto di altro formato<sup>88</sup> ed è comodamente scaricabile sul proprio pc. Al momento la versione più aggiornata è la 2.38, compatibile con Fedora 3.8.



Figura 43 – Installazione Phaidra Importer

È possibile importare immagini diverse e legarle tra loro oppure importare direttamente un pdf già pronto, il quale viene trasformato in immagini. I formati di immissione sono: TIFF, JPG, PNG e XML. Conseguentemente è presupposta la fase di composizione del libro che può essere diviso tra copertina (Abdeckung) e capitoli. Questi ultimi possono venire creati automaticamente attraverso una cartella di pagine singole in formato immagine. Altre due importazioni possibili sono i metadati in XML, ossia il file uwmetadata.xml generato da Phaidra Importer tramite una precedente esportazione oppure la struttura del libro (o della collezione) in XML. In questo caso si tratta del file phaidraimporterstructure.xml generato dall'esportazione della struttura del libro in XML. Una volta

<sup>88</sup> [https://phaidra.univie.ac.at/detail\\_object/o:471960](https://phaidra.univie.ac.at/detail_object/o:471960) [Ultimo accesso 30/09/2017]

caricato tutto, l'applicazione mostra la scomposizione strutturata in capitoli, con tutte le singole inserite. Tramite l'interfaccia dell'applicazione si può facilmente modificare l'ordine delle pagine e spostare i singoli oggetti da un capitolo all'altro. Se le immagini vengono prima sottoposte a scansione OCR, è possibile caricare insieme il file XML con il contenuto testuale delle stesse. Questo può venire poi ricercato, selezionato e copiato. La resa grafica del software si presenta così:

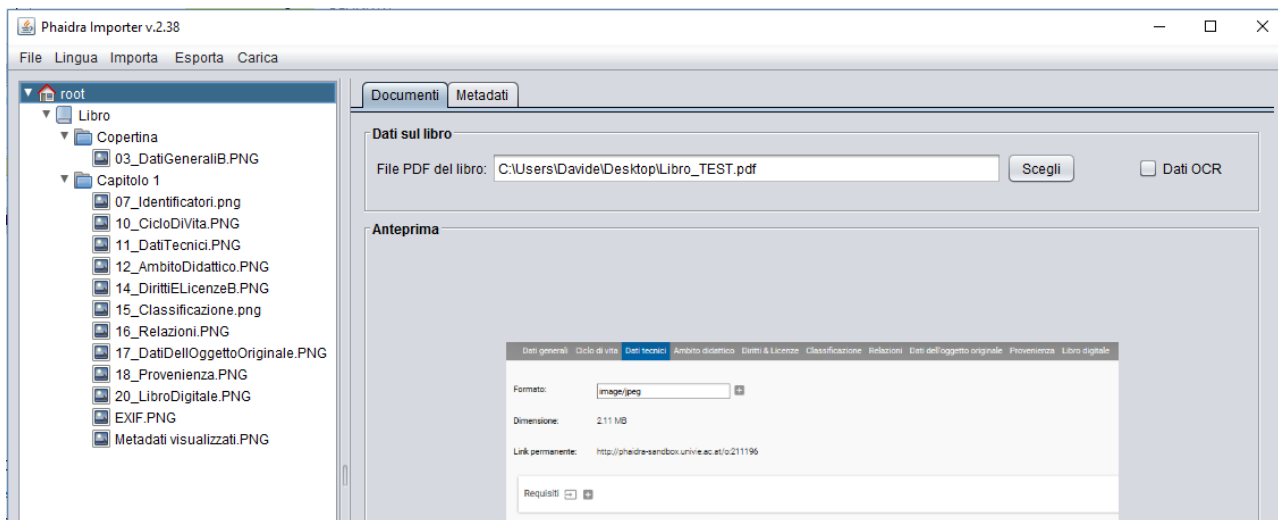


Figura 44 – Phaidra Importer: composizione del libro

Dopo aver completato la composizione del libro, i tempi sono maturi per l'aggiunta dei metadati, tra cui la lingua del testo.

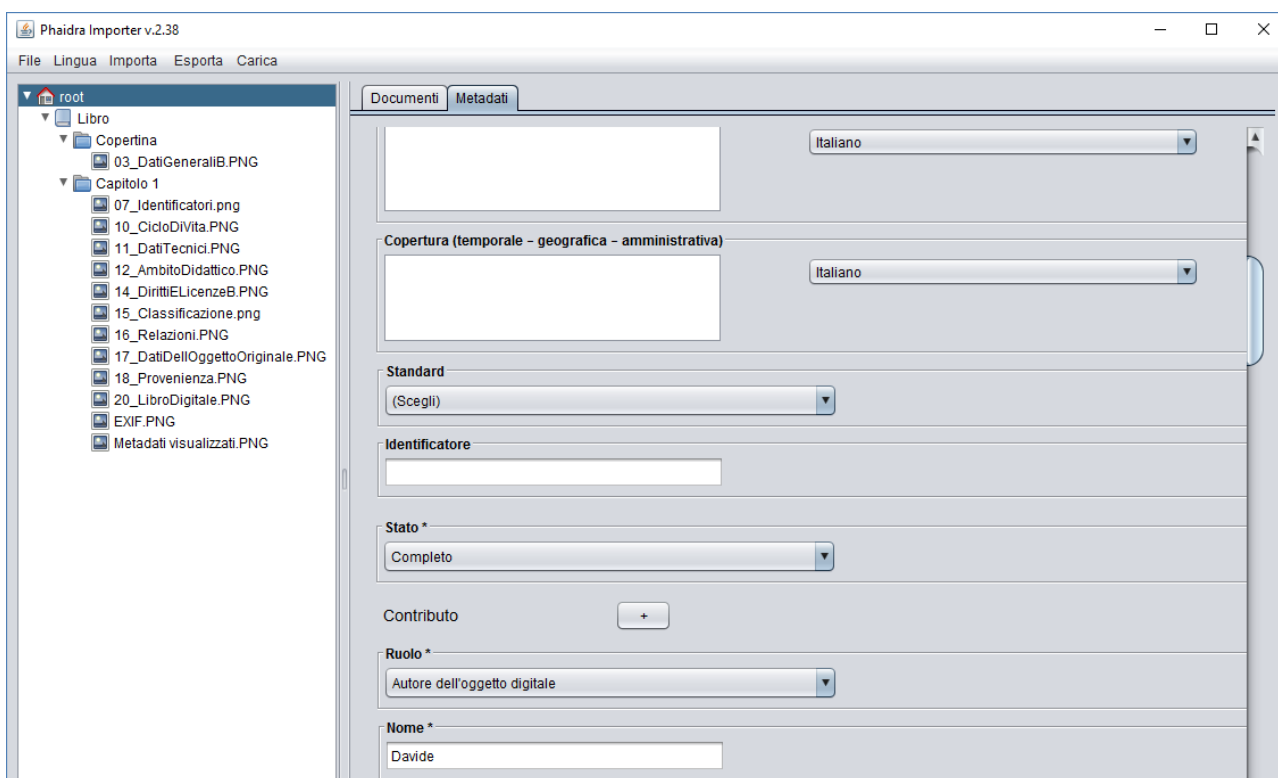


Figura 45 – Phaidra Importer: inserimento metadati



È molto interessante il fatto che ci sia un identificatore che rende univoco sia il libro intero, sia le singole pagine. Ciò può far sì che, nel caso l'oggetto debba venire citato, ci si possa riferire direttamente a tutto il complesso del testo se, eventualmente, si tratti di una bibliografia di contesto. Nel caso invece, di un singolo dato, è possibile riferirsi esclusivamente alla pagina che lo contiene. Una volta caricato il libro, può essere reso disponibile universalmente.

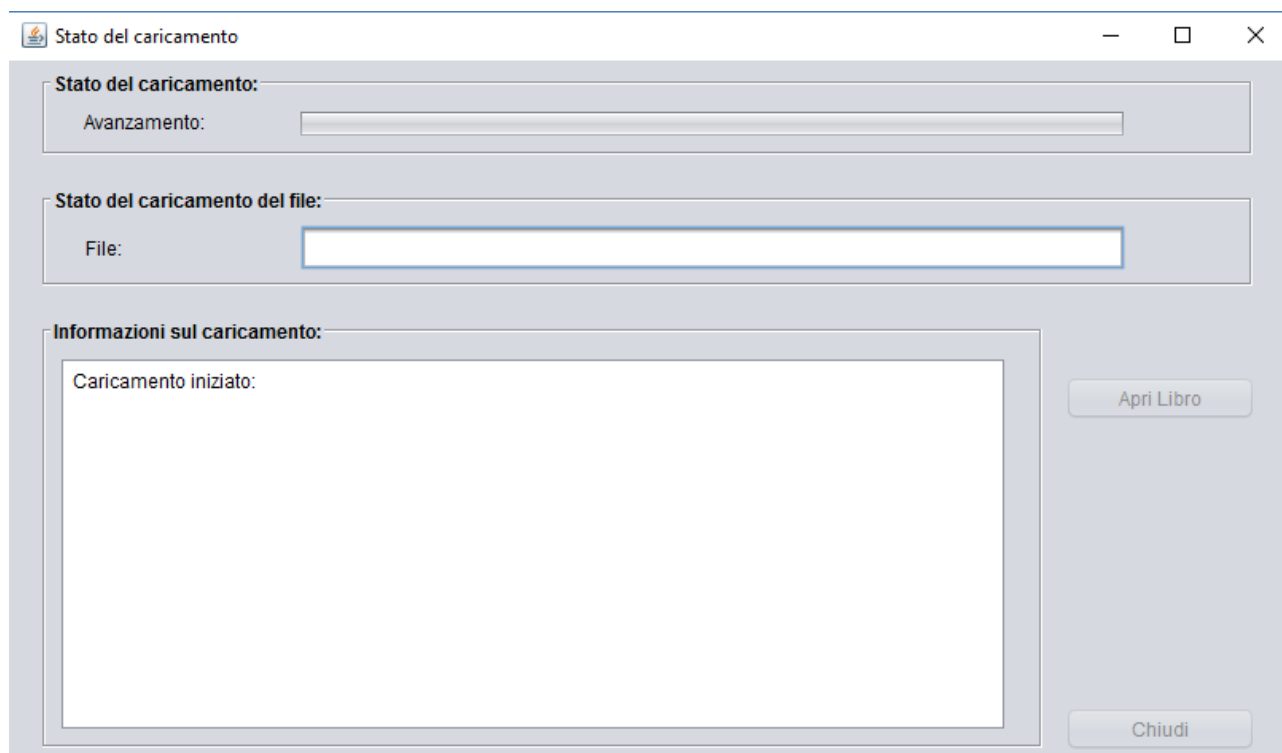


Figura 46 – Phaidra Importer: caricamento dell'oggetto

Un'altra opzione offerta da Phaidra Importer è quella dell'esportazione del libro o di alcuni suoi dati. Si può infatti creare il libro in PDF o in EPUB e salvarlo sul proprio computer o qualsiasi altro dispositivo in locale. In alternativa possono venire salvati i metadati inseriti in XML, in vista di una loro futura importazione nell'applicazione oppure può venire salvata la struttura del libro o della collezione. In questo caso parliamo sempre di un file XML che contiene l'organizzazione delle singole pagine all'interno del libro. Interessante considerare i risvolti nell'ambito della pubblicazione degli oggetti "libro": componendo infatti il proprio libro tramite questi strumenti appena descritti e caricandolo sulla piattaforma, si ha a disposizione una pubblicazione perfettamente citabile e disponibile universalmente senza passare attraverso il mondo editoriale.

### Appendice “immagini che si muovono”:

Una breve attenzione la meritano anche i filmati, i quali altro non sono che una lunghissima sequenza di immagini. Naturalmente vengono supportati da diverse tipologie di formati rispetto ai precedenti casi e, come si può facilmente intuire, hanno un impatto molto più pesante a livello di spazio occupato. Per questo esiste anche un servizio di streaming, che permette a Phaidra di reindirizzare l'utente presso un altro server che detiene il file, ma mantenendo il visitatore all'interno della propria interfaccia.

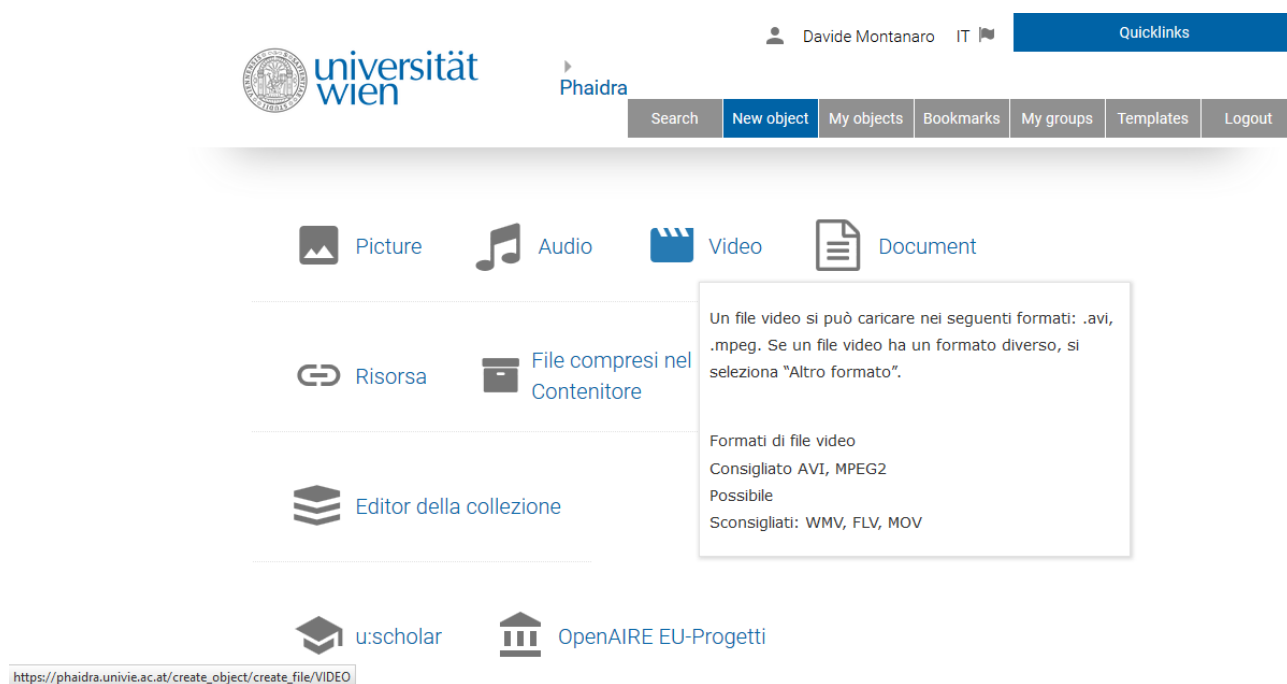


Figura 47 – Caricamento di un oggetto Video

Partiamo da una breve analisi dei formati visualizzati nella figura 47. Come consigliati troviamo MPEG2 e AVI. Il primo è uno standard introdotto nel 1994 da MPEG (Moving Pictures Experts Group), ha avuto particolare successo perché utilizzato nei DVD Video. AVI (Audio Video Interleave) è un formato contenitore realizzato da Microsoft nel 1992 come standard Windows. Il fatto di essere un formato contenitore implica che possa supportare diversi formati video e audio: Un file AVI potrebbe contenere ad esempio, il video MPEG2 e l'audio MP3, oppure il video FLV e l'audio DTS. Questi due formati consigliati hanno avuto un successo tale da essere compatibili con praticamente tutti i software di lettura video. Al contrario, i formati sconsigliati WMV e MOV sono proprietari di singole applicazioni come Windows Media Player e Quicktime. Diverso il caso di FLV (Flash Video) che è stato a lungo lo standard video per la rete, ma oggi sta cadendo in disuso a favore di HTML5. Una volta seguita la procedura di caricamento dell'oggetto digitale, lo si può visualizzare come nella figura 48.

« Back to search results



### Identifiers

<http://phaidra.univie.ac.at/o:429684>  
Handle: 11353/10.429684

#### Proprietario

ZBPH Video

#### Tipo di oggetto

VIDEO

#### Versione

Versione 1

#### Detail page views

### Titolo (deu)

Workshop "DOI für Forschungsergebnisse"; Teil 3: Angelina Kraft

Figura 48 – Visualizzazione di un oggetto Video

Phaidra dispone di un lettore video integrato tramite l'utilizzo di un plug-in, che permette di visualizzare la sequenza di immagini all'interno della pagina. Si chiama JWPlayer<sup>89</sup> ed è utilizzato in tutto il mondo per l'inserimento di video in pagine web, venendo anche utilizzato da aziende di notizie, video hosting e altre ancora. È molto versatile in quanto ed è molto leggero, dal momento che è stato pensato per essere utilizzato sul web. Si tratta di un player Open Source personalizzabile e compatibile con differenti CMS, oltre ad essere completo di un set API. Supporta sia Flash che HTML5 e offre possibilità di mostrare contenuti tramite video streaming e live webcasting consentendo una distribuzione di alta qualità. Risponde alle nuove tendenze del playback video che si sta spostando dal flash all'HTML5 così come i browser, i principali portali di video sharing e i dispositivi mobili (Tablet, Android, iPad).

Una breve menzione è meritata dal software Opencast<sup>90</sup>, utilizzato presso l'Università di Vienna per registrazioni parzialmente automatizzate e trasmissioni live di eventi didattici. Alcune sale conferenze selezionate sono state equipaggiate con le apparecchiature necessarie e l'elaborazione e la pubblicazione avviene tramite Opencast. Se necessario, verranno sviluppati altri moduli software

<sup>89</sup> <https://www.jwplayer.com/about-jwplayer/> [Ultimo accesso 30/09/2017]

<sup>90</sup> <http://www.opencast.org/> [Ultimo accesso 30/09/2017]

oltre a questo nello streaming AG della collaborazione accademica Moodle (AMC)<sup>91</sup>. L'idea di base è quella di raggruppare le attività di diverse università sullo sviluppo e il funzionamento di una soluzione di streaming basata su Opencast. L'Università TU di Vienna e l'Università di Vienna stanno attualmente collaborando su questo punto. Gli scenari su cui si sta lavorando sono la registrazione e la pubblicazione di eventi didattici e scientifici, la pubblicazione di contenuti AV prodotti in automatico, la pubblicazione di contenuti AV esterni e le pubbliche relazioni, dato che i video sono sempre più prodotti per questa motivazione e dovrebbero essere forniti attraverso l'infrastruttura di streaming universitaria. Il software Opencast soddisfa molti dei requisiti posti in un sistema di archiviazione, tuttavia è ancora visto principalmente come un sistema produttivo con cui i video vengono creati, modificati e pubblicati per un tempo determinato. L'utilizzo di questi sistemi sembra utile per l'archiviazione permanente a causa del forte utilizzo atteso e dell'esistenza di adeguati sistemi di archiviazione a lungo termine come Phaidra.

---

<sup>91</sup> BERGER C. et al., *Digitale Archivierung und Bereitstellung von AV-Medien – Erfahrungen und Praxisbeispiele aus dem tertiären Bildungssektor*, BoD – Books on Demand, 2017.

## Conclusioni

Partendo dall'analisi svolta, mi sembra doveroso riportare quanto studiato e appreso anche nell'ambito della mia istituzione di provenienza. È importante tenere sempre in considerazione il contesto entro il quale si muove lo sviluppo della piattaforma, ossia la commissione europea con l'European Open Science Cloud e i suoi obiettivi principali che sono la conservazione a lungo termine dei documenti digitali e la condivisione dei dati di ricerca; senza dimenticare tutto ciò che ruota attorno all'idea di Open Data. Viene quindi spontaneo, partendo da questi presupposti, domandarsi se una piattaforma come Phaidra sia utile ed utilizzata all'interno dell'università Ca' Foscari. Va subito detto, che essa non viene usata per caricare i dati di ricerca, bensì per caricare altri materiali provenienti da biblioteche e archivi, o riassunti sotto la guida dello SBA – Sistema Bibliotecario di Ateneo. Un'altra questione poi da porre in primo piano, riguarda la conoscenza della piattaforma da parte dei potenziali produttori di dati dell'Ateneo e se questi hanno inoltre una sufficiente sensibilità riguardo alla gestione dei dati. Può essere utile ai fini dell'analisi, utilizzare come metro di paragone alcuni dati emersi da un sondaggio tra i ricercatori all'interno del progetto E-Infrastructures Austria. In ogni caso per riuscire ad avere una panoramica completa sull'ecosistema veneziano, potrebbe essere interessante uno studio approfondito a partire dai casi d'uso attuali, seguendo in parte il processo descritto nel capitolo 3 parlando dell'università di Vienna. Va considerato subito che, i casi di Venezia e Vienna, non possono essere confrontati a livello paritario su alcuni temi date le importanti differenze fra i due Atenei. Pensiamo ad esempio alla tradizione storica, essendo l'università austriaca fondata nel 1365, rispetto a quella di Venezia nel 1868; o al numero di iscritti, 93855<sup>92</sup> del primo caso contro i 20039<sup>93</sup> del secondo. Ciò comporta differenze piuttosto considerevoli relativamente al patrimonio storico che può essere caricato e alla quantità di soggetti che hanno la possibilità di creare risorse digitali. Siamo dunque di fronte ad un'analisi dei due obiettivi chiave della piattaforma che vediamo in sequenza:

- Per quanto riguarda la *conservazione dei dati digitali* emergono in particolare questioni relative al deposito dei Digital Objects presso un altro Ateneo e relative all'effettivo uso e di conseguenza conoscenza delle potenzialità della piattaforma da parte dei membri dell'università. Inoltre altri interrogativi da non sottovalutare sono i seguenti: esistono i requisiti minimi per considerare l'utilizzo indipendente di una piattaforma come Phaidra? Ci sono di conseguenza, gli oggetti e le potenziali risorse per sfruttare al meglio questa possibilità?

---

<sup>92</sup> Nel semestre invernale 2016/2017. Fonte: <https://www.univie.ac.at/ueber-uns/auf-einen-blick/zahlen-daten-broschueren/> [Ultimo accesso 04/10/2017]

<sup>93</sup> Nell'anno accademico 2015/2016. Fonte: [anagrafe.miur.it](http://anagrafe.miur.it) [Ultimo accesso 04/10/2017]

- Nell'ambito della *condivisione dei dati di ricerca* esiste invece un problema di fondo. In Italia esiste già una piattaforma costruita a questo scopo da CINECA che si chiama IRIS (Institutional Research Information System)<sup>94</sup> e che nel caso di Ca' Foscari viene utilizzata dall'archivio istituzionale ARCA (Archivio Ricerca Ca' Foscari)<sup>95</sup>. In questo caso Phaidra potrebbe venire poco considerato, in quanto a livello istituzionale il repository usato per questo scopo è un altro. Non è questo il luogo però per un confronto tra le due piattaforme.

Proviamo ad approfondire tramite una riflessione che dovrà entrare nel merito dei tre concetti cardine indicati dalla commissione europea e sviluppati nel corso dell'elaborato come *governance*, *servizi* e *dati*. A tal proposito può far comodo tornare a vedere brevemente le figure 4 e 5 che illustrano la prospettiva dell'ecosistema digitale.

Entrando nel merito delle politiche (*governance*), non si può fare a meno di partire da una valutazione dei costi che la piattaforma comporta e di come il denaro stanziato venga utilizzato per lo sviluppo del sistema. Con questo si intende il mantenimento delle infrastrutture, la formazione RDM, la promozione del servizio, ecc. Tutto ciò è strettamente legato, in quanto si può avere a disposizione la piattaforma perfetta, ma se i soggetti produttori non la sanno usare o non sanno neanche che essa esista risulta poco utile. Nel caso dei ricercatori austriaci ad esempio, un settimo di loro dichiara di utilizzare data archives/repositories, mentre i due terzi predilige le soluzioni cloud o le applicazioni sul sito web personale<sup>96</sup>. È bene interrogarsi dunque sulla visibilità delle campagne in favore dell'Open Access e su l'impatto che queste hanno sui ricercatori.

Nel campo dei *servizi*, in primo piano va considerato il recente passaggio da repositories a ecosistemi digitali, i quali hanno aperto nuove possibilità. Queste però premettono anche l'ampliamento delle precedenti competenze, entrando nel merito di una terminologia comune e assodata, di un'assistenza legale, di supporto all'area di ricerca. Si parla quindi di programmi di formazione, i quali dovrebbero intervenire su tutti gli ambiti disciplinari in modo da coinvolgere tutti i dipartimenti nella gestione senza che si intralcino tra loro. Sempre in Austria, molti soggetti si aspettano maggiore supporto da parte della propria istituzione in ambito gestionale, legale, etico così come un servizio di help-desk centrale.

Per quanto riguarda il livello dei *dati*, ci troviamo in un periodo storico nel quale la maggior parte del volume di ricerca risulta essere direttamente realizzato in formato digitale. Siamo ancora inoltre in una fase di studio per quanto riguarda le tecnologie ideali per l'archiviazione nonostante i notevoli passi avanti. Ci sono varie attività avviate a livello europeo, come quelle relative ai principi FAIR e

<sup>94</sup> <https://www.cineca.it/it/content/IRIS> [Ultimo accesso 04/10/2017]

<sup>95</sup> <http://www.unive.it/pag/10391> [Ultimo accesso 04/10/2017]

<sup>96</sup> BUDRONI P., *Riflessioni sugli esiti delle iniziative BISA e il progetto E-Infrastructures Austria*, 7 agosto 2017. Disponibile a: <https://phaidra.univie.ac.at/view/o:556423> [Ultimo accesso 04/10/2017]

all'uso del Data Management Plan, ma sembrano ancora lontane dal venire utilizzate quotidianamente dai ricercatori. Ciò incide negativamente sulla ricerca e riutilizzo dei dati. Qui entra nuovamente in gioco la formazione dei soggetti produttori i quali, sentendosi più sicuri nell'uso di strumenti nuovi, possono dare il loro contributo nel migliore dei modi. Sarebbe sicuramente curioso avere una panoramica statistica dell'interesse mostrato dalla comunità accademica veneziana nei confronti della piattaforma. Sempre relativamente ai dati, è mia intenzione esporre un paio di riflessioni relative al mondo delle immagini, trattato nel capitolo 5. Nonostante tutti gli aspetti positivi già descritti, ho trovato la ricerca degli oggetti un po' limitante, in quanto ci sono a disposizione esclusivamente i tag inseriti dall'utente per titolo e descrizione. Esistono a tal proposito le classificazioni preimpostate, ma non essendo obbligatorie vengono spesso ignorate dal produttore di dati e non sono sempre di facile intuizione. Inoltre volendo fare una ricerca a partire dall'autore, non risulta possibile cercare tramite nome o cognome ma solo per la sigla assegnata. Per quanto riguarda il formato di archiviazione viene suggerito TIFF, il quale mi trova concorde sulla scelta. Certo è che, nonostante sia probabilmente il migliore per le finalità della piattaforma, ha comunque qualche difetto. Dunque non risulta esserci, al momento, un formato ideale per la conservazione a lungo termine delle immagini. Riguardo ai metadati, sarà interessante vedere lo sviluppo che prenderanno nel prossimo futuro. Dopo un lungo utilizzo dello standard LOM, è stato provato MODS per le mappe geografiche e sembra un buon compromesso tra gli aspetti relativi all'interoperabilità e alle necessità descrittive. Per concludere, vorrei porre nuovamente l'attenzione sulle politiche europee sulla scienza aperta che ogni giorno evolvono. Nonostante ogni Ateneo debba giustamente seguire le proprie vocazioni scientifiche sviluppando politiche interne adeguate, è necessario che queste volgano sempre uno sguardo alle linee guida UE arrivando a creare un quarto livello sovra istituzionale chiamato anche "Virtual Layer for Management of the Complete Life Cycle of Scientific Data"<sup>97</sup>. In caso contrario c'è un forte rischio di isolamento che poco gioverebbe alla comunità intellettuale locale e internazionale, mantenendo gli istituti di ricerca isolati.

---

<sup>97</sup> WEIGEL T., WITTENBURG P., *Recommendations for Implementing a Virtual Layer for Management of the Complete Life Cycle of Scientific Data*, 2017. Disponibile a:  
<<https://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf>> [Ultimo accesso 12/09/2017]





## Bibliografia

ALIPRANDI S., *Creative commons: manuale operativo. Guida all'uso delle licenze e degli altri strumenti cc*, Ledizioni/Copyleft-Italia.it, 2008.

AMSTERDAM CALL FOR ACTION ON OPEN SCIENCE, basato sulla conferenza *Open Science – From Vision to Action*, tenutasi ad Amsterdam il 4 - 5 aprile 2016. Disponibile a: <<http://www.openaccess.nl/sites/www.openaccess.nl/files/documenten/amsterdam-call-for-action-on-open-science.pdf>> [Ultimo accesso 12/09/2017]

BERGER C. et al., *Digitale Archivierung und Bereitstellung von AV-Medien – Erfahrungen und Praxisbeispiele aus dem tertiären Bildungssektor*, BoD – Books on Demand, 2017.

BUDRONI P., BLUMESBERGER S., PAUSZ R., PREZA J. L., SÁNCHEZ SOLÍS B., *Data Management Plan for projects at the University of Vienna. Recommended Repository: Phaidra (V2.0/customised for UNIVIE) (English): Version 2.0 / customised for UNIVIE*, 2015. Disponibile a: <<http://phaidra.univie.ac.at/o:407974>> [Ultimo accesso 12/09/2017]

BUDRONI P., PREZA J. L., PAUSZ R., SÁNCHEZ B., BLUMESBERGER S., *Data Management Plan - Eine Anleitung zur Erstellung von Data Management Plänen. Projekt e-Infrastructures Austria (Deutsch): Eine Anleitung zur dauerhaften Sicherung von digitalen Beständen*, 2015. Disponibile a: <<http://phaidra.univie.ac.at/o:367863>> [Ultimo accesso 12/09/2017]

EUROPEAN COMMISSION, *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020*, 2016. Disponibile a: <[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)> [Ultimo accesso 12/09/2017]

EUROPEAN COMMISSION, *Realising the European Open Science Cloud. First report and recommendations of the Commission High Level Expert Group on the European Open Science Cloud*, 2016. Disponibile a: <<https://ec.europa.eu/digital-single-market/en/news/first-report-high-level-expert-group-european-open-science-cloud>> [Ultimo accesso 12/09/2017]

EUROPEAN COMMISSION, *Europe's future: Open Innovation, Open Science, Open to the World - Reflections of the Research, Innovation and Science Policy Experts (RISE) High Level Group*, 2017. Disponibile a: <<https://ec.europa.eu/research/openvision/pdf/publications/ki0217113enn.pdf>> [Ultimo accesso 12/09/2017]

FERUS A., *Wissenschaftskommunikation, e-Infrastrukturen und wissenschaftliche Bibliotheken*, 2016. Disponibile a: <<http://phaidra.univie.ac.at/o:441295>> [Ultimo accesso 12/09/2017]

GANGULY R., BUDRONI P., SÁNCHEZ SOLÍS B., "Living Digital Ecosystems for Data Preservation: An Austrian Use Case Towards the European Open Science Cloud." In *Expanding Perspectives on Open Science: Communities, Cultures and Diversity in Concepts and Practices*, ELPUB. Limassol, Cyprus, 2017.

LEARN PROJECT, *20 RDM Best-Practice Recommendations*, 2017. Disponibile a: <<http://learn-rdm.eu/wp-content/uploads/20-RDM-Policy-Recommendations.pdf>> [Ultimo accesso 12/09/2017]

LEARN PROJECT, *LEARN Toolkit of Best Practice for Research Data Management*, 2017. Disponibile a: <<https://doi.org/10.14324/000.learn.00>> [Ultimo accesso 12/09/2017]

LERU RESEARCH DATA WORKING GROUP, *LERU Roadmap for Research Data*, 2013.

Disponibile a:

<[http://www.leru.org/files/publications/AP14\\_LERU\\_Roadmap\\_for\\_Research\\_data\\_final.pdf](http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf)>  
[Ultimo accesso 12/09/2017]

MIKSA T., SCHRAUF C., SÁNCHEZ SOLÍS B., *Data Management Plans (DMP)*, 2016.

Disponibile a: <<http://phaidra.univie.ac.at/o:441308>> [Ultimo accesso 12/09/2017]

ROVIRA C., *Technology Watch and Competitive intelligence for SEM-SEO*, "Hipertext.net", num.

6, 2008. Disponibile a: <<https://www.upf.edu/hipertextnet/en/numero-6/vigilancia-tecnologica.html>> [Ultimo accesso 25/09/2017] e <

<http://www.dpconline.org/knowledge-base/tech-watch-reports>> [Ultimo accesso 25/09/2017]

THE CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEM PRACTICES, *Reference Model for an Open Archival Information System (OAIS)*, 2012. Magenta Book. Disponibile a:

<<https://public.ccsds.org/pubs/650x0m2.pdf>> [Ultimo accesso 12/09/2017]

WEIGEL T., WITTENBURG P., *Recommendations for Implementing a Virtual Layer for*

*Management of the Complete Life Cycle of Scientific Data*, 2017. Disponibile a: <<https://www.rd-alliance.org/sites/default/files/recommendation-jan-2017-v8.pdf>> [Ultimo accesso 12/09/2017]

WILKINSON M. D. et al., *The FAIR Guiding Principles for scientific data management and stewardship*, 2016. Disponibile tramite DOI: 10.1038/sdata.2016.18