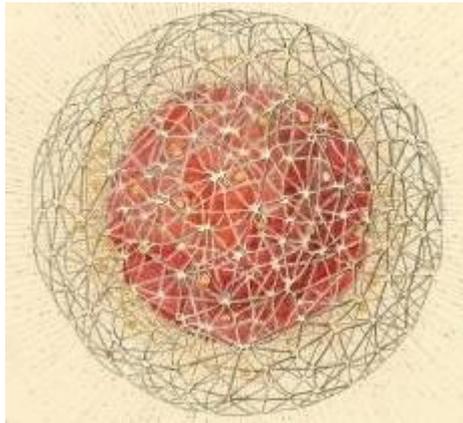


**Associazione Nazionale Archivistica Italiana – Direzione Generale per gli Archivi**



**Il Mondo degli Archivi STUDI – A. III, aprile 2015**

**Sapienza Digital Library:  
strategie e *policies* di digitalizzazione e conservazione**

*di Alessandro Cuomo*

Il saggio è pubblicato al seguente indirizzo:  
<http://www.ilmondodegliarchivi.org/index.php/studi/item/541-sapienza-digital-library-strategie-e-policies-di-digitalizzazione-e-conservazione>

## **Abstract**

Il presente saggio è una sintesi della tesi magistrale “Sapienza Digital Library: strategie e *policies* di digitalizzazione e conservazione”, frutto diretto dell’esperienza lavorativa maturata presso il Laboratorio di digitalizzazione e metadattazione del Centro DigiLab, struttura dell’Università di Roma “La Sapienza”. L’obiettivo della ricerca ha riguardato la descrizione, l’analisi e la proposta di possibili o necessarie implementazioni alle strategie e alle *policies* attuate in merito ai processi di digitalizzazione e conservazione digitale nell’ambito del progetto Sapienza Digital Library (SDL).

*The article is an extract of the M.A. thesis “Sapienza Digital Library: strategie e policies di digitalizzazione e conservazione” (“Sapienza Digital Library: digitization and preservation strategies and policies”). This study is the result of the work experience I had at DigiLab’s digitization laboratory for the SDL project. The research aimed to define digitization workflows and digital preservation policies at “La Sapienza” University of Rome.*

## **Introduzione**

L’obiettivo del lavoro da cui è tratto questo saggio è descrivere e analizzare le strategie e le *policies* attuate in merito ai processi di digitalizzazione e conservazione digitale nell’ambito del progetto *Sapienza Digital Library (SDL)*, e, laddove possibile, suggerire implementazioni necessarie affinché l’insieme delle operazioni effettuate risulti pienamente conforme agli standard di settore, garantendone in tal modo l’esito positivo. Il cuore dello studio è quindi duplice: da un lato, sulla scorta di standard elaborati negli ultimi anni e attraverso modelli – da questi derivati – appositamente concepiti per il caso SDL, ci si propone di indagare il complesso delle attività costituenti il macro-processo di digitalizzazione, insieme ai fattori necessari e immediatamente precedenti alla fase di acquisizione digitale vera e propria; dall’altro, grazie alla relativa documentazione progettuale e facendo sempre ricorso ai paradigmi scientifici di settore, si espone un’analisi complessiva del modello, logico e tecnologico, che consente una conservazione digitale scientificamente corretta. La premessa a uno studio del genere è costituita da una panoramica del progetto SDL così come concepito e gestito dalle istituzioni e dai centri coinvolti.

## **Il progetto SDL**

Il progetto Sapienza Digital Library nasce dall’idea di raccogliere in un unico sistema digitale di conservazione, gestione e *dissemination* la produzione intellettuale della Sapienza passata e futura, già nata digitale, cioè *born digital*, o tradotta successivamente in formato digitale attraverso processi di digitalizzazione. Il progetto

SDL ha avuto come obiettivo iniziale quello di integrare, organizzare e catalogare diversi tipi di materiali, al fine di renderli disponibili all'intera comunità accademica: nell'Università di Roma "La Sapienza", infatti, sono attualmente presenti, o potenzialmente realizzabili, numerosi e rilevanti archivi e collezioni digitali relativi al patrimonio culturale, scientifico e ambientale. Grazie a questo progetto è possibile assicurare la transizione da un contesto caratterizzato da forte frammentazione e dispersione a una situazione in grado di consentire l'accessibilità all'intero patrimonio disponibile, tramite il recupero delle collezioni esistenti e la creazione continua di nuove collezioni e oggetti, garantendo un supporto concreto a quelle istituzioni, come dipartimenti e biblioteche universitarie, che incontrano problemi analoghi a quelli della maggior parte degli enti locali e dei musei.

Le motivazioni alla base di un lavoro simile sono dunque immediatamente comprensibili: il complesso di strutture, centri e istituzioni di cui è composta l'Università "La Sapienza" possiede, relativamente sia alle aree umanistiche che a quelle scientifiche, un patrimonio quantitativamente e qualitativamente di assoluto rilievo. Il passaggio dal mondo degli atomi al cosiddetto mondo dei bit, insieme a interrogativi spinosi e problematici, ha dischiuso enormi possibilità di valorizzazione e riutilizzo costante dei patrimoni digitali, universalmente riconosciuti come *asset* strategici dalle maggiori università a livello internazionale, e per questa ragione oggetto di specifiche azioni e investimenti.

Il progetto SDL nasce sin dall'inizio con una visione consortile: il portale SDL<sup>1</sup> è il risultato di un progetto di ricerca e sviluppo condotto dalla Sapienza, nello specifico dal Centro DigiLab<sup>2</sup>, in collaborazione con il Sistema Bibliotecario della Sapienza (SBS) e con InfoSapienza, in partnership con il CINECA<sup>3</sup>. Il portale costituisce l'infrastruttura digitale per la comunicazione, in un contesto possibilmente non solo italiano ma anche europeo e internazionale, dei patrimoni culturali, scientifici e ambientali appartenenti alla Sapienza e a enti collegati (eventualmente, anche a donatori). In tutte le fasi di progettazione di questa infrastruttura sono state non solo preservate, ma anzi implementate, le potenzialità di estensione dell'intero sistema ad altre università e enti.

Inoltre, il progetto SDL ambisce ad assicurare la conservazione a lungo termine di un patrimonio caratterizzato non soltanto da contenuti immediatamente prodotti o posseduti dalle istituzioni di Sapienza, relativi al loro lavoro o al lavoro di altri soggetti concernente l'*heritage*, ma anche contraddistinto, in prospettiva, da quei contenuti riferiti alle attività di ricerca e formazione (pubblicazioni, tesi, convegni, videoregistrazioni, ecc.) e dagli *user generated content*, i quali, soprattutto nel caso di

---

<sup>1</sup> Raggiungibile all'indirizzo <http://sapienzadigitallibrary.uniroma1.it> (aprile 2015).

<sup>2</sup> DigiLab, centro interdipartimentale di ricerca e servizi dell'Università "La Sapienza" è raggiungibile all'indirizzo <http://digilab.uniroma1.it> (aprile 2015).

<sup>3</sup> Il consorzio interuniversitario senza scopo di lucro al servizio del sistema accademico italiano è raggiungibile all'indirizzo <http://www.cineca.it> (aprile 2015).

una grande università, costituiscono in prevalenza materiali prodotti nell'ambito delle attività di ricerca, a scopi divulgativi oppure come prodotto diretto delle attività degli studenti.

### **Digitalizzazione in SDL: modelli, *policies* e strategie**

L'analisi, la descrizione e l'ottimizzazione delle strategie e *policies* attuate in merito ai processi di digitalizzazione hanno trovato le proprie fondamenta in standard, *case studies* e linee guida internazionali, tra cui hanno costituito fonte privilegiata le *guidelines* internazionali elaborate in proposito: il *Preservation Management of Digital Materials: the Handbook* elaborato dalla DPC (*Digital Preservation Coalition*)<sup>4</sup>, il *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale* redatto nell'ambito del progetto europeo Minerva<sup>5</sup>, la *NINCH Guide to good practice in the Digital Representation and Management of Cultural Heritage Materials* a firma HATII (*Humanities Advanced Technology and Information Institute*) e NINCH (*National Initiative for a Networked Cultural Heritage*)<sup>6</sup>, la *Technical Guidelines for Digitizing Archival Materials for Electronic Access* realizzata dagli statunitensi NARA (*National Archives and Record Administration*)<sup>7</sup>, le *Guidelines for Digitization Projects* dell'IFLA (*International Federation of Library Associations*) in collaborazione con l'ICA (*International Council on Archives*)<sup>8</sup> e, soprattutto, tutta la documentazione prodotta nell'ambito dell'iniziativa americana FADGI (*Federal Agencies Digitization Guidelines Initiative*)<sup>9</sup>.

Proprio il *FADGI Still Image Working Group* ha elaborato un'utilissima struttura organizzativa, comprensiva di tutte le attività e la fasi di cui un teorico – e quindi, per forza di cose, generico – progetto di digitalizzazione dovrebbe essere costituito<sup>10</sup>. A partire da questo esempio è stato realizzato un modello organizzativo più snello e semplice – consultabile all'allegato 1 “*Overview and sequencing of activities for SDL digitization project*” – maggiormente adatto a riassumere non solo il complesso delle attività costitutive, ma anche gli ambiti professionali e disciplinari interessati dal

---

<sup>4</sup> DPC, *Preservation Management of Digital Materials: the Handbook*, <http://www.dpconline.org/advice/preservationhandbook>, ottobre 2008 (aprile 2015).

<sup>5</sup> MINERVA WORKING GROUP 6, *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale*, [http://www.minervaeurope.org/structure/workinggroups/goodpract/document/buonepratiche1\\_3.pdf](http://www.minervaeurope.org/structure/workinggroups/goodpract/document/buonepratiche1_3.pdf), marzo 2004 (aprile 2015).

<sup>6</sup> NINCH, HATII, *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, <http://goo.gl/pLA5LS>, 2002, (aprile 2015).

<sup>7</sup> NARA, *Technical Guidelines for Digitizing Archival Materials for Electronic Access*, <http://www.archives.gov/preservation/technical/guidelines.html>, giugno 2004 (aprile 2015).

<sup>8</sup> IFLA, ICA, *Guidelines for Digitization Projects*, <http://www.ifla.org/publications/guidelines-for-digitization-projects-for-collections-and-holdings-in-the-public-domain>, marzo 2002 (aprile 2015).

<sup>9</sup> Storia e obiettivi dell'iniziativa, oltre alla documentazione prodotta negli anni, sono disponibili al sito: <http://www.digitizationguidelines.gov> (aprile 2015).

<sup>10</sup> Consultabile all'indirizzo: <http://www.digitizationguidelines.gov/guidelines/DigActivities-FADGI-v1-20091104.pdf> (aprile 2015).

progetto SDL: sono individuate quattro fasi principali di svolgimento (*Project Planning, Pre-Digitization, Digitization e Post-Digitization*), ognuna analizzata nello specifico delle attività di cui si compone, mentre la suddivisione tipologica in *AV Resources*, cioè risorse audiovisive, e *Still Images*, ovverosia risorse testuali, mappe, stampe fotografiche, negativi, ecc., mutuata direttamente dal lavoro dell'iniziativa FADGI, è apparsa sin da subito molto utile nella sua concretezza<sup>11</sup>.

In particolar modo il modello e lo studio qui esposti, a differenza dello schema FADGI, insistono maggiormente sull'analisi dei fattori necessari e immediatamente precedenti alla fase di acquisizione digitale vera e propria, indagando in che misura possano definirsi propedeutici ad essa e come possano essere implementati sulla scorta delle *recommendations* internazionali e degli esempi forniti dai più importanti *case studies* in proposito. Per ogni operazione, ambito o settore specifico individuato nello schema viene quindi operato un confronto fra un ipotetico "modello ideale" e la realtà del progetto SDL. Conseguentemente, emergono non soltanto tutte le peculiarità caratterizzanti il progetto in questione, ma anche le sue criticità e i punti di forza: mentre si è cercato di elaborare e proporre soluzioni alle prime, ricorrendo sempre ai paradigmi di volta in volta citati ed analizzati, i secondi sono stati opportunamente sottolineati, con lo scopo e l'ambizione di proporli come esempi e modelli di riferimento.

Non è possibile in questa sede presentare in forma discorsiva gli elementi emersi ed evidenziati nell'ambito di ogni specifica sezione all'interno della tesi. Tuttavia, il modello appena descritto di base a questo studio, con la sua suddivisione in fasi e attività, si presta agevolmente a una schematizzazione delle criticità e dei punti di forza emersi dall'analisi delle varie operazioni e ambiti di cui consiste la digitalizzazione nel progetto SDL. È dunque possibile consultare la tavola "Criticità e punti di forza del processo di digitalizzazione in SDL" (allegato 2) per una visione d'insieme, riassuntiva, sia degli aspetti caratterizzanti in senso positivo il progetto, sia delle problematiche di varia natura che lo affliggono, nonché delle soluzioni proposte.

Volendo qui sintetizzarne le caratteristiche più rilevanti, è necessario sottolineare, tra le criticità, almeno la gestione delle risorse umane e dell'ambiente di digitalizzazione: a proposito della prima, vanno rilevate la mancanza di un vero e proprio *project manager*, cioè una figura a cui assegnare, in via prioritaria rispetto ad altre sue possibili mansioni, le responsabilità di gestione quotidiana dell'insieme di operazioni previste, e l'impossibilità di trattenere le risorse umane impiegate nel Laboratorio di digitalizzazione per un periodo di tempo necessario a stabilire un certa continuità nel lavoro svolto, mentre sarebbe utile individuare almeno un piccolo nucleo di risorse da poter confermare stabilmente. Relativamente alla gestione del *workspace* di

---

<sup>11</sup> La categorizzazione è stata proposta dal *FADGI Still Image Working Group* nel suo atto costitutivo: FADGI STILL IMAGE WORKING GROUP, *Still Image Digitization Working Group Charter*, <http://www.digitizationguidelines.gov/guidelines/StillImageCharter.pdf>, luglio 2008 (aprile 2015).

digitalizzazione, invece, la mancata progettazione di un ambiente *ad hoc*, così come indicato dalle raccomandazioni internazionali esaminate, è riconducibile non a responsabilità di gestione ma semplicemente a motivazioni economiche: una problematica che, tra l'altro, finisce per precludere al Laboratorio possibili sviluppi tecnologici, posti al di là della soglia di spesa sostenibile.

Nonostante queste e altre lacune, la maggioranza degli aspetti relativi alla digitalizzazione in SDL risulta conforme agli standard e alle linee guida di riferimento, in particolar modo per quanto riguarda la scelta e l'utilizzo di specifici *file formats*, la gestione dei diritti legali e, soprattutto, la gestione dei metadati: la scelta di riunire i diversi standard di settore, integrando i modelli di metadati e i linguaggi strutturati in un'unica scheda descrittiva, flessibile e compatibile con la grande varietà di risorse catalogabili, ha consentito da un lato di rispettare le esigenze di ogni dominio scientifico, mantenendone le caratteristiche peculiari e tutelando la multidimensionalità delle collezioni e dei fondi, dall'altro di preservare l'interoperabilità dei patrimoni digitali, garantendo la possibilità di esportazione e *mapping* in ulteriori aggregatori.

### **Conservazione digitale: il modello SDL – CINECA**

Se è vero che una conservazione digitale scientificamente corretta può essere definita come «la stabilizzazione fisica e tecnologica e la protezione del contenuto intellettuale dei documenti di archivio, tesa ad assicurarne una catena di conservazione continua, duratura, stabile, costante ed ininterrotta, senza un termine prestabilito»<sup>12</sup>, assume di conseguenza un'importanza fondamentale l'architettura, logica e materiale, che permette una conservazione a pieno titolo "archivistica". Non potendo in questa sede dar conto dell'intero processo di *ingestion* previsto, comunque risultante nel complesso conforme alle indicazioni fornite dallo standard *OAIS*, soprattutto in merito alla costruzione dei pacchetti *SIP*<sup>13</sup>, ci si concentra sull'analisi della struttura informativa realizzata dal modello SDL – CINECA.

Tuttavia, prima di affrontare nel dettaglio lo studio dell'infrastruttura tecnologica e informativa necessaria alla realizzazione e gestione di un sistema *digital library* è indispensabile analizzare il modello su cui si basa tutto il suo sviluppo: Il *DELOS Digital Library Reference Model* presenta la *digital library* come «*an evolving organization that comes to existence through a series of development steps that bring*

---

<sup>12</sup> THE INTERPARES 3 PROJECT, *Conservazione archivistica*, in Team Italy Terminology Database, [http://www.interpares.org/ip3/ip3\\_terminology\\_db.cfm?term=1227](http://www.interpares.org/ip3/ip3_terminology_db.cfm?term=1227), 2007/2012 (aprile 2015).

<sup>13</sup> ISO 14721:2012 "Space data and information transfer systems - Open Archival Information System (OAIS) - Reference model", disponibile all'indirizzo: [http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=57284](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284). In lingua italiana: G. MICHETTI (a cura di), *OAIS – Sistema Informativo Aperto per l'Archiviazione*, ICCU, 2007.

together all necessary constituents»<sup>14</sup>. Questi step evolutivi si concretizzano in una struttura tripartita, modellata su tre distinte nozioni di “sistemi”, a cui corrispondono tre differenti livelli di concettualizzazione dell’universo delle *digital library*: *Digital Library*, *Digital Library System* e *Digital Library Management System*.

- *Digital Library (DL)*: un’organizzazione, anche virtuale, che complessivamente colleziona, gestisce e conserva a lungo termine contenuti digitali, e in riferimento a questi, in accordo a *policies* codificate, offre alla sua comunità di utenti specifiche funzionalità dalla qualità valutabile.
- *Digital Library System (DLS)*: un sistema software basato su un’architettura definita, possibilmente distribuita, che fornisce tutte le funzionalità richieste da una particolare *digital library*; la comunità di utenti interagisce con una *digital library* attraverso il corrispondente *digital library system*.
- *Digital Library Management System (DLMS)*: un software di sistema che fornisce l’infrastruttura software appropriata sia per la produzione e amministrazione di un *digital library system*, incorporando il set di funzionalità considerate fondamentali per una *digital library*, sia per l’implementazione di software aggiuntivi che offrono funzionalità maggiormente avanzate, rifinite o specialistiche.

Tale modello di riferimento ha ispirato il disegno del Servizio *Digital Library* CINECA, il quale realizza un *Digital Library Management System* che fornisce l’infrastruttura tecnologica e informativa necessaria alla creazione e alla gestione di un universo *digital library*. Tale *DLMS* da un lato incorpora una suite di funzionalità di base, un nucleo minimo di servizi caratterizzanti qualunque sistema *DL*, quali catalogazione, archiviazione, indicizzazione di metadati, di contenuti e accesso, dall’altro integra applicazioni e strumenti per il supporto a servizi specializzati, indispensabili alla realizzazione di specifiche funzioni avanzate.

Il servizio, progettato e realizzato in base all’adozione di modelli, specifiche e standard internazionali (come OAI, ISO-16363, METS, MODS, PREMIS, OAI-PMH), ha come obiettivi primari garantire la massima interoperabilità con analoghi sistemi alla base di progetti nazionali e internazionali (*in primis* Europeana)<sup>15</sup> e supportare tipologie di risorse digitali e metadati eterogenei. Per questi motivi l’architettura tecnologica di sistema, basata sull’interoperabilità, sull’integrazione e sulla modularità, è stata realizzata prediligendo l’impiego di software *open source*. Inoltre, queste caratteristiche rendono possibile l’integrazione con altri servizi già

---

<sup>14</sup> DELOS, *The DELOS Digital Library Reference Model. Foundations for Digital Libraries*, <https://www.coar-repositories.org/files/D3-2b-Digital-Library-Reference-Model.pdf>, novembre 2007 (aprile 2015), p. 17.

<sup>15</sup> Europeana (<http://www.europeana.eu>) è una biblioteca digitale europea che riunisce contributi già digitalizzati da diverse istituzioni dei 28 paesi membri dell’Unione europea in 30 lingue. La sua dotazione include libri, film, dipinti, giornali, archivi sonori, mappe, manoscritti ed archivi.

presenti in Sapienza, come portali di ateneo, piattaforme *e-Learning*, ecc. L'obiettivo dell'interoperabilità accresce indubbiamente il valore complessivo dell'intero progetto, che risulta scientificamente valido in quanto affonda le proprie radici negli standard internazionali precedentemente elencati.

L'architettura è di tipo *SOA (Service – Oriented Architecture)*<sup>16</sup>, organizzata su tre livelli: *repository*, livello di integrazione e livello applicativo. Fra questi, il livello di integrazione, denominato *Digital Library Service Delivery Platform (DL SDP)*, costituisce il fulcro centrale del sistema: permette, infatti, di realizzare un modello astratto e semantico delle entità e delle loro relazioni nella *digital library*, e, al tempo stesso, consente sia la comunicazione interna tra tutte le componenti e i servizi integrati al livello inferiore, sia l'accesso da parte del livello applicativo a tutte le funzionalità *digital library* – prima fra tutte il *repository* – esposte verso utenti o applicazioni esterne (ad esempio aggregatori OAI-PMH).

Il livello applicativo prevede l'impiego di un'omogenea interfaccia *digital library API (Application Programming Interface)*<sup>17</sup> basata su *web services* di tipologia RESTful, la quale, essendo l'unico punto di accesso verso tutti i servizi offerti, semplifica e rende trasparente l'utilizzo del *repository* e dei sistemi e servizi ad esso connessi. Dunque questo *layer* può in realtà potenzialmente concretizzarsi in una qualsiasi applicazione, anche *mobile*, capace di sfruttare servizi web. Questa struttura permette l'erogazione di un set di funzionalità basilari, un nucleo minimo di servizi generali caratterizzanti pressoché qualsiasi *digital library*:

- acquisizione delle risorse digitali: la funzionalità di *ingestion* è sviluppata conformemente al modello *OAIS*; i *Submission Information Package (SIP METS)*, forniti secondo le modalità previste, vengono elaborati da procedure automatiche in modo coordinato con tutte le altre componenti del servizio. Questi processi, monitorabili da parte dell'ateneo, prevedono controlli di integrità dei file, controlli antivirus, *file characterization* e integrazione con il processo di transcodifica;
- transcodifica: questo servizio, utilizzando anche soluzioni *open source* esterne come ConceptMapper e Mediamosa, permette di ottenere, in base alla tipologia di risorsa in ingresso, nuove forme di rappresentazione digitale utili alla *dissemination* (transcodifiche audio-video, conversioni di immagini, estrazione di *thumbnail*) o alla fase di ricerca e *discovery* (transcodifiche OCR, analizzatori della lingua, estrattori di testo o analizzatori semantici);

---

<sup>16</sup> Ricorrendo a questo termine si indica un'architettura software adatta a supportare l'uso di servizi web per garantire l'interoperabilità tra diversi sistemi, in modo tale da consentire l'utilizzo delle singole applicazioni come componenti del processo di business e soddisfare le richieste degli utenti in modo integrato e trasparente.

<sup>17</sup> Con "*Application Programming Interface*" si indica ogni insieme di procedure disponibili al programmatore, di solito raggruppate a formare un set di strumenti specifici per l'espletamento di un determinato compito all'interno di un certo programma. Spesso con tale termine si intendono le librerie software disponibili in un certo linguaggio di programmazione.

- deposito digitale: le funzioni di memorizzazione e gestione dei contenuti, dei relativi metadati, dei derivati e delle informazioni di autorizzazione sono svolte dal *repository open source* Fedora Commons integrato nel servizio;
- indicizzazione e ricerca: queste funzioni sono affidate al motore di ricerca open source SOLR<sup>18</sup>. Una volta realizzata l'integrazione, l'indice dei contenuti, dei metadati, delle relazioni e dei derivati (trascrizioni OCR, annotazioni semantiche, testi) risulta allineato con il *repository*; inoltre, attraverso la API di accesso, le informazioni possono essere sfruttate dal portale per la ricerca e il *browsing* delle risorse digitali;
- *dissemination*: l'accesso alle risorse digitali e ai servizi da parte dell'utente avviene attraverso il Portale Digital Library, realizzato con il CMS *open source* Drupal; il Portale è stato integrato anche per condividere con il servizio *digital library* le informazioni di autenticazione e autorizzazione sulle risorse digitali.

## Conclusioni

«*Constant monitoring, planning, and maintenance of the repository, as well as conscious actions and strategy implementation will be required of repositories to carry out their mission of digital preservation*»<sup>19</sup>: da questa constatazione si evince che i fattori individuati come essenziali per una conservazione digitale scientificamente soddisfacente consistono in un monitoraggio, una pianificazione e una gestione rigorosamente costanti, non soltanto relativamente al singolo *repository*, ma in generale all'intero archivio. Dallo studio del caso specifico SDL – CINECA è emersa in maniera piuttosto chiara quella che si potrebbe definire una direzione comune verso cui convergono le attività precedentemente enunciate, una sorta di *leitmotiv* che le caratterizza indistintamente: la garanzia dell'affidabilità dell'intero archivio, inteso, sempre sulla scorta del modello OAIS, come un complesso e ampio sistema informativo per l'archiviazione.

Grazie a questa analisi e agli studi del Gruppo di lavoro SDL sulla conservazione digitale, è possibile definire tre fattori generali a cui ricondurre tutte le operazioni e le procedure, concettualmente incluse nelle attività di monitoraggio, pianificazione e gestione, che concorrono ad assicurare la *trustworthiness* di un intero sistema:

1. La **sostenibilità** dell'intero progetto nel tempo. Ovviamente diversi fattori partecipano al raggiungimento di un obiettivo così complesso: dalla definizione di un chiaro modello organizzativo generale alla creazione di una ben precisa *governance*, dall'individuazione di precisi ruoli e responsabilità dei soggetti coinvolti alla condivisione di un modello di business (che possa tener conto, ad esempio, delle

<sup>18</sup> <http://lucene.apache.org/solr> (aprile 2015).

<sup>19</sup> CCSDS, *Audit and Certification of Trustworthy Digital Repositories*, Magenta Book (Issue 1), settembre 2011, <http://public.ccsds.org/publications/archive/652x0m1.pdf>, p. 15 (aprile 2015).

quote per i costi vivi di gestione, che sia in grado di prevedere i costi di sviluppo, ecc.), e così via.

2. La **certezza dei contenuti archiviati**. E quindi, considerando che «*the trustworthiness of a repository system relies on the ability of tracking back information about the custodianship of the objects*»<sup>20</sup>, acquista fondamentale importanza il *metadata framework* concepito per la generazione dei pacchetti informativi. A tal proposito vale la pena sottolineare come la struttura dei metadati realizzata nell'ambito della SDL risulti rispettosa dei requisiti necessari alla LTDP (*Long Term Digital Preservation*).

3. La definizione di **policies, documentazione e regolamentazione** di tutte le procedure rilevanti ai fini conservativi. Si tratta naturalmente di un obiettivo complesso, che investe anche ambiti e processi concettualmente esterni al modello OAIS, e che infatti prevede diverse operazioni:

- definizione delle policies rilevanti a fini conservativi, sia presso Sapienza (DigiLab e SBS) che presso CINECA, coerentemente con quanto previsto dallo standard ISO 16363<sup>21</sup> e con quanto indicato dalle regole tecniche sulla conservazione digitale: manuale di conservazione, rapporto di versamento e integrazione con il manuale di gestione, relativamente a Sapienza;
- documentazione degli interventi significativi gestiti dal CINECA nel sistema di conservazione/archiviazione, in riferimento a migrazioni, nuove aggregazioni, interventi di *refreshment* e di gestione dei supporti, insieme alla valutazione delle raccomandazioni del progetto APARSEN in materia di *authenticity evidence*<sup>22</sup>;
- predisposizione di regolamenti di servizio interni a DigiLab e a SBS (Sapienza), per la gestione dei flussi di digitalizzazione e di versamento – relativamente alle responsabilità per le collezioni, alle schede di lavorazione ecc. – nonché per la tenuta della documentazione rilevante.

Proprio alla luce di questo ultimo fattore si è innanzitutto reso necessario fornire un'accurata descrizione dei processi, delle operazioni e delle *policies* attuate in merito

---

<sup>20</sup> A. DI IORIO, M. SCHAERF, M. GUERCIO, S. ORTOLANI, M. BERTAZZO, *A Digital Infrastructure for Trustworthiness. The Sapienza Digital library Experience*, in *Bridging Between Cultural Heritage Institutions*, IRCDL 2013, Roma, Italia, 2013, [http://link.springer.com/chapter/10.1007%2F978-3-642-54347-0\\_7](http://link.springer.com/chapter/10.1007%2F978-3-642-54347-0_7) (aprile 2015), p. 65 (7/11).

<sup>21</sup> ISO 16363:2012 "*Space data and information transfer systems - Audit and certification of trustworthy digital repositories*", disponibile all'indirizzo: [http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510) (aprile 2015).

<sup>22</sup> Si fa qui riferimento, in generale, al lavoro svolto nell'ambito del progetto *Alliance for Permanent Access to the Records of Science in Europe Network* riguardo al concetto di *authenticity* (<http://www.alliancepermanentaccess.org/index.php/community/common-vision/authenticity>) e, in particolare, ai risultati del WP24 "*Authenticity and Provenance*": <http://www.alliancepermanentaccess.org/index.php/aparsen/aparsen-research/wp24-authenticity-and-provenance> (aprile 2015).

sia alle attività di digitalizzazione che di conservazione digitale nell'ambito del progetto SDL. Una descrizione non fine a se stessa, ma propedeutica a un'analisi delle strategie da cui quelle stesse *policies* sono derivate.

In riferimento allo studio delle *policies* di digitalizzazione, il modello organizzativo che costituisce la base del lavoro, appositamente elaborato a partire da *recommendations* internazionali, ha consentito, con la sua suddivisione in fasi e attività, di coprire in maniera pressoché totale l'intero quadro delle operazioni in questione, proponendosi in tal senso come un possibile paradigma organizzativo per futuri progetti di digitalizzazione, anche in contesti differenti da quello a cui la SDL appartiene. Inoltre, grazie al modello in questione è stato possibile elaborare, a conclusione dell'analisi, una puntuale schematizzazione delle criticità e dei punti di forza emersi dall'indagine dei diversi processi e ambiti di cui consiste la digitalizzazione in SDL, consultabile alla tavola "Criticità e punti di forza del processo di digitalizzazione in SDL".

Relativamente alla strategia e alle *policies* previste dal modello di conservazione digitale concepito nell'ambito del progetto, l'indagine della struttura informativa relativa al DLMS CINECA ha evidenziato, almeno sulla carta, una piena corrispondenza con lo specifico modello di riferimento internazionale, il *DELOS Digital Library Reference Model*. Inoltre, la struttura materiale, cioè tecnologica, del sistema, basata sull'integrazione e sulla modularità, risulta effettivamente in grado non solo di supportare tipologie di risorse digitali e metadati eterogenei, ma anche di consentire l'interoperabilità con analoghi sistemi alla base di progetti nazionali e internazionali, oltre all'integrazione con servizi già presenti in Sapienza, come portali di ateneo, piattaforme *e-learning*, archivio PADIS e piattaforma OJS. Infine, l'obiettivo dell'interoperabilità accresce indubbiamente il valore complessivo dell'intero progetto, che risulta scientificamente valido in quanto affonda le proprie radici negli standard internazionali di settore.

Alessandro Ottavio Cuomo  
Dottore magistrale in Archivistica e Biblioteconomia

### **Bibliografia e Sitografia**

A. DI IORIO, M. SCHAERF, M. GUERCIO, S. ORTOLANI, M. BERTAZZO, *A Digital Infrastructure for Trustworthiness. The Sapienza Digital library Experience*, in *Bridging Between Cultural Heritage Institutions*, IRCDL 2013, Roma, Italia, <[http://link.springer.com/chapter/10.1007%2F978-3-642-54347-0\\_7](http://link.springer.com/chapter/10.1007%2F978-3-642-54347-0_7)>, 2013 (dicembre 2014).

ALLIANCE FOR PERMANENT ACCESS TO THE RECORDS OF SCIENCE IN EUROPE NETWORK (APARSEN), *Integrated view of digital preservation: Authenticity*, <<http://www.alliancepermanentaccess.org/index.php/community/common-vision/authenticity>>, (dicembre 2014).

CCSDS, *Audit and Certification of Trustworthy Digital Repositories, Magenta Book (Issue 1)*, <<http://public.ccsds.org/publications/archive/652x0m1.pdf>>, settembre 2011 (novembre 2014).

CINECA – CONSORZIO INTERUNIVERSITARIO, <<http://www.cineca.it/it>> (dicembre 2014).

DELOS, *The DELOS Digital Library Reference Model. Foundations for Digital Libraries*,

<[http://www.delos.info/files/pdf/ReferenceModel/DELOS\\_DLReferenceModel\\_096.pdf](http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_096.pdf)>, novembre 2007 (dicembre 2014).

DIGITAL PRESERVATION COALITION (DPC), *Preservation Management of Digital Materials: the Handbook*, <<http://www.dpconline.org/advice/preservationhandbook>>, ottobre 2008 (novembre 2014).

EUROPEANA, *Europeana Data Model (EDM) Documentation*, <<http://pro.europeana.eu/edmdocumentation>> (novembre 2014).

FADGI STILL IMAGE WORKING GROUP, *Still Image Digitization Working Group Charter*, <<http://www.digitizationguidelines.gov/guidelines/StillImageCharter.pdf>>, luglio 2008 (novembre 2014).

FEDERAL AGENCIES DIGITIZATION GUIDELINES INITIATIVE (FADGI), <<http://www.digitizationguidelines.gov/about>> (dicembre 2014).

IFLA, ICA, *Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives*, <<http://www.ifla.org/publications/guidelines-for-digitization-projects-for-collections-and-holdings-in-the-public-domain>>, marzo 2002 (novembre 2014).

ISO 14721:2012, *Space data and information transfer systems - Open Archival Information System (OAIS) - Reference model*, <[http://www.iso.org/iso/home/store/catalogue\\_tc/catalogue\\_detail.htm?csnumber=57284](http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=57284)> (dicembre 2014).

ISO 16363:2012, *Space data and information transfer systems - Audit and certification of trustworthy digital repositories*, <[http://www.iso.org/iso/catalogue\\_detail.htm?csnumber=56510](http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510)>, 2012 (novembre 2014).

G. MICHETTI (trad. it. a cura di), *OAIS – Sistema Informativo Aperto per l'Archiviazione*, ICCU, 2007.

MINERVA WORKING GROUP 6, *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale*, <[http://www.minervaeurope.org/structure/workinggroups/goodpract/document/buonepratiche1\\_3.pdf](http://www.minervaeurope.org/structure/workinggroups/goodpract/document/buonepratiche1_3.pdf)>, marzo 2004 (novembre 2014).

NARA, *Technical Guidelines for Digitizing Archival Materials for Electronic Access*, <<http://www.archives.gov/preservation/technical/guidelines.html>>, giugno 2004 (novembre 2014).

NINCH, HATII, *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials*, <<https://www.prestocentre.org>>, 2002 (novembre 2014).

THE INTERPARES 3 PROJECT, *Conservazione archivistica*, in Team Italy Terminology Database, <[http://www.interpares.org/ip3/ip3\\_terminology\\_db.cfm?term=1227](http://www.interpares.org/ip3/ip3_terminology_db.cfm?term=1227)>, 2007/2012 (dicembre 2014).

UNIVERSITÀ DEGLI STUDI DI ROMA “LA SAPIENZA”, *DigiLab – Centro interdipartimentale*, <[http://w3.uniroma1.it/biblioteche/chi\\_siamo.htm](http://w3.uniroma1.it/biblioteche/chi_siamo.htm)> (dicembre 2014).

UNIVERSITÀ DEGLI STUDI DI ROMA “LA SAPIENZA”, *Sapienza Digital Library*, <<http://sapienzadigitalibrary.uniroma1.it/>> (dicembre 2014).