

# **Digital Math Library DML: un progetto mondiale per la matematica**

di Antonella De Robbio<sup>1</sup>

*A good library is a place,  
a palace where the lofty spirits of all nations and generations meet.*  
— Samuel NIGER (1883-1956)

## **Abstract**

*Il presente elaborato descrive lo stato dell'arte del progetto mondiale DML Digital Mathematical Library il quale si pone l'obiettivo di rendere disponibili e liberamente accessibili, in formato digitale, la totalità dei documenti di matematica, stimati in circa 50 milioni di pagine. La DML comprenderà sia collezioni di documenti retro-digitalizzati (past literature) sia collezioni di documenti in formato nativo digitale. Numerosi i soggetti coinvolti, tra cui biblioteche, università, editori, progetti e le società matematiche nazionali e internazionali. Il tempo previsto per la realizzazione del progetto è di dieci anni, il budget totale di 100 milioni di dollari. Il progetto mira anche a creare un coordinamento tra le esperienze di digitalizzazione già esistenti, attraverso una rete che poggia su un modello distribuito. Numerosi i nodi da sciogliere tra cui il copyright, i formati, l'archiviazione. Molti gli aspetti su cui si dovrà decidere, tra cui la selezione dei contenuti, gli standard tecnici e i formati di metadati, tutti argomenti caldi dei quali si è discusso durante il meeting tenutosi a Berlingen (lago di Costanza) dal 31 gennaio-3 febbraio 2003. Nel brainstorming di Berlingen, a cui ha partecipato anche l'Italia, i gruppi di lavoro costituitisi allo scopo di predisporre documenti utili alla richiesta di finanziamento entro il 6. Programma Quadro dell'Unione Europea, hanno delineato le linee strategiche di pianificazione e sviluppo del progetto europeo UE Digital Math Library.*

## **Contenuti:**

**Il progetto DML Digital Mathematical Library**

**Gli scopi del progetto DML**

**Le motivazioni del progetto DML: i punti di forza**

**L'architettura organizzativa del progetto**

**Le fasi di progetto**

**Alcuni progetti di digitalizzazione già esistenti**

**Il progetto Europeo UE Digital Math Library**

**L'incontro di Berlingen per la richiesta di finanziamento europeo**

**Partner di progetto e stakeholder**

**Contributi finanziari dall'Unione Europea**

**Pianificazione del progetto UE DML**

**I contenuti**

**Standard tecnici e formati**

**Il formato DjVu<sup>2</sup>**

**Archiviazione**

**Copyright**

---

<sup>1</sup> Centro di Ateneo per le Biblioteche. Biblioteca del Seminario Matematico. Università degli studi di Padova

<sup>2</sup> Per una descrizione tecnica del formato DjVu vedi articolo di Yann LeCun, Léon Bottou, and Patrick Haffner *DjVu: a Compression Technique and Software Platform for Publishing Scanned Documents, Digital Documents and High-Resolution Images on the Web.*

<http://djvu.sourceforge.net/abstract.html>

## **Il modello economico**

### **Aspettative e previsioni di progetto**

### **Il progetto italiano ItalDML**

### **Bibliografia di riferimento**

## **Il progetto DML *Digital Mathematical Library***

Il progetto DML *Digital Mathematical Library* si pone l'obiettivo di rendere disponibili e liberamente accessibili a tutta la comunità, in formato digitale, la totalità dei documenti di matematica, in modo facilitato e unificato, stimati in circa 50 milioni di pagine. La DML comprenderà sia collezioni di documenti retro-digitalizzati sia collezioni di documenti in formato nativo digitale le quali potranno essere accessibili da chiunque e da qualsiasi parte del globo.

Il progetto, assai ambizioso, si colloca a livello mondiale in due anime strettamente interagenti, da una parte il progetto statunitense con l'American Mathematical Society (AMS), dall'altra parte il progetto europeo con l'European Mathematical Society (EMS), attorno una costellazione di altri soggetti tra cui i maggiori editori scientifici, le biblioteche di università e dei centri di ricerca, le associazioni di matematica nazionali ed internazionali. Sul versante statunitense la Cornell University, che cerca il supporto del National Science Foundation, e la Michigan University assieme a numerosi editori accademici e alle società scientifiche stanno iniziando una serie di progetti coordinati assai imponenti. Allo scopo viene commissionato dal NSF Division of Mathematical Sciences a John Ewing, Executive Director of the American Mathematical Society, il *white paper* "*Twenty Centuries of Mathematics,*" panoramica che descrive gli aspetti fondamentali e la particolare importanza della letteratura "past", o letteratura retro-digitalizzata, per la ricerca matematica.

Nel suo libro bianco "*Venti secoli di matematica: digitalizzando e disseminando il passato della letteratura matematica*", da cui si attinge per la stesura di alcune parti del presente lavoro, John Ewing individua quattro grossi ostacoli al progetto: contenuti, copyright, formati, archiviazione i quali vanno a costituire i rischi di progetto di cui bisognerà tener conto.

Su queste problematiche si è discusso durante il meeting di Berlingen (31 gennaio-3 febbraio 2003) nei gruppi di lavoro costituitisi allo scopo di predisporre documenti utili alla richiesta di finanziamento entro il 6. Programma Quadro dell'Unione Europea.

I risultati dei vari white papers commissionati dal NSF ai gruppi di esperti nelle varie aree di interesse sono stati discussi lo scorso agosto 2003 a Pechino durante l'International Congress of - ICM2002 Beijing. In tale contesto si è instaurata la collaborazione con la Tsinghua University Library (China), entro il progetto DML.

Inoltre la Cornell ha creato e mantenuto un sito mirror per i maggiori database bibliografici A&I europei, per esempio il *Zentralblatt für Mathematik*, ha stretto alleanze con editori accademici come la Duke University Press entro il progetto Euclid, un'infrastruttura per l'editoria elettronica per periodici matematici correnti e prossimamente anche per quelli da retro-digitalizzare e con la Springer-Verlag entro il progetto Archimedes al fine di assicurare la conservazione della letteratura matematica. E' probabile anche un prossimo coinvolgimento della Cornell entro il progetto JSTOR.

## **Gli scopi del progetto DML**

I tre scopi primari di questo progetto mondiale sono:

1. digitalizzare una parte preponderante della letteratura matematica scientifica che non esiste già in formato digitale e che viene definita "past literature", la quale è collocata a livello temporale prevalentemente prima del 1997

2. predisporre un insieme di standard tecnici per rendere i documenti di matematica digitalizzati accessibili online a chiunque
3. negoziare un protocollo per rendere la letteratura matematica digitale disponibile anche in futuro

Ciò significa servire le comunità scientifiche, gli studenti e, più generalmente, tutti i cittadini del mondo, offrendo un accesso facilitato ed efficace ad un largo corpus di documenti matematici, ma in particolare all'importante patrimonio del materiale a stampa. I dati grezzi entro la DML dovrebbero essere liberamente disponibili a tutti via Internet, regole del copyright permettendo.

## **Le motivazioni del progetto DML: i punti di forza**

La matematica è l'unica scienza, a differenza delle altre, che ha una stretta dipendenza dalla letteratura scientifica passata, per questa ragione la comunità matematica mondiale si propone di digitalizzare l'esistente letteratura su carta in modo da creare una Biblioteca Matematica Digitale disponibile in Rete a tutti i matematici del mondo.

In analogia con le altre scienze, la matematica si sta sviluppando in modo incrementale, ma ha anche una configurazione particolare, in quanto i risultati e i metodi di tutti i periodi rimangono di interesse e formano parte della base dei nuovi sviluppi. Perciò un alto livello di documenti matematici non diviene mai obsoleto e i dati richiesti per conseguire nuovi risultati di ricerca sono contenuti nell'intera collezione comprendente tutte le pubblicazioni passate che le presenti, che se dotata di un opportuno motore di ricerca, potrebbe offrire un potente strumento bibliometrico, per l'analisi delle citazioni. L'emivita delle citazioni di un lavoro matematico (con l'esclusione della teoria computazionale) è quasi sempre superiore ai 10 anni.

Le idee e le tecniche per risolvere problemi contemporanei in matematica e nelle altre scienze spesso vengono da lavori matematici pubblicati nel passato.

Dal diciassettesimo secolo la matematica è stata una congiunzione indispensabile per le scienze fisiche e per la tecnologia e nei periodi più recenti ha assunto un ruolo simile negli aspetti quantitativi delle scienze della vita.

Secondo la comunità dei matematici che ha ideato questo progetto, l'intera letteratura matematica consiste approssimativamente di 50 milioni di pagine contenute in monografie, periodici, atti di convegni e altre tipologie di documenti. Il volume è alto, però può essere discretamente gestibile se lo sforzo diviene globale; ci si aspetta che le istituzioni matematiche, le società accademiche, gli editori, vorranno accrescere le collezioni offrendo valore aggiunto come il cross-linking tra documenti.

## **L'architettura organizzativa del progetto**

Tutto il sistema si basa su un modello distribuito, sostenuto dalla base costituita da numerosi soggetti.

Oltre alla Steering Committee e alla IMU Liaison Committee, entro il progetto DML operano vari Working Groups su tematiche specifiche:

- ❑ WG sui contenuti
- ❑ WG sugli standard tecnici
- ❑ WG sui metadati
- ❑ WG sulle licenze e copyright
- ❑ WG sull'archiviazione
- ❑ WG sul modello economico

Le azioni si svilupperanno attorno a queste linee di pianificazione:

- ❑ Obiettivi di programma
- ❑ Contenuti
  - ❑ Scopi
  - ❑ Selezione dei materiali da digitalizzare
  - ❑ Priorità per la digitalizzazione
- ❑ Specifiche tecniche
  - ❑ Standard per la digitalizzazione
  - ❑ Standard di metadati
  - ❑ Standard per l'interoperabilità
  - ❑ Standard per il reference linking
  - ❑ Standard per l'archiviazione
- ❑ Copyright e proprietà intellettuale
- ❑ Project management
- ❑ Costi e modello economico
- ❑ Gestione e mantenimento

Il nodo dell'archiviazione, che peraltro è un problema spesso frainteso anche dagli esperti, non è tanto un problema delle fasi di avvio o di applicazione del progetto quanto per chi manterrà la collezione dopo la chiusura del progetto.

Purtroppo in un'area come l'archiviazione digitale, dove non ci sono veri esperti in quanto nessuno ha molta esperienza, la questione dell'archiviazione viene confusa il correlato "problema del formato". Ridurre il grande problema dell'archiviazione facendolo confluire nel problema del formato banalizza la discussione, riducendo la questione a un mero problema di presentazione. Di fatto, il problema centrale risiede nell'archiviazione.

Quando usiamo il termine archiviazione che ha sfumature diverse rispetto al correlato termine "conservazione" è necessario comprendere che con l'avvento delle nuove tecnologie e nel nuovo ambiente digitale il ruolo dell'archiviazione è mutato.

Come sottolinea Ewing nel suo libro bianco, "fino a tempi recenti, non c'era un gran bisogno di archiviare in modo consapevole riviste o libri scientifici, l'archiviazione era quasi automatica perché molte copie erano distribuite alle istituzioni in diverse località. Ci si affidava alle leggi della probabilità per assicurarsi che almeno una copia avrebbe continuato a esistere, negli anni a venire. Quell'unica copia avrebbe potuto essere utilizzata come fonte per la riproduzione simultanea di molte copie, anni dopo la pubblicazione iniziale."

I nuovi modelli proposti dall'editoria elettronica per la pubblicazione di materiale digitale hanno comportato due ribaltamenti nei processi di archiviazione. Il primo si riconduce al concetto di "robustezza" ed è di diretta derivazione dal processo di archiviazione.

Il secondo si riferisce all'impossibilità di riprodurre in futuro, a distanza di anni, copie dell'opera pienamente funzionali e identiche a quelle originali.

Il concetto di "robustezza" nasce appunto come problema tipico dell'ambiente digitale in quanto le copie "originali" possono non essere a larga distribuzione come avviene invece con la carta, ma molto spesso risiedono in forma elettronica in uno o due siti in tutto il mondo. E' molto facile quindi che possano succedere catastrofi digitali e qui entra in gioco la questione della sicurezza dei dati, dove gli unici esemplari possono andare irrimediabilmente distrutti.

Per il secondo punto va considerato che riviste e libri elettronici sono spesso costituiti di file contestualizzati entro un sistema più ampio che utilizza programmi, file ausiliari e hardware specifico per presentare l'opera e riprodurre copie in modo distribuito. In sostanza la copia ha bisogno del suo ambiente operativo originario. Come dice Ewing il contesto in cui l'opera è inserita

è spesso essenziale per realizzare una copia fedele, e l'archiviazione richiede la capacità di riprodurre quel determinato contesto.

## **Le fasi di progetto**

Il tempo previsto per la realizzazione del progetto è di dieci anni, suddivisi in tre fasi:

### **1. Fase di disegno: 1-2 anni**

Durante questa fase tre commissioni si occuperanno di contenuti, standard e indicazioni. La prima lavorerà sulla selezione dei contenuti e deciderà cosa dovrà essere digitalizzato. La seconda si occuperà degli standard tecnici in stretta connessione con gli standard per l'archiviazione. La terza commissione, più ridotta, dovrebbe rappresentare la comunità dei matematici, fornendo indicazioni sulle maggiori decisioni del progetto e avrà la responsabilità di stabilire i protocolli per aggiungere nuovo materiale da digitalizzare per l'accrescimento delle collezioni future. Si costruiranno le linee guida per la struttura del progetto coinvolgendo i partner e i soggetti interessati (stakeholders). Si stabilirà il modello economico da adottare, i criteri di selezione e le priorità per la determinazione dei contenuti, la messa a punto degli standard tecnici su metadati, interoperabilità e linking, le politiche sul copyright, di archiviazione a lungo termine e il supporto tecnico.

Le attività della prima fase previste sono:

- ❑ individuare le realizzazioni nazionali per progetti di digitalizzazione di ambito nazionale al fine di evitare duplicazioni, come purtroppo già avvenuto nel passato
- ❑ predisporre le linee guida per i progetti di digitalizzazione
- ❑ predisporre le politiche di copyright e di IPR (Intellectual Property Rights) che dovranno essere calate nelle realtà nazionali (leggi nazionali)
- ❑ stabilire standard e formati di metadati
- ❑ individuare i formati idonei ai documenti da trattare
- ❑ stabilire gli standard tecnici per la digitalizzazione, le modalità e i protocolli per l'archiviazione e per il recupero dei documenti

### **2. Fase di implementazione: 2-8 anni**

Dopo un periodo di iniziale di due anni, i lavori in questa fase continueranno approssimativamente per otto anni. In questo periodo il materiale digitalizzato che giunge dai vari progetti sarà reso disponibile alle varie organizzazioni disseminandolo alla comunità, incrementando le collezioni digitali disponibili. Poiché differenti gruppi di lavoro potrebbero procedere in modo disomogeneo, in merito a qualità e in rapporto alla tipologia di progetto, il controllo sulla qualità sarà coordinato da uno staff centrale che funge da corpo d'autorità.

- ❑ Digitalizzare circa 50 milioni di pagine (200 Gigabyte)

### **3. Fase operativa: 1-2 anni**

Saranno adottati protocolli per le nuove collezioni che via via si aggiungeranno. In questa fase si collocheranno anche i processi per le specifiche e eventuali modifiche a standard e protocolli.

- ❑ Mantenimento e aggiornamento dei dati

Alla fine dei dieci anni, questo sistema dovrebbe essere sostenuto dalle varie organizzazioni nel mondo, ciascuna con un proprio interesse locale, ma al contempo con l'interesse comune volto all'intera comunità di ricerca dei matematici. Aggiungere materiale alle collezioni diverrà una normale prassi del processo di pubblicazione, con benefici nei costi dovuti alla standardizzazione dei processi e all'amministrazione della collezione che sarà su piccola scala.

## **Alcuni progetti di digitalizzazione già esistenti**

E' stato stimato che il costo per la digitalizzazione si attesta sui 2 Euro per pagina, inclusi scansione, inserimento di linking e di keyword ricercabili, per un totale di 100 milioni di Euro. Una piccola parte di questi 50 milioni di pagine è già stata digitalizzata, entro alcuni progetti non coordinati tra loro e comunque in modo frammentario e non standardizzato.

Il progetto mira anche a creare un coordinamento tra le esperienze già esistenti.

Poiché la letteratura storica è stata digitalizzata da separate istituzioni ed è accessibile attraverso differenti interfacce utenza, con differenti metodologie di navigazioni e vari livelli di funzionalità, gli utenti hanno difficoltà sia nella scoperta sia nel recupero e conseguente utilizzo della letteratura online. Non vi è un approccio sistematico alla conservazione digitale di questo materiale online. Vi è quindi bisogno di coordinare progetti e ancora più importante è collaborare alla costruzione di una somma che risulterà più grande e preziosa delle singole parti, non soltanto perché aumenta la quantità di materiale a livello numerico, ma perché aggiungendo processi comuni di distribuzione, omogeneizzando le funzionalità, si offrono strumenti di più facile utilizzo all'utente finale, aumentando il valore aggiunto in termini di efficienza ed efficacia generali.

Per una lista delle maggiori collezioni digitalizzate di matematica si rimanda al repertorio curato da Steven Rockey "*Mathematics Digitization*"

<http://www.math.cornell.edu/~library/digitalization.html>

Si citano di seguito alcuni progetti già esistenti:

DIEPER Digitised European PERiodicals

Numerose università europee e biblioteche nazionali, coordinate da SUB Gottingen

<http://gdz.sub.uni-goettingen.de/dieper/>

Gottinger Digitalisierungszentrum

La collezione "Mathematica" contiene 249 documenti tra periodici e monografie

<http://gdz.sub.uni-goettingen.de/>

ERAM Electronic Research Archive for Mathematics

<http://www.emis.de/MATH/JFM/JFM.html> in Germania

Nell'ambito dei progetti europei per la matematica in seno all'EMIS vi è il database "**The Jahrbuch-project - Electronic Research Archive for Mathematics (ERAM)**"<sup>3</sup>

Fondato nel 1868 dai matematici Carl Ohrtmann e Felix Müller, si compone di 68 fascicoli per il periodo dal 1868 al 1943 con oltre 200.000 recensioni di pubblicazioni matematiche dell'epoca.

JFM al momento contiene i seguenti volumi già digitalizzati, comprensivi di citazioni bibliografiche e abstract:

1-20 (1868-1888); 25 (1893/94); 28-30 (1897-1899); 32-34 (1901-1903); 38-39 (1907-1908); 42 (1911); 45 (1914/1915); 50 (1924); 44, 46-49,51-56 (1913-1930): solo citazioni.

The Electronic Library of Mathematics

Mathematical Monographs and Lecture Notes

<http://www.emis.de/ELibM.html>

Mathematical Conference Proceedings. The Electronic Library of Mathematics

<http://www.emis.de/proceedings/index.html>

Classical Works, Selecta, and Opera Omnia The Electronic Library of Mathematics

<http://www.emis.de/classics/index.html>

---

<sup>3</sup> La pagina del progetto è raggiungibile al seguente indirizzo: <<http://www.emis.de/projects/JFM/>>  
L'interfaccia di accesso al database in versione test: <<http://www.emis.de/MATH/JFM/JFM.html>>

Gallica (Bibliothèque nationale de France) (circa 250 monografie)  
<http://gallica.bnf.fr/>

NUMDAM (NUMerisation de Documents Anciens Mathematiques)  
<http://www-mathdoc.ujf-grenoble.fr/NUMDAM/> Francia

JSTOR (periodici)  
<http://www.jstor.org/>

Cornell University and Michigan University  
Collezione di monografie di matematica digitalizzate in formato Dienst (577 monografie pre-1914)  
<http://www.math.cornell.edu/>  
<http://www.lib.umich.edu/>

Progetto transatlantico tra Cornell, Michigan e Gottingen "Distributed Digital Library of Mathematical Monographs"  
<http://www.library.cornell.edu/mathbooks/>, finanziato da NSF National Science Foundation e DFG.

[The On-line Books Page](#) (University of Pennsylvania [precedentemente alla Carnegie Mellon University])

E' una directory non un progetto, ma da qui si raggiungono parecchi testi in formato digitale  
<http://onlinebooks.library.upenn.edu/>

[Mathematics and Computer Science \(call numbers beginning QA\)](#) (382 monografie)

[Physical Sciences \(call numbers beginning QB-QE\)](#) (248 monografie)

[Science \(call numbers beginning Q\)](#) (1013 monografie)

## **Il progetto Europeo UE Digital Math Library**

Al fine di poter attuare questo ambizioso progetto è necessario avere supporto, entro l'architettura di un ambiente internazionale, delle società matematiche nazionali e dell'Unione Matematica Internazionale (IMU). Per questa ragione l'European Mathematical Society cerca il supporto dell'Unione Europea nel contribuire alla creazione e sviluppo di una Digital Math Library europea. Il Progetto europeo si sta muovendo in questo scenario in stretta connessione con i progetti statunitensi entro il quadro mondiale e alimentandosi dei progetti nazionali che vi aderiranno.

Lo scopo principale del progetto sia di parte statunitense sia di parte europea è creare una collezione di materiale che rappresenti la "past literature" di matematica, prevedendo anche un meccanismo di continuo aggiornamento per il mantenimento e sviluppo delle collezioni digitali correnti che andranno man mano aggiungendosi al corpo principale. Creare una collezione di materiale retro-digitalizzato richiede necessariamente che vi sia un meccanismo di aggiornamento e accrescimento delle collezioni in futuro. Sarà quindi essenziale predisporre un protocollo per l'aggiornamento delle collezioni nel tempo, differenziato per monografie e periodici, e questo comporterà delle alleanze con gli editori e le organizzazioni a seguito di negoziazioni che dovranno essere attivate per tempo, plasmate entro un modello economico che rispetti il contesto di riferimento. Per tale ragione il modello economico del progetto europeo potrebbe differire da quello statunitense, ragionevolmente per una serie di condizioni che vanno opportunamente considerate. La prima condizione è legata alla dimensione strutturale e organizzativa delle biblioteche europee rispetto a quelle oltreoceano. La seconda si riferisce alle differenti normative sulla proprietà intellettuale, che vedono il modello copyright da una parte e quello basato sul diritto d'autore in contesto europeo dall'altra, dove nel regime copyright la negoziazione dei diritti avviene tra editore e biblioteca digitale, nel sistema

normativo a diritto d'autore, tipico europeo, entrano in gioco anche gli autori, i quali detengono tutti i diritti e quindi devono essere interpellati al fine di ottenere il permesso alla digitalizzazione. La terza condizione è quella legata alla formalizzazione dei consorzi che, per quanto riguarda per esempio l'Italia, risulta assai problematica.

Alla fine la questione dei costi di accesso deve essere considerata. La produzione, gestione e mantenimento di una biblioteca digitale globale comporta dei costi i quali devono essere coperti in modo equo dagli utenti e dai singoli stati a supporto della ricerca. La versione digitale permette di offrire differenti prezzi per differenti regioni, ciò può evitare che l'Europa dell'Est o i Paesi in via di sviluppo siano esclusi dall'accesso, si dovrà quindi prevedere griglie differenziate per l'accesso ai materiali.

## **L'incontro di Berlingen per la richiesta di finanziamento europeo**

Durante il brainstorming UE Digital Math Library al meeting tenutosi in Berlingen (Lago di Costanza) dal 31 gennaio al 2 febbraio 2003 il programma dei lavori prevedeva la stesura di alcuni documenti tra cui la prima bozza delle linee guida per l'accordo consortile del DML-FP6, la richiesta di finanziamento alla comunità europea con la stesura del progetto, i possibili partners per il consorzio, la scrittura della proposta sulla base dell'espressione di interesse (EoI Expression of Interest) presentata all'Unione Europea da Rolf Jeltsch (ex presidente della EMS) e la bozza di statuto. Di fatto poi durante le varie sessioni e i gruppi di lavoro che si sono freneticamente succeduti fino a notte, e nelle conseguenti discussioni collettive dei circa trenta partecipanti sui singoli lavori di gruppo, si è ravvisata l'esigenza di suddividere i partecipanti per fasi successive dapprima in quattro e poi in tre gruppi su specifiche tematiche.

Rolf Jeltsch ha illustrato l'idea del progetto: fase primaria di due anni, fase di implementazione di tre anni, strumenti, focus, eccellenza, impatto e tecnologia, accesso al patrimonio culturale, fase finale di consolidamento del progetto, ulteriori tre anni.

Sulla base di questi aspetti si sono formati i primi nuclei lavoro del primo giorno.

Un gruppo si è riunito per lavorare sul Consorzio e sulle linee guida di pianificazione strategiche, organizzazione, individuazione dei partner, accordi consortili e linee di finanziamento. Un secondo gruppo ha lavorato sugli strumenti da mettere a disposizione entro una piattaforma interoperabile per la DML. Un terzo si è occupato degli aspetti della digitalizzazione, standard tecnici su formati e su metadati. Il quarto gruppo ha trattato la questione "discovery & retrieval" che comprende le interfacce, le problematiche dell'indicizzazione e del reference linking.

Va detto che molti degli argomenti sono stati trattati nei quattro gruppi in modo ricorsivo e che vi sono state molte sovrapposizioni, utili in fase di discussione plenaria. Certe volte ci sarebbe stata la necessità di procedere in modo più sequenziale, ma i tempi erano stretti e l'idea di far ruotare le persone da un gruppo ad un altro gruppo nei giorni successivi è risultata vincente.

In tutti i gruppi è emersa sempre come una costante la questione del copyright a lungo dibattuta durante i tre giorni.

Dall'incontro di Berlingen è emerso chiaramente che i vari Paesi dovranno muoversi attraverso partner forti che siano in grado di coordinarsi a livello europeo da una parte e dall'altra siano in grado di individuare linee di finanziamento per far girare i meccanismi delle unità operative di digitalizzazione. Tra i giocatori forti dell'arena sicuramente le società matematiche nazionali dovranno farsi carico del ruolo di sostenitori e fornitori di risorse, stringendo alleanze e negoziando finanziamenti.

## **Partner di progetto e stakeholder**



I membri si accordano di partecipare alla digitalizzazione di una gran parte della letteratura matematica, direttamente o attraverso convenzioni con altri soggetti, entro un'architettura quadro internazionale. Il progetto avrà rappresentanti europei nelle commissioni che saranno inseriti nel contesto di un'architettura globale.

Tra i soggetti coinvolti vi sono le più importanti biblioteche e istituzioni accademiche e scientifiche di matematica del mondo i quali hanno messo a disposizione strumenti e esperienza ai fini della costruzione di una piattaforma comune.

Per la parte europea i soggetti interessati e che hanno fatto parte di gruppi lavoro entro i meeting e workshop europei e che hanno in qualche modo contribuito alla stesura dei documenti di progetto sono: Technische Universität di Berlino, Sobolev Institute of Mathematics di Novosibirsk, Russia, Danish Technical Knowledge-Center (DTV) Library, Library of University Bielefeld, Oxford University Library Services, SUB Goettingen State and University Library, Mathematical Institute of SANU (Belgrade), Osterreichische Zentralbibliothek fuer Physik, Réseau National des Bibliothèques de Mathématiques (France), ETH-Bibliothek in coordinamento con il Consortium of Swiss Academic Libraries, l'Università degli Studi di Padova (Centro di Ateneo per le Biblioteche. Biblioteca del Seminario Matematico).

Tra gli editori oltre all'AMS e EMS che sono anche editori, vi sono Springer, Cambridge University Press, SIAM, e numerosi progetti già avviati o in corso di avvio tra i quali Cellule MathDoc, Limes, EMANI (Electronic Mathematics Archiving Network Initiative), ERAM (Electronic Research Archive for Mathematics), la banca dati bibliografica europea Zentralblatt Math, la rivista Documenta Mathematica, esempio di pubblicazione accademica alternativa.

Le regole che governano la DML pianificheranno i ruoli, le responsabilità e i benefici dei membri in relazione alle seguenti condizioni:

- ❑ I membri avranno un maggiore accesso ai fondi, nazionali e internazionali al fine di digitalizzare i loro archivi matematici
- ❑ Ai membri non sarà richiesto di trasferire la proprietà intellettuale dei contenuti o i diritti editoriali o anche solo concedere licenze per la rappresentazione di periodici o volumi digitali alla DML, ma sarà invece richiesto loro di creare copie di archivio dei loro contenuti digitali e stabilire sezioni DML sui loro siti Web con link al server DML e agli altri server laddove esistenti
- ❑ I membri saranno d'accordo di pagare una tassa annuale per appartenere alla DML
- ❑ I visitatori che accedono alla DML saranno dirottati sui siti Web dei proprietari dei contenuti aumentando la propria visibilità e laddove esistente un modello di business, aumentando i guadagni (per esempio nelle funzionalità di PoD Print on Demand)
- ❑ DML offrirà strumenti, standard e diagrammi per i flussi di lavoro di digitalizzazione ai membri; perciò i membri non dovranno sviluppare da sé strumenti o predisporre standard, a meno che non si offrano loro stessi come sviluppatori di strumenti e/o procedure e in tal caso sarà necessario integrare tali nuovi processi entro i flussi di lavoro generali incorporandoli entro la DML.

## **Contributi finanziari dall'Unione Europea**

I contributi finanziari al progetto europeo arriveranno dai membri e dall'Unione Europea, sulla base di quanto sarà definito in sede di contratto consortile. Questi contributi finanziari copriranno le spese per il potenziale umano per il disegno architettonico e l'implementazione dell'infrastruttura, le spese generali e di avvio per il trattamento ordinario dei dati e per i software per la scansione, il riconoscimento ottico dei caratteri (OCR), la predisposizione di interfacce di ricerca e specializzate.

Questi contributi copriranno un'ampia massa critica della letteratura europea, compreso il materiale proveniente da progetti volti al mantenimento della propria identità locale.

Il totale del budget previsto per il progetto è di 16 milioni di Euro.

E' subito evidente come i partner entro il Consorzio dovranno garantire delle linee di finanziamento agganciandosi a progetti già esistenti, chiedendo finanziamenti ai governi, o mettendo a disposizione strutture, piattaforme e strumenti software, capitale umano per il lavoro di digitalizzazione e unità di coordinamento e operative.

Partner ideali sono le società matematiche nazionali che dovranno garantire supporto in termini di coordinamento, ma soprattutto nel reperimento di fondi per il funzionamento del progetto entro un modello di sistema distribuito.

Per quanto riguarda l'Europa vi sono esperienze entro alcune università o enti scientifici che possono assicurare piattaforme e strumenti software, già predisposti o in corso di sperimentazione, offrendo un supporto tecnologico a tutti i membri che afferiscono al progetto europeo. In tale dimensione i partecipanti dovranno operare entro un ambiente di test-bed per l'utilizzo degli strumenti software in relazione alla tipologia di documenti da trattare per la digitalizzazione e alle collezioni da costruire.

## **Pianificazione del progetto UE DML**

In analogia con il progetto mondiale, sono state individuate tre fasi di pianificazione del progetto europeo:

### **Fase 1. Disegno: 2 anni**

- ❑ Identificazione dei contenuti
- ❑ Decisione delle priorità
- ❑ Indirizzamento delle questioni del copyright
- ❑ Predisposizione standard tecnici e metadati

### **Fase 2. Implementazione: 3 anni**

- ❑ Set up del database
- ❑ Digitalizzazione grezza del materiale
- ❑ Costruzione di interfacce di ricerca specializzate
- ❑ Costruzione del sistema su modello distribuito

### **Fase 3. Operazioni di sostegno**

- ❑ Archiviazione
- ❑ Migrazione dei formati

I gruppi di lavoro a Berlingen si sono mossi lungo alcune direttrici. Sono stati redatti dei documenti base utili alla composizione del progetto europeo.

I punti trattati da più gruppi hanno riguardato le seguenti tematiche: contenuti, standard tecnici, formati, archiviazione, copyright.

## **I contenuti**

In merito alla questione dei contenuti è indubbio che uno dei problemi maggiori è insito nella selezione dei contenuti da digitalizzare. E' risaputo per esempio che ogni anno arrivano al Mathematical Reviews circa 110.000 item, e che di questi solo 75.000 sono considerati accettati ai fini dell'inclusione. Un problema non secondario sta nello scegliere gli articoli di matematica all'interno di periodici di altre discipline, per esempio economia, fisica, psicologia,...

La situazione per le monografie è ancora più complicata, in quanto la questione dei livelli o granularità informativa è un aspetto assai spinoso che anche i bibliotecari ben conoscono, per non parlare dell'interdisciplinarietà, della questione delle edizioni successive o multiple, dell'aspetto della lingua.

Va sottolineato che prendere decisioni caso per caso è oltremodo costoso come potrebbe risultare assai costoso anche all'interno del processo di digitalizzazione.

Ci si aspetta un supporto prezioso proveniente dai progetti di digitalizzazione nazionali i quali per loro conto dovranno essere in grado di selezionare il materiale maggiormente rappresentativo per quell'area linguistica e geografica.

## Standard tecnici e formati

Uno dei gruppi di lavoro a Berlingen si è occupato di vagliare la questione relativa agli standard tecnici per la digitalizzazione.

Allo scopo si è utilizzato il report finale del DML Working Group sugli standard tecnici presentato da Ulf Rehmann di Bielefeld che faceva parte anche del nostro gruppo di lavoro a Berlingen.

Anche per il progetto europeo sarà essenziale stabilire delle "raccomandazioni" rispetto ai seguenti argomenti specifici:

- ❑ qualità della scansione
- ❑ formati per l'archiviazione
- ❑ convenzioni dei nomi dei file
- ❑ formati di distribuzione
- ❑ unità download
- ❑ modalità tecniche sui server

Per permettere ai molti contributori di fornire materiale digitalizzato di **qualità** sufficientemente alta, è necessario fornire loro strumenti software pubblici per trasformare il materiale nel formato più corretto. Per esempio il **software per il riconoscimento ottico dei caratteri OCR** dovrebbe essere ottimizzato per la lingua originale del manoscritto da trattare, quindi sarebbe auspicabile avere server pubblici per le varie aree linguistiche. Inoltre dovrebbe essere facile per i contributori fornire il materiale scansito con **metadati elementari** come per esempio quelli relativi alla classificazione internazionale per la matematica MSC Mathematics Subject Classification, parole chiave e frasi sulla base degli standard Dublin Core e Open Archivi Iniziative.

Server pubblici di **conversione dei formati** potrebbero anche contribuire al problema dell'archiviazione a lungo termine, considerato che essi offrono solitamente strumenti dinamici utili a questo scopo.

In merito alle **convenzioni dei nomi dei file** si è stabilito che questo argomento dovrà essere affrontato in cooperazione con il gruppo sui metadati. In ogni caso dovrebbe essere garantito un unico e significativo nome per tutti i file, URL stabili per tutti i documenti, uniformità stilistica delle pagine web per tutti i server, possibilmente organizzati utilizzando l'ordinamento governato dallo schema MSC Mathematical Subject Classification 2000 questo anche al fine di poter offrire delle tecniche di accesso uniformi per tutti i documenti.

Il **browsing attraverso tavole** dei contenuti dovrebbe essere condizione desiderabile e dovranno essere anche stabilite le modalità del **downloading delle unità** (pagine singole, range di pagine, ...)

La definizione di **standard per contenuti** che sono già in forma digitale è un problema ben noto, se non proprio ben compreso. Questo richiederà un duro lavoro e una consistente negoziazione. Ma anche il problema apparentemente semplice di decidere il **formato del materiale** scansito è estremamente difficile. Non molto tempo fa, molti avrebbero indicato di usare qualche forma di file TIFF compresso incapsulato in file Adobe PDF. Ma, sebbene PDF abbia al momento un'ampia

disponibilità di strumenti di supporto, per certi sistemi operativi come Unix questo supporto è diventato problematico. Ancora più importante, ci sono formati nuovi, estremamente efficienti per le immagini scansionate, che riducono la dimensione dei file di un fattore da 3 a 8 o più.

Il "garbage text" per esempio derivante dal materiale non riconosciuto dall'OCR come le formule matematiche, dovrebbe essere rimosso. Inoltre si dovrebbero rendere ricercabili sia il formato PDF che il DjVu e dovrebbero essere aggiunti link al MathSciNet e al Zentralblatt a partire dalle referenze bibliografiche estratte in modo automatico dal testo pieno.

## Il formato DjVu<sup>4</sup>

Ci sono molti modi di digitalizzare la *letteratura past* (cioè, letteratura che è già in forma digitale) ma l'unico modo conveniente in termini di costi/benefici è combinare la scansione con una parziale lettura ottica dei caratteri, creando una combinazione di immagini di pagine scansionate e file di testo associati, che rendano possibile la ricerca per parola o stringa. Naturalmente il processo reale è più complesso. Si devono catturare, di solito digitando manualmente, i dati bibliografici rilevanti di ciascun documento; i documenti devono essere analizzati per individuarne le varie parti (articoli, capitoli, ecc.); deve essere effettuata un'attenta correzione dei dati critici nei file di testo.

Le stime del costo di attuazione di queste fasi in un'operazione a grande scala sono variabili, ma una grossa approssimazione è di Euro 2 (o dollari) per pagina risultando, in un costo totale per la digitalizzazione di 50 milioni di pagine, attorno ai 100 milioni di Euro o dollari.

Tra i nuovi formati efficienti il più rappresentativo è DjVu, pronuncia "déjà vu", un formato sviluppato negli AT&T Laboratories, che usa la tecnica delle ondine per una maggiore compressione e un algoritmo progressivo per la decompressione delle immagini, il quale presenta immediatamente un'immagine grossolana, in progressivo miglioramento. I vantaggi sulla tecnologia più tradizionale sono notevoli. E' un formato utilizzato da molteplici siti accademici, commerciali e governativi in tutto il mondo.

I prodotti che implementano DjVu sono attualmente di proprietà e in vendita da Lizard Tech. Come PDF, DjVu richiede un software speciale per vedere le immagini nei browser chiamato DjVuLibre, un open source rilasciato con licenza GPL il quale include il visualizzatore, il plug-in per il browser, il decodificatore o il codificatore e le utilities.

DjVu è un formato web-centrico e una piattaforma software per distribuire documenti e immagini. Il formato DjVu può tranquillamente rimpiazzare i formati PDF, PS, TIFF, JPEG, e GIF nella distribuzione di documenti scansionati, documenti digitali o immagini ad alta risoluzione, utilizzando minori risorse client quindi risultando più veloce e leggero e con un più piacevole layout visivo.

Risulta essere un'alternativa superiore in termini di qualità al PDF e PostScript per i documenti digitali, al TIFF e PDF per i documenti bitonali scansionati, al JPEG per le fotografie e al GIF per le immagini ampie. E' adatto per la distribuzione di documenti a colori scansionati ad alta risoluzione.

La visualizzazione è istantanea e uniforme per tutte le piattaforme in merito a font di caratteri, colori, ecc... Il documento può essere percorso nella sua lunghezza e zoomato senza problemi in ordine di lunghezza o dimensione.

## Archiviazione

L'archiviazione richiede decisioni su quale informazione sarà necessaria in futuro, e tali decisioni devono essere prese in assenza di conoscenze dettagliate. In effetti, attualmente e per molti anni a venire, tali decisioni devono essere prese senza alcuna esperienza al riguardo. Ci sono moltissimi esempi di decisioni scorrette prese negli ultimi venti anni, che si sono concretizzate nella perdita del

---

<sup>4</sup> Per una descrizione tecnica del formato DjVu vedi articolo di Yann LeCun, Léon Bottou, and Patrick Haffner *DjVu: a Compression Technique and Software Platform for Publishing Scanned Documents, Digital Documents and High-Resolution Images on the Web.*  
<http://djvu.sourceforge.net/abstract.html>

lavoro; non c'è ragione per credere che possiamo evitare decisioni scorrette in futuro, sottolinea Ewing nel suo libro bianco.

Selezionare il giusto formato iniziale, eventualmente un formato proprietario, in un ambiente che è in costante cambiamento, per un progetto che dura più di dieci anni, è un lavoro praticamente impossibile. Ciò è strettamente connesso al problema dell'archiviazione, anche se non è la stessa cosa: il formato iniziale indicato per la presentazione può non essere quello giusto per l'archiviazione. Ci sono numerosi schemi semplici per assicurare la "robustezza", compresa la banale tecnica della replicazione per creare copie multiple, proprio come per la carta e considerato che fare copie elettroniche è molto più facile che fare copie su carta questo compensa in parte il costo di eventuali passaggi aggiuntivi.

Dato che i supporti elettronici possono deteriorarsi più rapidamente della carta, ci deve essere un passo in più di replicazione di routine per produrre copie su supporti nuovi, ma va considerato che cambiare formati non è equivalente a fare una copia. Infatti se fare una copia è un'operazione di routine e di facile esecuzione per grandi masse di materiale, cambiare formato richiede interventi speciali, almeno per una frazione di materiale. Anche se solo una piccola parte del materiale necessita di interventi speciali realizzati da personale in possesso di tecniche specifiche, questo può essere enormemente costoso per una grande collezione. Spesso chi tratta con piccole collezioni personali ignora questo punto.

Le difficoltà dipendono dal *vecchio* formato, qualcosa che conosciamo in anticipo, come pure dal *nuovo*, qualcosa che non conosciamo quando creiamo l'archivio. Si può sperare che il problema del formato venga risolto in una maniera simile a quello dell'archiviazione, ovvero cambiando regolarmente formati ogni volta che se ne presentano di nuovi, considerato che la replicazione di routine richiede sempre un cambiamento dei supporti e in considerazione del fatto che una copia si può muovere su qualsiasi supporto sia al momento in uso. Sarà nell'interesse di tutti di rendere certo che al cambiamento dei formati vengono prese decisioni affidabili.

Ma vi è una ragione più sottile che ci dice che il cambiamento di formati non è una soluzione adeguata. Il problema del formato è più di una mera preservazione del formato di un'opera; è decidere quale informazione sull'ambiente in cui un'opera è presentata dovrebbe essere salvata inizialmente, e poi decidere, in ogni stadio successivo dell'archiviazione, quale informazione viene mandata avanti. E' virtualmente impossibile salvare ogni singola informazione sull'ambiente. Per esempio, potremmo far riferimento agli standard ISO per il riconoscimento dei caratteri e assumere determinate convenzioni per i caratteri di capolinea.

## Copyright

La questione copyright è spesso sconosciuta e forse anche sottostimata persino nei progetti di digitalizzazione più importanti, sebbene recentemente si stia incominciando a parlare di diritti di proprietà tra gli addetti ai lavori nelle varie comunità coinvolte in tali progetti. Il copyright è il capitolo centrale di ogni Biblioteca Digitale e questo non dovrebbe mai essere sottovalutato.

La cessione o trasferimento dei diritti dall'autore all'editore si attua attraverso dei contratti, ma si deve tenere in considerazione la questione delle diverse normative nazionali, delle leggi comunitarie (per quanto concerne l'Europa) e dei trattati internazionali. In sostanza digitalizzare un periodico comporta chiedere il permesso a migliaia di autori, o loro eredi, trattare con centinaia di editori, molti dei quali non facilmente identificabili o che non sono più presenti nel mercato. In tutto questo le leggi non aiutano la costruzione di Biblioteche Digitali che dovrebbero essere i templi del patrimonio scientifico da offrire alla comunità intera.

Questo aspetto fa aumentare notevolmente i costi amministrativi del progetto, come ben descrive Clifford Lynch nel suo articolo "*Convertire i vecchi volumi in formato digitale*".

Come suggerisce Ewing, una possibile soluzione al problema del copyright è decidere di includere solo letteratura che è già di dominio pubblico o quella il cui permesso è facilmente ottenibile. Da una stima effettuata a grandi linee risulta però che dei 50 milioni di pagine di matematica da

digitalizzare, oltre il 90% è protetto da copyright, di questo una buona metà richiederebbe una ricerca e conseguente negoziazione sulla questione dei diritti.

Risolvere il problema del copyright ignorandolo perciò richiede un forte compromesso rispetto all'obiettivo originale del progetto di rendere accessibile la maggior parte della letteratura matematica.

E' quindi necessario che si muovano i governi nel tutelare le biblioteche digitali. I governi dovrebbero comprendere che gli interessi delle comunità scientifiche e della popolazione globale nell'accesso ai contenuti del patrimonio dell'umanità prevalgono sugli interessi dei singoli che, molto spesso, sono piuttosto gli interessi di alcune lobbyies di mercato.

## **Il modello economico**

Nel modello economico europeo andranno focalizzati una serie di aspetti tra cui i presupposti base di scenario, gli attori in gioco o *players*, la struttura generale e i finanziamenti.

Al fine di salvaguardare la letteratura detta di *legacy* sarà importante creare e pianificare una struttura che possa attrarre i fornitori e possessori di contenuto i quali attualmente detengono la tradizionale letteratura matematica. Questi detentori includono società, organizzazioni o editori no-profit e editori commerciali. La maggioranza degli editori sono piuttosto scettici nel trasferimento della proprietà dei materiali o nell'assegnazione del copyright elettronico a favore delle biblioteche digitali. Inoltre gli editori commerciali preferirebbero ospitare i contenuti sui loro propri siti di modo da integrare questi materiali con i contenuti digitali esistenti, per aumentare la loro visibilità in termini di business.

Non tutti gli editori commerciali sono disposti inoltre a digitalizzare materiale considerato letteratura di *legacy* in quanto non si assumono i costi che possono anche essere rilevanti a fronte di scarsi profitti. Sarà necessario quindi trovare finanziamenti esterni per la digitalizzazione di questo patrimonio storico.

Le biblioteche possono essere l'agente naturale per archiviare questi materiali.

## **Aspettative e previsioni di progetto**

1. Ci si aspetta prevedibilmente che almeno 8 milioni di pagine saranno digitalizzate entro il progetto europeo, in stretta dipendenza dai trattamenti che effettivamente saranno applicati al materiale a stampa
2. Saranno sviluppati strumenti, programmi e standard, che renderanno facile per la comunità matematica mondiale continuare il programma di digitalizzazione nel futuro e che si occuperanno di supportare le operazioni di mantenimento e aggiornamento corrente
3. La creazione di una biblioteca digitale per la matematica che rappresenta una parte significativa del patrimonio culturale europeo, da una parte migliorerà sensibilmente le condizioni di lavoro della comunità scientifica, ma d'altro canto, ci si aspetta che questo progetto avrà nel 21.mo secolo un forte impatto sul modo in cui la matematica è fatta e usata. In particolare ci si aspetta che vengano messe a punto, a beneficio dei ricercatori, tecnologie d'interfaccia e di gestione della conoscenza. Le applicazioni effettivamente interessanti risulteranno dai programmi di digitalizzazione per il riconoscimento ottico delle formule o delle strutture, dell'interazione senza giunture fra i dati delle biblioteche ed i sistemi di computer algebra, delle macchine di ricerca intelligenti, di nuovi formati ecc...
4. La digitalizzazione totale della letteratura di una disciplina nella sua interezza potrebbe servire come modello di riferimento per altre scienze o altre aree disciplinari. Gli strumenti e i metodi messi a punto e potrebbero essere adattati ed utilizzati per aumentare l'infrastruttura scientifica e sviluppare così un percorso importante verso una società della conoscenza.

## Il progetto italiano ItalDML

Nell'ottica di quanto esposto sopra, entro un quadro progettuale europeo coordinato dall'European Mathematical Society, finanziato dall'Unione Europea entro il 6. Programma Quadro, e in asse coordinata con il progetto statunitense capeggiato dalla Cornell University, appare quanto mai indispensabile comparire sulla scena offrendo un progetto ItalDML per lo sviluppo di un archivio accessibile delle pubblicazioni matematiche italiane.

Questo presuppone un forte coinvolgimento di partner italiani da individuare come membri del consorzio europeo, partner che supportino un'attività di coordinamento nazionale e che dovranno garantire risorse economiche, finanziarie e di personale. Tra questi partner parrebbe logico che l'UMI Unione Matematica Italiana si candidasse come partner ufficiale a rappresentare l'Italia e che si proponesse quale attore nello scenario europeo offrendo supporto ad un'unità di coordinamento che potrebbe essere individuata entro il Sistema Bibliotecario di Ateneo dell'Università degli Studi di Padova, Biblioteca del Seminario Matematico la quale ha presenziato e attivamente partecipato ai lavori di Berlingen e possiede le competenze adeguate per svolgere attività di coordinamento.

Accanto all'unità di coordinamento dovrebbero essere attivate delle unità operative individuate laddove vi siano già dei centri di digitalizzazione operativi.

Sarebbe quindi opportuno procedere con la stesura del progetto *Italian Digital Math Library* con i principali requisiti in linea con quelli del progetto europeo nel quale esso si incardinerebbe:

- ❑ archivio aperto e accessibile a livello internazionale su web
- ❑ l'accesso geografico dovrebbe essere facilitato
- ❑ le copie distribuite dovrebbero garantire la sicurezza dei dati
- ❑ ItalDML dovrebbe far parte di una rete più ampia (europea) che offre le pubblicazioni digitali matematiche
- ❑ lo sviluppo di ItalDML dovrebbe integrarsi con la cooperazione internazionale
- ❑ coordinamento con la base dati bibliografica europea Zentralblatt fur Math per link a doppio senso, dalla banca dati al testo pieno e vice-versa e per il reference linking dalle referenze bibliografiche di ciascun articolo scansato
- ❑ predisposizione del modello tecnologico-organizzativo in merito a piattaforme e strumenti software (in raccordo con il gruppo europeo)
- ❑ definizione di fasi operative necessarie al raggiungimento degli obiettivi (da stabilire entro il progetto): numero di pagine da digitalizzare, presumibilmente sull'ordine di 1 milione di pagine da suddiversi in 4-6 unità operative (250.000-170.000 ciascuna).
- ❑ individuazione delle linee di finanziamento: circa 2 milioni di Euro per la digitalizzazione di 1 milione di pagine
- ❑ tempi: programmazione e calendario

L'unità di coordinamento dovrà occuparsi di:

- ❑ individuare, entro il territorio italiano, quali progetti di digitalizzazione esistono e sono già attivi, per opere o periodici matematici
- ❑ creare un registro delle opere digitalizzate (in raccordo con un registro europeo laddove esistente)
- ❑ definire chi digitalizza cosa (in raccordo con la commissione selezione)
- ❑ stilare un documento per le politiche di copyright e IPR in ottemperanza alle leggi nazionali, europee e ai trattati internazionali
- ❑ distribuire gli standard tecnici per una digitalizzazione uniforme e assicurarsi che siano rispettati (in connessione con il gruppo europeo)

- ❑ stilare il protocollo per l'utilizzo dei formati di distribuzione (in connessione con il gruppo europeo)
- ❑ coordinarsi a livello europeo in merito ai formati e standard di metadati e per i formati di archiviazione

La gran massa di lavoro per la digitalizzazione di tutte le pubblicazioni importanti di ambito italiano, potrebbe condurre, in una fase successiva, alla conclusione che potrebbe essere auspicabile l'integrazione di partner addizionali nel lavoro di digitalizzazione.

Per sostenere questo progetto, si deve trovare il modo di coprire i costi potenzialmente elevati dell'aggiornamento dei formati nel futuro, così come di prendere decisioni ragionevoli su quali collezioni digitalizzare. Mantenendo collezioni su molti siti, ciascuno con interessi professionali o finanziari al materiale, si crea la sicurezza che un vasto gruppo sarà motivato a dividerne i costi. Nondimeno, questi sono problemi che si estendono per lunghi periodi di tempo, andando spesso oltre la durata della carriera delle persone coinvolte: ci deve essere quindi un meccanismo per garantire che i problemi dell'archiviazione, dei formati e degli standard, come pure il lavoro di selezione dei contenuti, vengano affrontati su una base di continuità.

Chiudo riprendendo ancora una volta Ewing che cita in epigrafe al suo libro bianco una frase di Henry David Thoreau "*Se avete costruito castelli in aria, il vostro lavoro non deve essere perso. Quelli sono là dove dovrebbero essere. Ora mettete le fondamenta sotto di essi*"<sup>5</sup>

## Bibliografia di riferimento

Ewing, John (2002, March). Twenty centuries of mathematics: Digitizing and disseminating the past mathematical literature. Version 12.3.  
[http://www.ams.org/ewing/Twenty\\_centuries.pdf](http://www.ams.org/ewing/Twenty_centuries.pdf)

Kuperberg Greg. Scholarly mathematical communications at a crossroads.  
 Nieuw Arch. Wisk. (5) 3 (2002), 262-264,  
<http://front.math.ucdavis.edu/math.HO/0210144>

Branin, Joseph J. and Case, Mary. (1998, April) Reforming scholarly publishing in the sciences: a librarian perspective. Notices of the AMS, 45(4), 475-486  
<http://www.ams.org/notices/199804/branin.pdf>

A Digital Mathematics Library: Expression of Interest to the European Union submitted by the European Mathematical Society  
<http://www.library.cornell.edu/dmlib/DML-EoI-draft6.pdf>

Lynch, Clifford  
 The Battle to Define the Future of the Book in the Digital World,  
[http://firstmonday.org/issues/issue6\\_6/lynch/index.html](http://firstmonday.org/issues/issue6_6/lynch/index.html)

Conference on retrodigitization of mathematical journals and their integration into searchable digital libraries. Institute for Experimental Mathematics, Essen, August 1-3, 2000. (Presentations). Institut für Experimentelle Mathematik. (IEM). University of Essen.  
<http://www.exp-math.uni-essen.de/algebra/veranstaltungen/retro.htm>

---

<sup>5</sup> Henry David Thoreau, Walden, Chap.18



Workshop on Linking and Searching in Distributed Digital Libraries. University of Michigan, Ann Arbor, University Library March 18 - 20, 2002. (Presentations). Institut für Experimentelle Mathematik (IEM). University of Essen.

<http://www.exp-math.uni-essen.de/algebra/veranstaltungen/program4.htm>

Digital Library Forum, Institute of Museum and Library Services (IMLS). (2001, Nov. 6). A Framework of Guidance for Building Good Digital Collections.

<http://www.ims.gov/pubs/forumframework.htm>

Garrett, John, ed. ISO Archiving Standards - Overview. NASA/Science Office of Standards and Technology (NOST).

<http://ssdoo.gsfc.nasa.gov/nost/isoas/>