

1.2

la documentazione culturale online

DOCUMENTAZIONE BIBLIOGRAFICA, ARCHIVISTICA E DEI BENI CULTURALI

prof. PIERLUIGI FELICIATI

da 2022/23

Obiettivi e strumenti per comunicare beni culturali sul web

Il quinto principio di qualità del [manuale MINERVA](#) del 2004 affermava che un sito web culturale di qualità deve «selezionare, digitalizzare, indicizzare, presentare e controllare i contenuti per creare un sito Web efficace per tutti gli utenti» .

Le **presentazioni di contenuti culturali** realizzate in digitale e destinate al web assumono sempre più spesso una **rilevanza istituzionale** e un forte **profilo pubblico** per i servizi di musei, archivi, biblioteche e istituzioni culturali.

Una **mostra digitale** non sostituisce una mostra reale, come un **catalogo digitale** non completa la tutela e un **inventario d'archivio online** non soddisfa tutti gli utenti, ma tutti questi hanno il vantaggio di essere disponibili **senza tempo e senza luogo, per tutti** (prima, durante e dopo la fruizione reale, se possibile, o comunque pronti in caso di emergenza).

l'utente del web culturale

Sempre secondo il *Manuale MINERVA*, **l'utente del web culturale** è:

[...] colui, professionista o no, specialista o no, che utilizza in modo casuale, saltuario, finalizzato, sistematico l'applicazione web culturale pubblica.

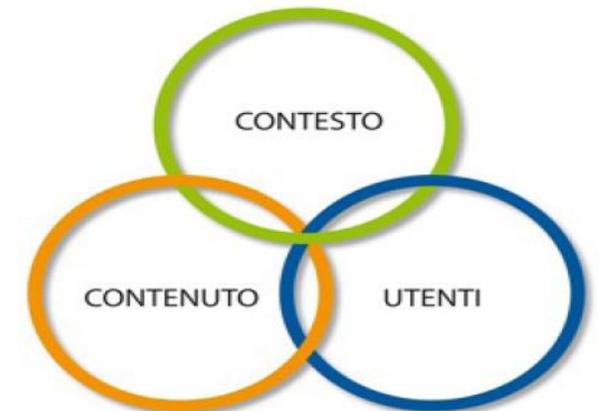
Egli è portatore di esigenze estremamente variabili che dipendono sia dal proprio profilo culturale, sia dalle proprie aspirazioni di crescita culturale, sia infine dalle proprie curiosità, anche momentanee.

Edificare informazione

Ogni **presentazione di contenuti virtuale** dovrà poggiare su un'**architettura** (dal termine inglese *information architecture*), che costituisce la **struttura organizzativa logica e semantica** delle informazioni, dei contenuti, dei processi e delle funzionalità del progetto.

L'architettura delle informazioni è il cuore di un qualsiasi progetto di *interaction design*. Integrando informazioni e processi, svolge un ruolo chiave nel definire il reale grado di fruibilità e di usabilità da parte dell'utente finale. Comprende **l'analisi, la scelta e la progettazione degli strumenti tecnici e culturali per l'organizzazione, la catalogazione, la ricerca, la navigazione e la presentazione di contenuti e dati.**

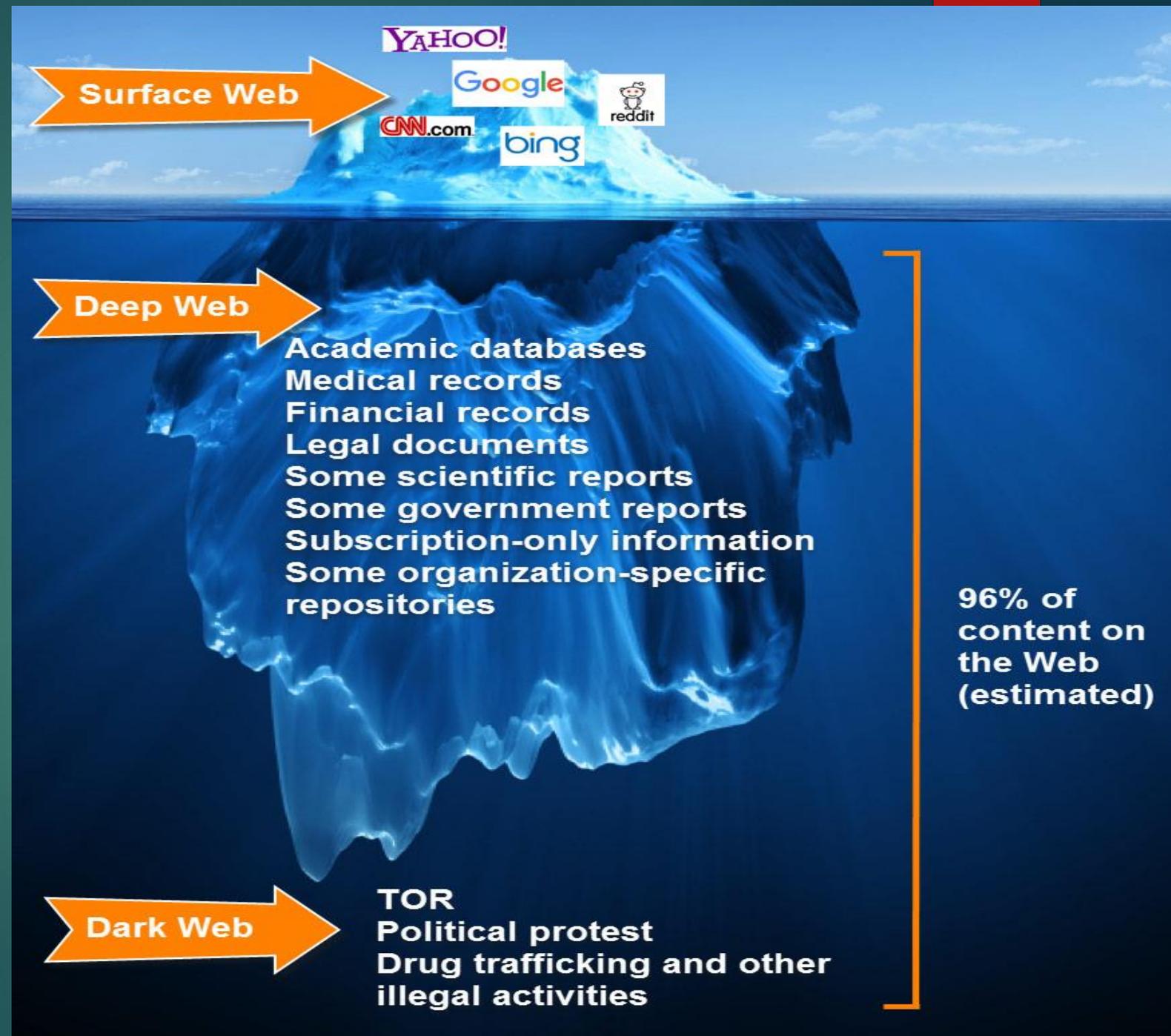
L'architettura dell'informazione rappresenta lo spazio virtuale all'interno del quale si collocano i progetti di valorizzazione e diffusione della conoscenza che abbiano come obiettivo quello di mettere in relazione utenti, contesti e contenuti, anche con finalità di business, attraverso un sistema o un'applicazione informatica.



Surface web (siti web html)

Deep web (database)

Dark web (criptato)



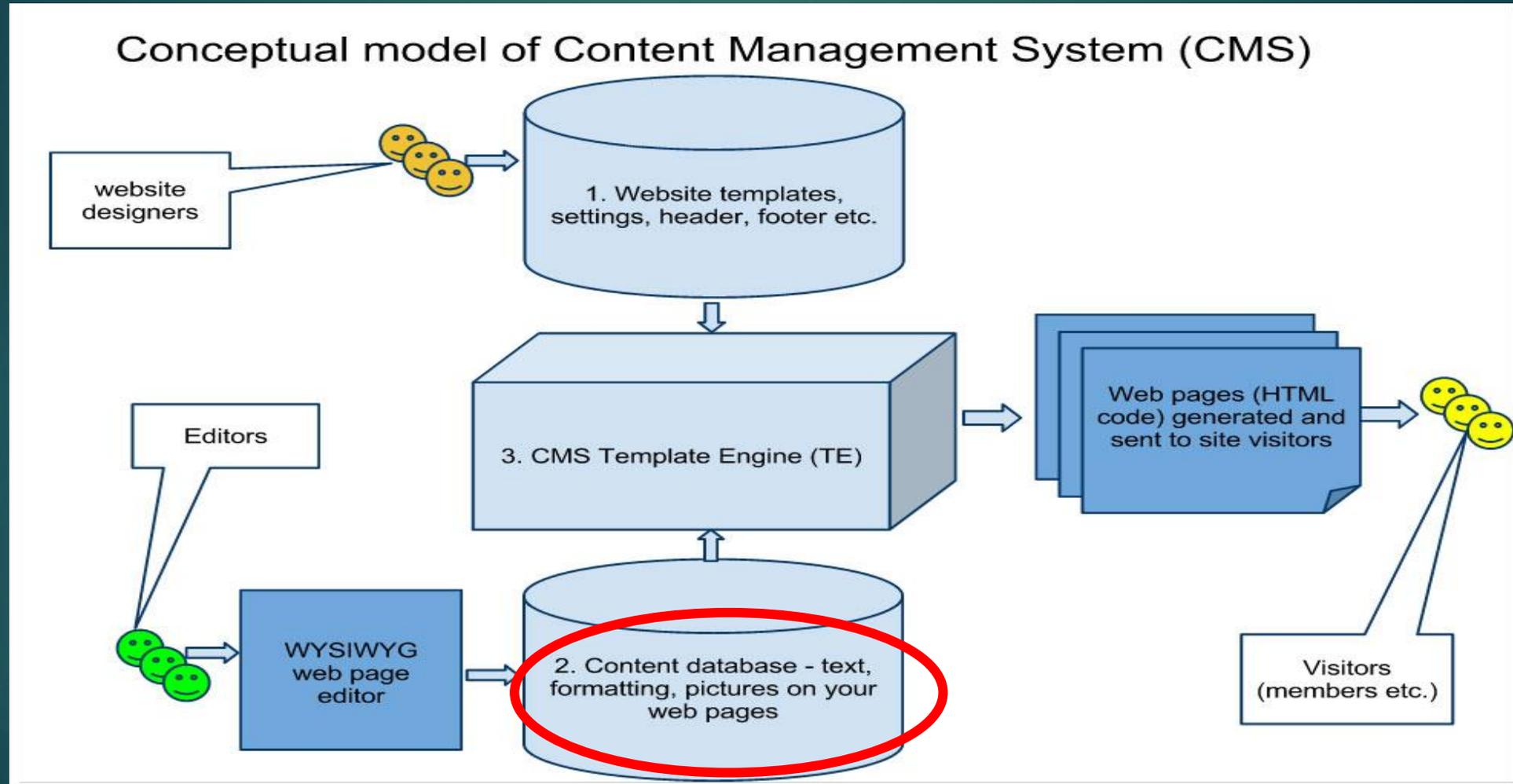
Cosa è un database (in 1 slide)

un **database** è una **raccolta di dati permanenti, gestita da un elaboratore** e suddivisa in:

- **Metadati o schema:** definizioni che ci dicono quali regole valgono sui dati, quali valori possono essere validi (vincoli di integrità), come i dati sono strutturati e collegati tra loro. Lo schema può cambiare nel tempo, non ha alcun vincolo con i programmi che accedono al database e deve essere definito prima dei dati.
- **Dati:** rappresentazione (astrazione) dei fatti, ma può anche includere previsioni, supposizioni ed ipotesi conformi ai Metadati (o schema).

I dati sono organizzati in **insiemi omogenei** (stessa struttura) e sono collegabili con altri insiemi. Il **modello dei dati** è un insieme di astrazioni che descrivono le associazioni tra i dati.

Dai contenuti strutturati al sito web: i CMS



Il catalogo come database

Il **catalogo online** è una *collezione* di rappresentazioni standardizzate di oggetti bibliografici e dei loro contesti.

Tipicamente, nel Web, assume la forma di una **banca dati**, *nascondendo* le rappresentazioni dietro alla mediazione di una *query* (sul *deep web*), OPAC, e a scelte grafiche per rendere i risultati usabili (e talvolta anche **riusabili**, ad esempio per le nostre bibliografie, vedi Zotero).

Il catalogo digitale può assumere anche la forma di **set di dati**, disponibile indirettamente e solo per le macchine (*open data*)

I portali culturali

Nell'ultimi decenni si è assistito ad una vera e propria rivoluzione nel campo dello scambio e della diffusione delle informazioni. Al fine di strutturare in maniera comprensibile il flusso di dati inseriti – spesso non contestualizzati – è nata l'idea del “**portale culturale**”.

Il portale è stato definito come un: **prodotto on-line che svolge la funzione di punto privilegiato di accesso al Web per gli utenti e che fornisce loro risorse informative, servizi di comunicazione personale e strumenti con cui localizzare e raggiungere i contenuti e i servizi on-line di cui hanno comunemente bisogno.**

Un portale **aggrega contenuti** sulla base di un modello di metadati condiviso e **aggiunge servizi** rispetto a quelli offerti dai suoi *content provider*, fornitori di contenuti (oppure è uno spreco).

I portali culturali: Europeana



Europeana è la **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, risorse sonore, mappe, manoscritti ed archivi. La logica è quella di Culturaitalia: pubblicare metadati sintetici, anteprime dei contenuti e rinvio alla fonte. Non una digital library, quindi!

Europeana è un progetto della Commissione Europea ed ha l'intento di convergere verso un **modello digitale di rappresentazione della ricca diversità culturale** dell'Europa (cfr. le dichiarazioni di Manuel Barroso sul portale: ponte fra "le tecnologie della comunicazione e il nostro ricco patrimonio culturale").

Europeana: come funziona?



I diversi *content providers* «espongono» i metadati relativi ai propri contenuti esportandoli dai database e adattandoli al *data model* di Europeana, che li raccoglie e pubblica.

Ogni «scheda» contiene così un link al sito di origine, così che gli utenti possano approfondire ciò che trovano nel portale andando alla fonte.

Il problema principale è che le fonti sono talvolta altri progetti quindi i contesti culturali si possono perdere

I portali culturali: **Culturaitalia**



cultura
italia UN PATRIMONIO
DA ESPORARE

Culturaitalia è un portale promosso dal Mibact-ICCU.

Mira all'aggregazione e alla diffusione in rete di oggetti culturali digitalizzati provenienti dal patrimonio culturale italiano.

Offre agli utenti diversi contenuti redazionali (itinerari tematici, articoli, focus, eventi, rubriche) che valorizzano l'ampio **catalogo di metadati descrittivi** messi a disposizione da diversi enti di interesse culturale come, ad esempio, tutte le opere digitalizzate di archivi e collezioni private, l'OPAC SBN e le banche dati regionali.

Scopo di **Culturaitalia** è dunque quello di consentire una maggiore fruibilità delle **risorse culturali digitali online**, favorendo promozione del patrimonio italiano e il turismo culturale; mira inoltre ad una valorizzazione in chiave europea, stimolando di fatto i diversi enti portatori di interesse.

I portali culturali: **Culturaitalia**

cultura
italia UN PATRIMONIO
DA ESPORARE

Culturaitalia parte dai risultati raggiunti dall'attività di due grandi progetti europei:

MINERVA

a cui hanno fatto seguito due estensioni denominate MINERVAPlus e MINERVAeC,

e **MICHAEL** (*Multilingual Inventory of Cultural Heritage in Europe*), con la sua estensione MICHAELPlus.

Il portale **permette l'adesione** attraverso differenti canali di immissione: i contenuti delle banche dati possono essere messi a disposizione attraverso i **feed RSS, un formato per la distribuzione di contenuti web**; più precisamente si tratta di uno scambio di file con estensione XML che rende possibile il trasferimento dei contenuti. È inoltre possibile segnalare le risorse digitali utilizzando la funzione **“segnala al portale”**, nonché comunicare con la redazione di Culturaitalia per l'invio di materiale redazionale o multimediale.

I portali culturali: Internet culturale



Il portale [Internet Culturale. Cataloghi e collezioni digitali delle biblioteche italiane](#), affidato all'Istituto centrale per il catalogo unico delle biblioteche italiane ICCU, nasce negli anni 2000 (ancora la legge speciale del 2003) per creare un accesso unico e integrato al patrimonio delle biblioteche italiane attraverso i Cataloghi e attraverso le Collezioni digitali delle biblioteche (**Biblioteca Digitale Italiana**).

«Obiettivo primario è promuovere la conoscenza del patrimonio bibliografico e documentario offrendo anche approfondimenti culturali sulle raccolte librerie attraverso risorse multimediali, dedicati alla cultura letteraria, scientifica, musicale. ecc.»; come aggregatore tematico italiano di contenuti digitali aspira ad accogliere i risultati di queste attività realizzate dalle biblioteche di qualunque tipo.

Biblioteca digitale

Cosa è una biblioteca digitale (digital library)?

- una biblioteca immateriale, in cui vengono conservati e resi disponibili esclusivamente **documenti digitali**, siano essi nativi digitali o convertiti da originali cartacei, gestiti e catalogati elettronicamente (Metitieri, Ridi 2003).
- Uno spazio in cui mettere insieme **collezione, servizi e persone** a supporto dell'intero ciclo di vita della creazione, uso, conservazione di dati, informazione e conoscenza (*Santa Fe Planning Workshop on Distributed Knowledge Work Environments: Digital Libraries*, 1997)
- Le biblioteche digitali sono conversazioni (**Manifesto AIB sulle biblioteche digitali**, <http://www.aib.it/aib/cg/gbdigd05a.htm3>)

Biblioteca digitale e catalogo

Se assumiamo queste definizioni di **Digital Library** (*servizio + utenti + risorse*), dovremmo includere però anche gli le bibliografie sul Web, i cataloghi di oggetti culturali, i sistemi archivistici, gli "archivi digitali", addirittura **tutti i buoni siti web interattivi...**

Una distinzione utile può essere se la DL include:

- **una collezione di oggetti digitali** (riproduzioni o no) autoconsistenti, oltre alle loro rappresentazioni
- **solo rappresentazioni** (come un OPAC o una bibliografia, che rinvia a risorse digitali esterne)

La riproduzione digitale

La grande versatilità e potenza delle tecnologie digitali ci consente facilmente di produrre **versioni digitali di oggetti della realtà** (documenti, opere, libri...). Questa operazione si chiama **digitalizzazione**.

Non è pensabile però di **riprodurre tutto il reale** e neanche **solo i beni culturali**, specie in un paese come l'Italia. Intanto, per la loro quantità, poi per lo sforzo che richiede creare copie che siano:

- **trovabili e identificabili** in rete
- dotate di **descrizioni/metadati comprensibili** che ne restituiscano i contesti (culturale, geografico, cronologico, relativo all'origine...)
- agilmente **elaborabili e riproducibili** con strumenti domestici.

La riproduzione digitale: il PND

Le misure di ripristino e rinnovamento del patrimonio fisico culturale saranno accompagnate da un programma di digitalizzazione volto a virtualizzare con approccio standard e ispirato alle migliori pratiche internazionali il patrimonio culturale e turistico italiano.

In questo modo, da un lato si garantirà un accesso universale alle opere d'arte e dall'altro si abiliteranno iniziative di approfondimento e di divulgazione innovative. Verrà potenziata la piattaforma web centrale del turismo italiano che, su ispirazione delle migliori pratiche messe in atto da altri paesi, funga da volano per una comunicazione di qualità del patrimonio e dell'offerta del nostro Paese e da strumento di aggregazione delle informazioni e dei servizi necessari all'incontro della domanda-offerta del turismo in Italia.

Piano Nazionale di Ripresa e Resilienza (PNRR), p.109

La riproduzione digitale: il PND

Gli interventi sul patrimonio “fisico” saranno accompagnati da un importante sforzo per la digitalizzazione di quanto custodito in musei, archivi, biblioteche e luoghi della cultura, così da consentire a cittadini e operatori di settore di esplorare nuove forme di fruizione del patrimonio culturale e di avere un più semplice ed efficace rapporto con la pubblica amministrazione.

Una infrastruttura digitale nazionale raccoglierà, integrerà e conserverà le risorse digitali, rendendole disponibili per la fruizione pubblica attraverso piattaforme dedicate. Sarà inoltre sostenuta la creazione di nuovi contenuti culturali e lo sviluppo di servizi digitali ad alto valore aggiunto da parte di imprese culturali/creative e start-up innovative, con l'obiettivo finale di stimolare un'economia basata sulla circolazione della conoscenza.

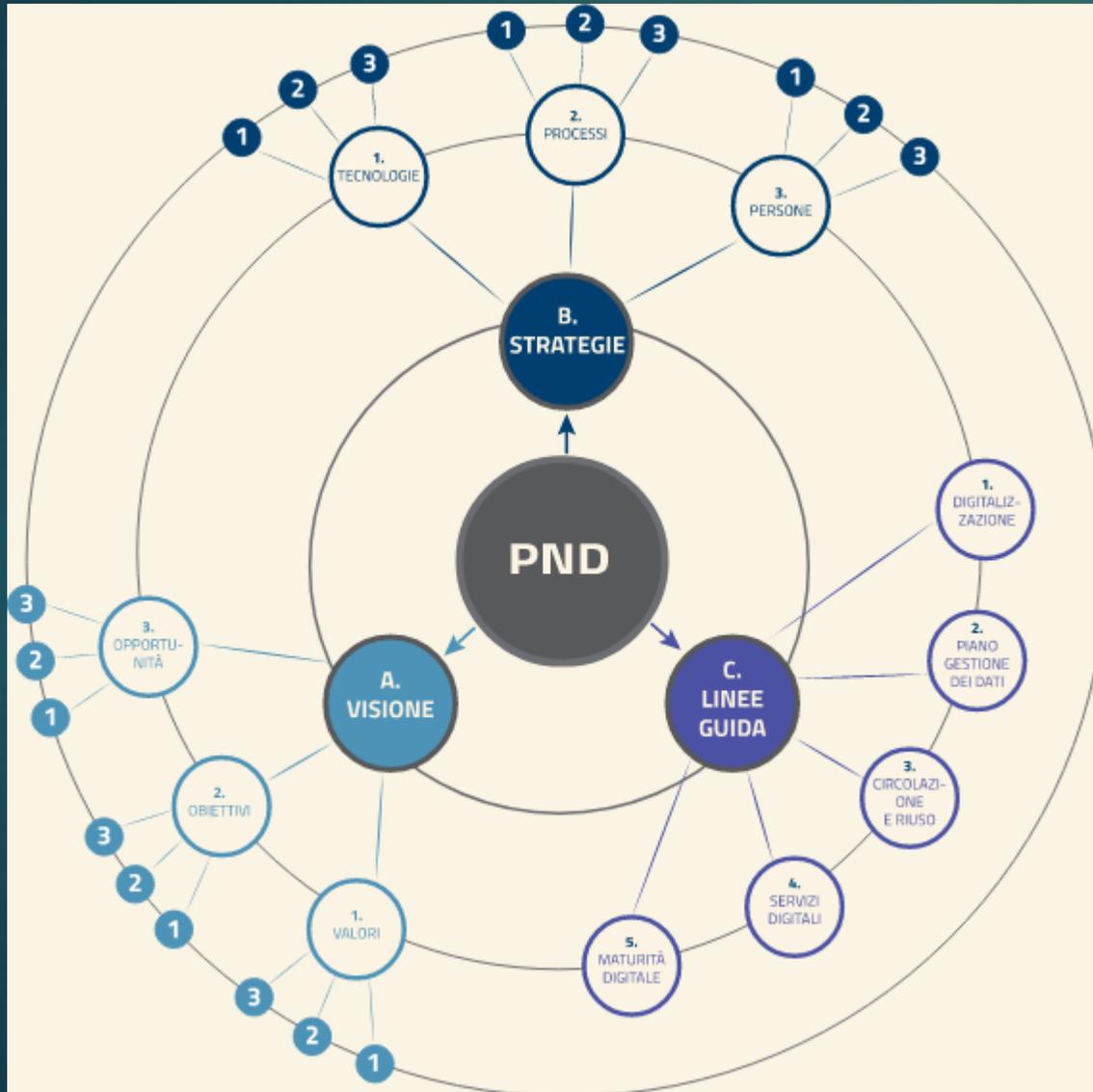
La riproduzione digitale: il PND

Il PND – Piano Nazionale di Digitalizzazione – è il documento teso a **creare il contesto strategico** (intellettuale e professionale) di riferimento per la **realizzazione degli obiettivi del PNRR** relativamente all'investimento M1C3 1.1. **“Strategie e piattaforme digitali per il patrimonio culturale”**.

Il PND è stato redatto dall'Istituto centrale per la digitalizzazione del patrimoni culturale - Digita Library del Ministero della Cultura.

- 18 maggio 2022 rilascio bozza di consultazione
- 15 giugno 2022 chiusura della consultazione
- 30 giugno 2022 pubblicazione del PND

La riproduzione digitale: il PND



Il PND si articola in tre sezioni:

1. **VISIONE**: dove si prefigura la trasformazione e le opportunità di cambiamento, indicando gli obiettivi a lungo termine
2. **STRATEGIA**: sezione nella quale si definisce il percorso per implementare e conseguire gli obiettivi
3. **LINEE GUIDA**: strumenti operativi che supportano la pianificazione e l'esecuzione delle attività legate alla digitalizzazione del patrimonio e alla trasformazione digitale dei luoghi e degli istituti della cultura

La riproduzione digitale: il PND

	VALORI	OBIETTIVI	OPPORTUNITÀ	
TRAIETTORIE DI CAMBIAMENTO	TRAIETTORIA 1 Un patrimonio dai confini aperti	Contesti come paesaggi culturali	Ampliare le forme di accesso al patrimonio culturale	Estensione del patrimonio culturale per nuovi pubblici
	TRAIETTORIA 2 Il digitale come ambiente	Patrimonio culturale digitale	Digitalizzare per operare una trasformazione digitale	Processi per il design di nuovi servizi
	TRAIETTORIA 3 Dagli oggetti alle relazioni	Il capitale semantico delle relazioni	Abilitare ecosistemi interdipendenti	Modelli di conoscenza per nuove organizzazioni

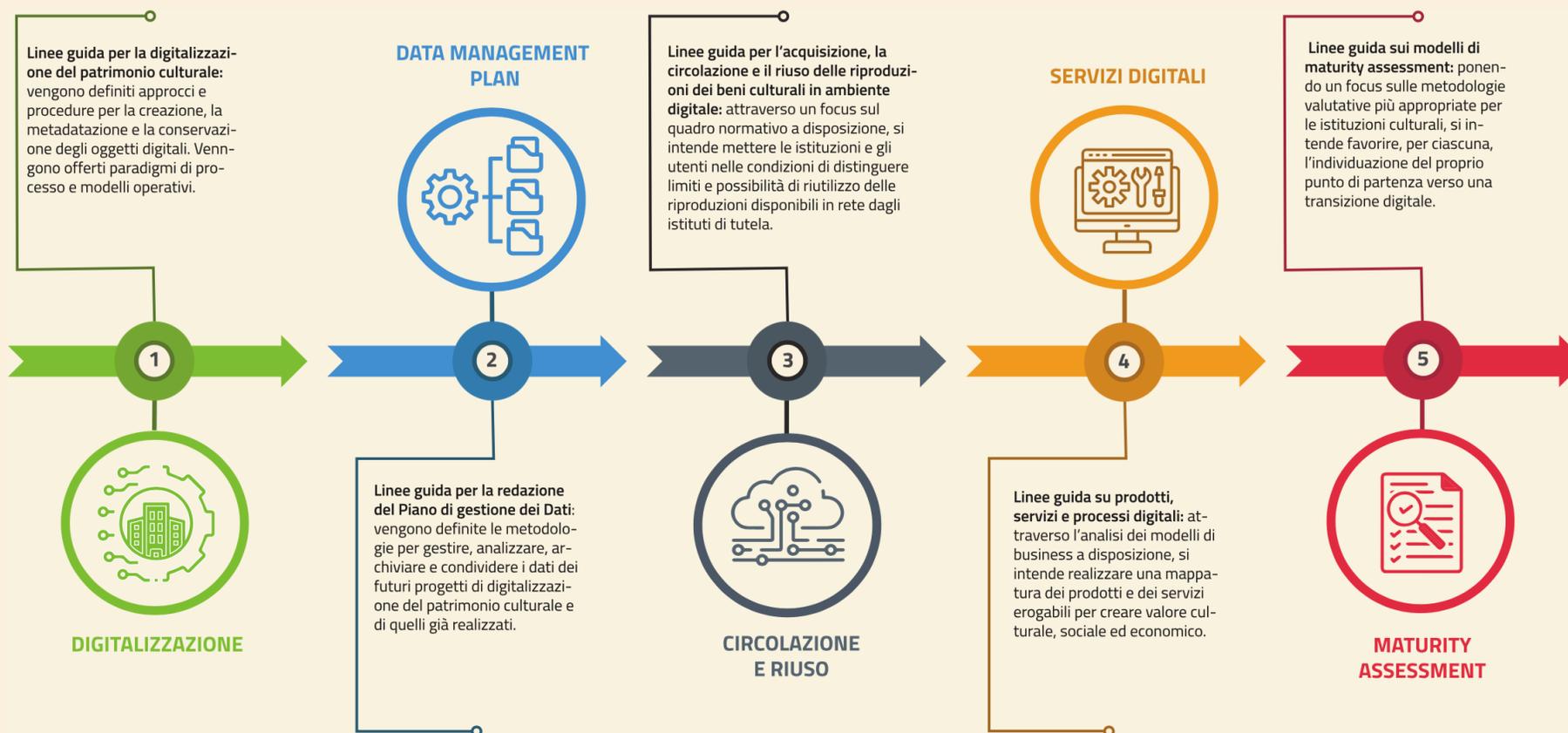
Figura 2. Schema dei contenuti della sezione Visione del PND

La riproduzione digitale: il PND

La seconda parte, **STRATEGIA**, è a sua volta divisa in tre parti

1. **Tecnologie abilitanti** (Infrastruttura nazionale dei dati del patrimonio culturale, Sistema di certificazione dell'identità digitale dei beni culturali, user-centered design)
2. **Processi** (Digitalizzazione e ciclo di vita della risorsa digitale, Politiche di accesso e riuso, Design dei servizi e modelli per la creazione del valore)
3. **Persone** (Formazione e crescita delle competenze, Disseminazione culturale e condivisione sociale, Co-creazione e crowdsourcing)

La riproduzione digitale: il PND



La terza parte, **LINEE GUIDA**, è articolata così:

La riproduzione digitale: il PND

Sono insomma previste:

- › Linee guida per la digitalizzazione del patrimonio culturale
- › Linee guida per la redazione del piano di gestione dei dati
- › Linee guida per l'acquisizione, la circolazione e il riuso delle riproduzioni dei beni culturali in ambiente digitale
- › Linee guida per la classificazione di prodotti e servizi digitali, processi e modelli di gestione
- › Introduzione alla metodologia per la valutazione della maturità digitale degli istituti culturali

La riproduzione digitale: il PND

Per concludere:

Nei prossimi anni saranno disponibili ingenti fondi per la digitalizzazione del patrimonio culturale (sempre che la crisi dovuta alla guerra e alla pandemia non recuperi questi fondi per misure più urgenti).

Saremo tutti chiamati a partecipare. Sia sul piano strategico e di coordinamento, sia a livello operativo. Per ora le università non sono state coinvolte direttamente.

Il piano e le linee guida li trovate in [questo sito](#), *DocsiItalia*.

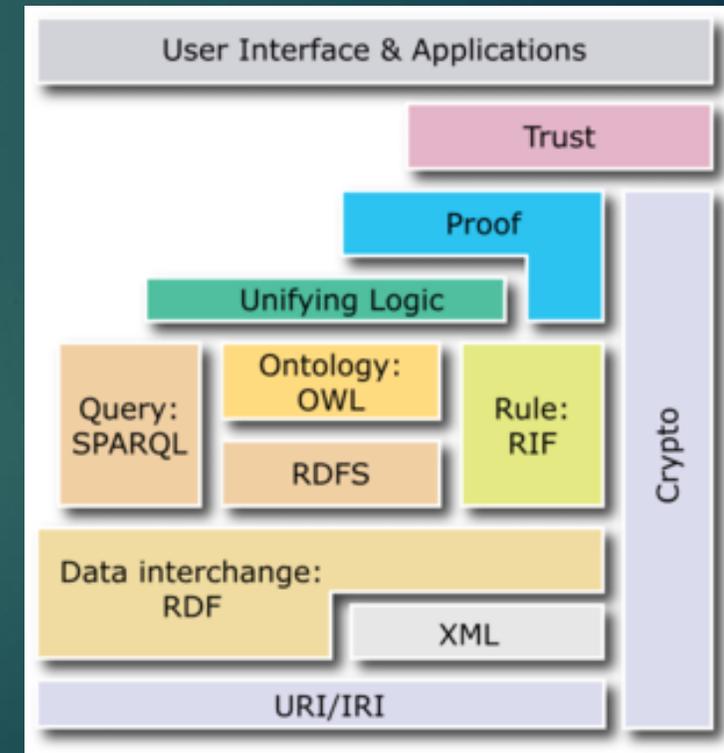
Il web da arcipelago di pagine a galassia di dati

Il Web delle origini è cambiato: non è più una rete fatta di **isole/siti** separati e indipendenti, al massimo *linkati*.

Ogni **oggetto di informazione testuale e multimediale** (un post, un'immagine, una news, una scheda scientifica, un articolo) agisce in modo **indipendente, dinamico, è usabile e ri-usabile, condiviso, visibile potenzialmente ovunque.**

L'*infosfera* è una **galassia di oggetti digitali.**

Questo fenomeno **mette in crisi** l'idea originaria del **sito auto-sufficiente** e l'**originale idea di uso/fruizione univoca:** si attiva infatti un ciclo di vita della conoscenza aperto, senza punti di arrivo, non ci sono più **utenti finali** ma **comunità complesse e dinamiche** che **creano, interagiscono, arricchiscono e modificano contenuti.**



I livelli del Web 3.0

Oltre il sito web: il web 3.0 o web of data

Alla fine, fin dalle sue origini, il Web è un immenso catalogo di **informazioni multimediali**, con strumenti molto potenti per navigare, cercare, usare e riusare le risorse.

Convive però con le risorse informative una quantità ancora maggiore di **dati**, molto spesso non direttamente accessibili (**deep web**).

La **differenza tra un dato e un'informazione** in ambiente di rete consiste soprattutto su **chi e cosa possono farne qualcosa**: i primi sono soprattutto per i software, le seconde per gli utenti umani.

La sfida attuale è affiancare alle informazioni un **Web of data**, che strutturi le informazioni in modo da renderle dati *comprensibili* anche ai software e **le porti in superficie per renderle ricercabili**.

Oltre il sito web: il Web 3.0

Il **Semantic Web** si basa su una architettura a livelli e l'intera comunità scientifica sta investendo molte energie per implementarlo

<http://www.semanticweb.org/>

La sfida del semantic web è fornire **un linguaggio e degli strumenti** per esprimere dati e regole per ragionare sui dati, con la condivisione aperta e decodificabile di qualunque sistema di rappresentazione della conoscenza. I pilastri sono:

- **XML** (con *namespace* e XMLS - *xmlschema*) consente di dare alle risorse informative una **struttura esplicita e decodificabile**.
- **RDF**, uno schema XML, si usa per **esprimere il significato**, asserendo cioè che alcuni particolari entità hanno delle proprietà (p.es. essere *autore-di*, *costruito-da*, *nato-il...*) collegate in modo significativo con altre entità.

Oltre il sito web: le ontologie

30

Un componente necessario per il web semantico, oltre a XML e RDF è l'**Ontology Vocabulary** (*livello ontologico*), inteso come il contenitore che definisce in modo formale le **relazioni significanti fra i termini**, le proprietà semantiche, insomma.

Il linguaggio definito dal W3C per scrivere ontologie strutturate, in architettura web, è **OWL** (Ontology Web Language).

Le ontologie devono rappresentare solo **un dominio di conoscenza definito** ed è preferibile che siano connesse tra di loro tramite **core ontologies**, di livello più generale.

Oltre il sito web: il web 3.0: ontologie

31

CIDOC-CRM <http://cidoc.ics.forth.gr/index.html> (CIDOC Conceptual Reference Model)

Un prodotto del Committee on Documentation of the International Council of Museums , **un' ontologia** di 81 classi e 132 **proprietà per il contesto artistico e culturale.**

Accettato dall'ISO nel settembre 2000, ora è ISO/CD 21127:2006 “*Reference ontology for the interchange of cultural heritage information*”

Una guida intellettuale per creare schemi, formati, profili; un linguaggio per analizzare e integrare fonti preesistenti di informazione.

Insomma, CIDOC-CRM **identifica elementi con lo stesso significato**

Nuova forma alle risorse: i Linked Open data

I dati, se isolati, hanno poco valore, lo abbiamo detto. Il loro valore aumenta quando **uno o più dataset** aperti, prodotti e pubblicati in modo indipendente e da soggetti diversi, offrono la possibilità di essere integrati e interrogati tra di loro in modo da creare nuova conoscenza condivisa.

Per riuscirci è necessario rispettare alcune regole base:

- ▶ I *dataset* devono necessariamente essere pubblicati in maniera **aperta e raggiungibile**.
- ▶ I dati devono essere **strutturati secondo un formato standard** riconoscibile e interpretabile dal calcolatore (RDF, RDF/XML, ...).
- ▶ Utilizzo di **URI univoci** per la definizione delle risorse.

i LOD - Linked Open data

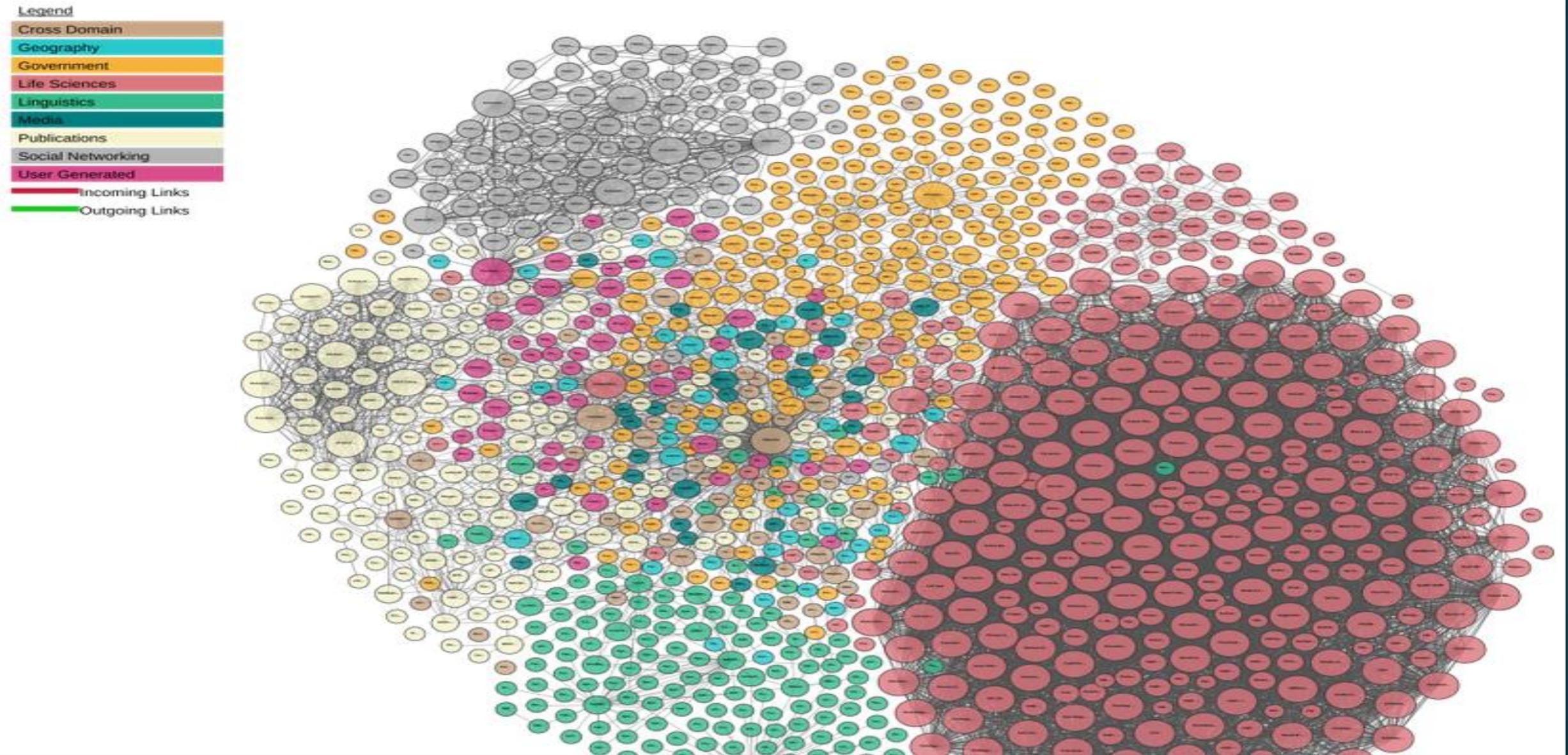
Linked Open Data= buone pratiche per pubblicare e collegare dati strutturati sul Web (Tim Berners Lee)

Attraverso i LOD è possibile pubblicare sul Web i dati in una **modalità leggibile e interpretabile da una macchina**, il cui significato è definito da una stringa di parole e marcatori ('tripla') per contribuire al reticolo di dati collegati appartenenti a un dominio e al tempo stesso collegabili ad altri *dataset* di altri domini nel Web.

Si costruisce così una **rete di dati globale**, i cui contenuti possono essere **scambiati e interpretati dalle macchine**, la base del Web semantico

L'obiettivo è creare dati che siano *del Web* e non solo *sul Web*

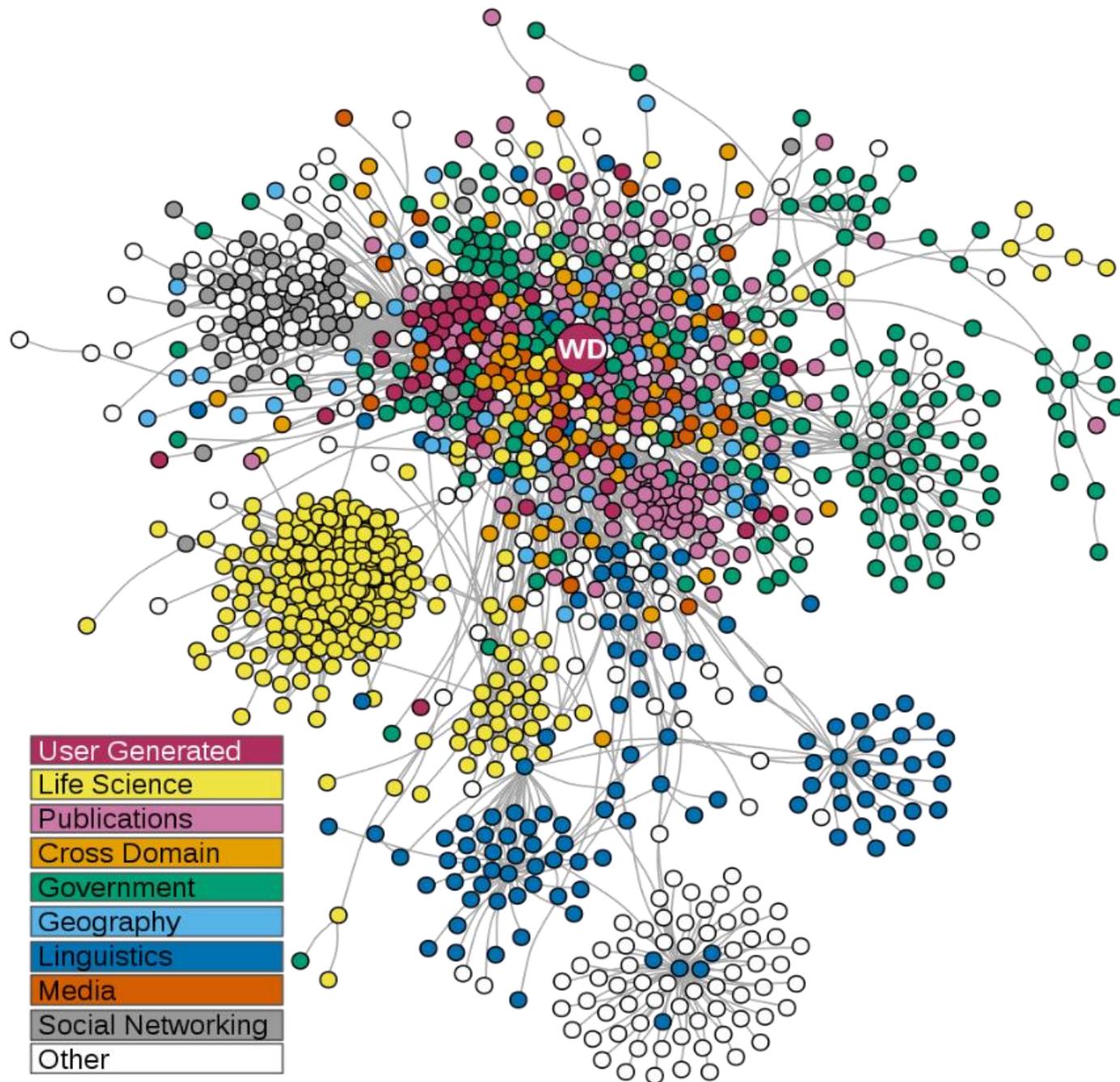
il paesaggio dei Linked Open Data



Wikidata

WIKIDATA è una BASE DI CONOSCENZA GRATUITA E APERTA che può essere letta e modificata allo stesso modo da esseri umani e macchine.

Wikidata, a cui chiunque può contribuire, fornisce un accesso centralizzato alla produzione e gestione di dati strutturati ai progetti Wikimedia tra cui Wikipedia, Wikivoyage, Wikisource e altri. Ma Wikidata fornisce supporto non solo ai progetti Wikimedia, ma anche per molti altri siti e servizi: il suo ricco contenuto semantico è esportabile utilizzando formati standard e può essere interconnesso con altri set di dati aperti (semantic web LOD cloud).



Wikidata
nella nuvola
dei Linked
Open Data

Wikidata

Wikidata è:

Collaborativo. I dati sono inseriti e gestiti dagli utenti di Wikidata, che decidono le regole per la creazione e la gestione dei contenuti. Anche gli agenti web automatizzati inseriscono dati su Wikidata.

Multilingue. La modifica, l'utilizzo, la consultazione e il riutilizzo dei contenuti è completamente multilingue. I dati inseriti in una lingua sono immediatamente disponibili in tutte le altre.

Un database secondario. Wikidata non solo registra le dichiarazioni, ma anche le loro fonti e collegamenti ad altri database; così facendo, riflette la varietà di conoscenze disponibili e supporta l'idea di verificabilità.

Raccolta di dati strutturati. Imponendo un alto grado di organizzazione strutturata, consente un facile riutilizzo dei dati da parte di progetti Wikimedia e terze parti e consente ai computer di elaborarli (Web of data).

Supporto per Wikimedia. Aiuta Wikipedia con dati più gestibili e collegamenti ad altre lingue, riducendo il carico di lavoro per la modifica e migliorandone la qualità. Gli aggiornamenti in una lingua sono immediatamente disponibili in tutte le altre lingue. Chiunque nel mondo può usare Wikidata per un gran numero di modi diversi usando la sua interfaccia di programmazione.