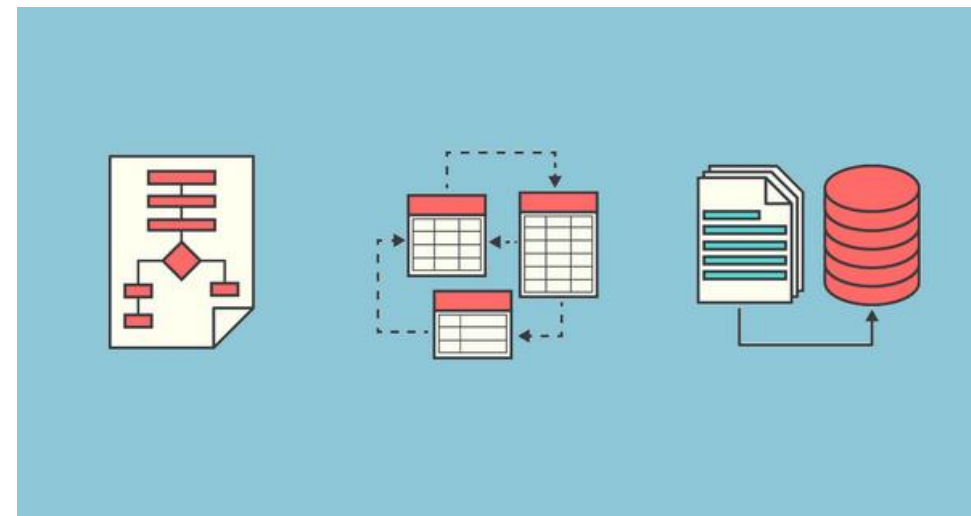




Documentazione digitale per i beni culturali

PROF. PIERLUIGI FELICIATI

PIERLUIGI.FELICIATI@UNIMC.IT

A.A. 2022/23

MODULO 2 –DOCUMENTAZIONE DIGITALE IN RETE

Platone diceva...

MENONE: E **in che modo cercherai**, o Socrate, **ciò che non sai assolutamente cosa sia**? Quale tra le cose che non sai proporrà come **oggetto della tua ricerca**? E se poi, nel migliore dei casi, ti imbattessi in essa, in che modo capirai che questa cosa è ciò che tu non sapevi?

SOCRATE: Capisco cosa vuoi dire, Menone. Vedi come svolgi un discorso eristico per il quale **all'uomo non è dato cercare né ciò che sa né ciò che non sa**? Infatti ciò che sa non lo cercherebbe - perché lo sa e non ha nessun bisogno di cercarlo - né cercherebbe ciò che non sa - e infatti non sa neppure cosa cercare.

Menone, o della Virtù, IV sec. a.C.

Rappresentazione

L'attività e l'operazione di rappresentare con figure, segni e simboli sensibili, o con processi vari, anche non materiali, *oggetti o aspetti della realtà, fatti e valori astratti*.

In **filosofia**, il processo mediante il quale un contenuto di percezioni, immaginazioni, giudizi e concetti, si presenta alla coscienza, e il contenuto stesso.

Con significato analogo, **in psicologia**, ciò che la mente presenta a sé stessa in sostituzione di qualcosa (oggetto, persona o evento) percepito in precedenza, e che costituisce il risultato di un processo percettivo e cognitivo caratterizzato da una relazione più o meno diretta o elaborata con lo stimolo percepito.

(Vocabolario on line Treccani)

Comunicazione

L'origine del termine *communicare* è latina e unisce la preposizione *cum* (con) e il sostantivo *munus*, che significa sia *compito*, dovere che *dono*. In latino lo si utilizzava per indicare l'atto di partecipazione, di concessione ad altri di usare qualcosa di proprio (invitare a pranzo, condividere un messaggio, una gioia...). Questa polisemia di *munus* che si riferisce allo scambio di beni, nel tempo si è ristretto semanticamente allo **scambio di conoscenza, di senso attraverso i segni**.

«il rendere partecipe qualcuno di un contenuto mentale o spirituale, di uno stato d'animo, in un rapporto spesso privilegiato e interattivo».

(Vocabolario on line Treccani)

Documentazione

il complesso delle attività, e delle operazioni intellettuali e tecniche, occorrenti per **raccogliere, classificare e mettere a disposizione** degli studiosi il materiale bibliografico, informativo, documentario, ecc., utile a far progredire e perfezionare il lavoro intellettuale.

Documentazione automatica

l'insieme delle tecniche che consentono di fare ricorso all'elaborazione elettronica e in genere all'informatica per la raccolta, classificazione e registrazione di informazioni in una banca di dati e per il loro rapido reperimento mediante l'esplorazione di opportuni indici e cataloghi di parole chiave

(Vocabolario on line Treccani)

Semantica e Semiotica

Semantica

Ramo della linguistica che si occupa dei fenomeni del linguaggio non dal punto di vista fonetico e morfologico, ma guardando al loro significato.

Semiotica

Scienza generale dei segni, della loro produzione, trasmissione e interpretazione, o dei modi in cui si comunica e si significa qualcosa, o si produce un oggetto comunque simbolico». La filosofia considera il segno in quanto concetto (bellezza, ad esempio, simboleggia qualcosa di complesso, non corrispondente con il suo suono); nella tradizione linguistica, il segno è parola/testo. Il segno ha bisogno di essere segno di qualcosa per qualcuno in grado di coglierlo e interpretarlo.

(Vocabolario on line Treccani)

Semantica e Semiotica (2)

Occuparsi di segni significa confrontarsi con le **relazioni**.

Relazioni **naturali** (tempo, spazio, ...), **reali** (una cosa contenuta in un'altra, relazioni di parentela etc.), relazioni di **causa-effetto** (i risultati di un'azione) ma anche relazioni **ideali**, definite dal pensiero e riconosciute/attivate con un accordo sociale.

Un altro concetto importante è quello di **simbolo**.

In greco antico, *symbolon* indicava **oggetti usati per riconoscere relazioni positive** tra persone, famiglie, città, come i doni recati dagli ospiti. Oggi è *qualsiasi cosa (segno, gesto, oggetto, animale, persona), la cui percezione suscita un'idea diversa dal suo immediato aspetto sensibile*

(Vocabolario on line Treccani)

Informatica

Lo sviluppo dell'informatica è direttamente dipendente dal continuo scambio con altre discipline, teoriche ed applicative, tra cui spicca la tecnologia elettronica, che rende disponibili le macchine capaci di mettere in pratica in modo ottimale i metodi e le soluzioni (algoritmi) elaborate dall'informatica.

L'informatica individua i problemi, ipotizza le soluzioni e mette a punto i metodi per fissare, trattare, elaborare e trasmettere dati e informazioni, facendone unità di conoscenza, in modo indipendente dalle macchine.

In Italia non è riconosciuta una disciplina come *l'Information Science*, lo studio dei metodi e tecniche per la concezione, strutturazione ed elaborazione di raccolte/collezioni di informazioni, al fine di gestirle e renderle disponibili, specie in ambiente digitale.

Informatica e algoritmi

L'informatica in senso proprio è la disciplina che si occupa dello studio degli algoritmi.

La parola algoritmo deriva dal matematico persiano al-Khwarizmi, vissuto nel IX secolo, e significa *sistema definito di regole e procedure di calcolo (sequenziali) che portano alla soluzione di un problema con un numero finito di operazioni.*

Non lo studio dei computer, non lo studio di come scrivere programmi per i computer, non lo studio degli usi di computer e software.



Rappresentare i beni culturali

Per i beni culturali, è importante fare alcune precisazioni importanti:

1. La rappresentazione/descrizione è sempre stata parte dello studio e trattamento degli «oggetti di cultura»
2. Ci sono oggetti di cultura che contengono scrittura, quindi apparentemente sono già portatori di significati decodificabili (L.Moro, dalla somiglianza alla sinergia...)
3. «Se un leone potesse parlare, noi non potremmo capirlo», ovvero che possibilità abbiamo di percepire o comunicare appieno il significato degli oggetti nell'ambiente digitale? (G. Michetti, Se un leone...)

Rappresentare i beni culturali

Discutiamole, una per volta:

1. La rappresentazione/descrizione è sempre stata parte dello studio e trattamento degli «oggetti di cultura»

I catasti, la copia dei manoscritti antichi, gli indici dei libri proibiti, i repertori archivistici, le guide delle città e dei musei del Grand Tour, i cataloghi delle case d'asta, gli inventari delle gallerie d'arte o delle antichità dei sovrani moderni...

è sempre esistita la rappresentazione di oggetti cui si attribuisce valore, per comunicarli/tramandarli, o più semplicemente per ritrovarli, per scambiarli, per pretenderli. Il catalogo nasce alla fine dell'ottocento per scavi, monumenti e gallerie, poi si estende a fotografie, beni naturalistici, beni demoetnoantropologici anche immateriali, oggetti scientifici.

Le tecnologie del presente hanno ampliato le possibilità comunicative e la ricchezza descrittiva, rendendo inoltre possibile l'integrazione tra linguaggi e tipologie di oggetti trattati. Parleremo di portali, di record d'autorità, di interoperabilità

Rappresentare i beni culturali

2. Ci sono oggetti di cultura che contengono scrittura, quindi apparentemente sono già portatori di significati decodificabili

Questo è apparentemente vero, ma abbiamo imparato che i significati contenuti in un oggetto, qualunque sia, devono esser rappresentati per poterlo identificare, selezionare, elaborare. **Va separata la selezione dalla fruizione.**

Da un lato non sarebbe possibile leggere il testo di tutti i documenti di un archivio o tutti i libri di una biblioteca alla ricerca di quello che ci interessa (ammesso che lo sappiamo già), dall'altro perderemmo il valore aggiunto della visione d'insieme, del contesto. Di questo parleremo trattando di cataloghi e inventari elettronici, che non si consultano ma ricercano...

Rappresentare i beni culturali

3. *Se un leone potesse parlare...* che possibilità abbiamo di percepire o comunicare appieno il significato degli oggetti nell'ambiente digitale?

La questione della **semantica adottata** e della reale **efficacia delle rappresentazioni** di beni culturali in ambiente digitale è tutt'altro che risolta. Ogni specialista ritiene che le parole che usa sono le uniche, le migliori possibili, proiettando in sostanza sé stesso e le proprie opinioni e percezioni sull'ignoto utente.

Ma l'unico modo per essere certo che un utente è soddisfatto dei servizi/contenuti in rete è **valutare** la sua esperienza, tendenzialmente **chiedendoglielo!**

Modelli concettuali e modelli di dati

Un **modello**, nella terminologia della scienza, è un insieme di teorie che descrive un fenomeno in modo oggettivo. L'obiettivo di un modello è poter analizzare, studiare, comprendere, quantificare ed elaborare il fenomeno in questione.

Un **modello di dati**, in informatica, rappresenta un insieme di strumenti concettuali, detto "formalismo", indipendente sia dal tipo che dallo specifico DBMS che verrà utilizzato, che comprende tre componenti essenziali:

- Insieme di **strutture dati**, con operatori opportuni.
- **Notazione** per specificare i dati tramite le strutture dati del modello.
- Insieme di **operazioni per manipolare** i dati.

Modello entità-relazioni

Il modello concettuale **Entità-Relazioni (modello E-R)**, descrive ad alto livello la porzione di realtà modellata, indipendentemente da come i dati verranno logicamente e fisicamente rappresentati.

Il modello E-R definisce uno **schema concettuale**. Uno schema descrive la **struttura o forma dei dati, ma non descrive i dati veri e propri**, vale a dire le **istanze o occorrenze** dello schema.

Ad esempio, lo schema ER può dire che l'entità museo ha gli attributi nome e indirizzo ma non dice quali sono gli effettivi valori di questi attributi. I valori verranno inseriti alla fine della progettazione in fase di popolamento della base di dati, creando tante istanze quante saranno necessarie per rappresentare la nostra porzione di realtà.

Modello entità-relazioni

Entità

classi di oggetti (fatti, cose, persone, idee, dati...) che hanno proprietà comuni ed esistenza autonoma ai fini dell'applicazione di interesse. Una **istanza di una entità** è un oggetto della classe rappresentata. Ad esempio, nella descrizione bibliografica, sono entità separate l'autore, l'opera e l'esemplare del libro. Talvolta una entità può essere specializzata in una o più **sotto-entità** che ereditano i suoi attributi e le sue relazioni ma possono definire nuovi attributi e nuove relazioni non presenti nell'entità genitore. Ad esempio, l'entità *persona* può avere la sotto-entità *studente*.

Relazioni

(dette anche associazioni) rappresentano un **legame tra due o più entità**. È possibile legare un'entità con se stessa (attraverso un'associazione ad anello) e legare le stesse entità con più associazioni. Per ogni entità partecipante ad una relazione viene specificata una **cardinalità di relazione**, una coppia di numeri naturali che specifica il numero minimo e massimo di istanze di relazione a cui una istanza dell'entità può partecipare. (0,N oppure 1,N etc.).

Modello entità-relazioni

Una **entità** ha un **nome univoco** nello schema concettuale e viene rappresentata nel diagramma ER con un **rettangolo con il nome** dell'entità all'interno:



Una **relazione** viene rappresentata mediante **un rombo con il nome all'interno** e da **linee che connettono** la relazione con le entità componenti:



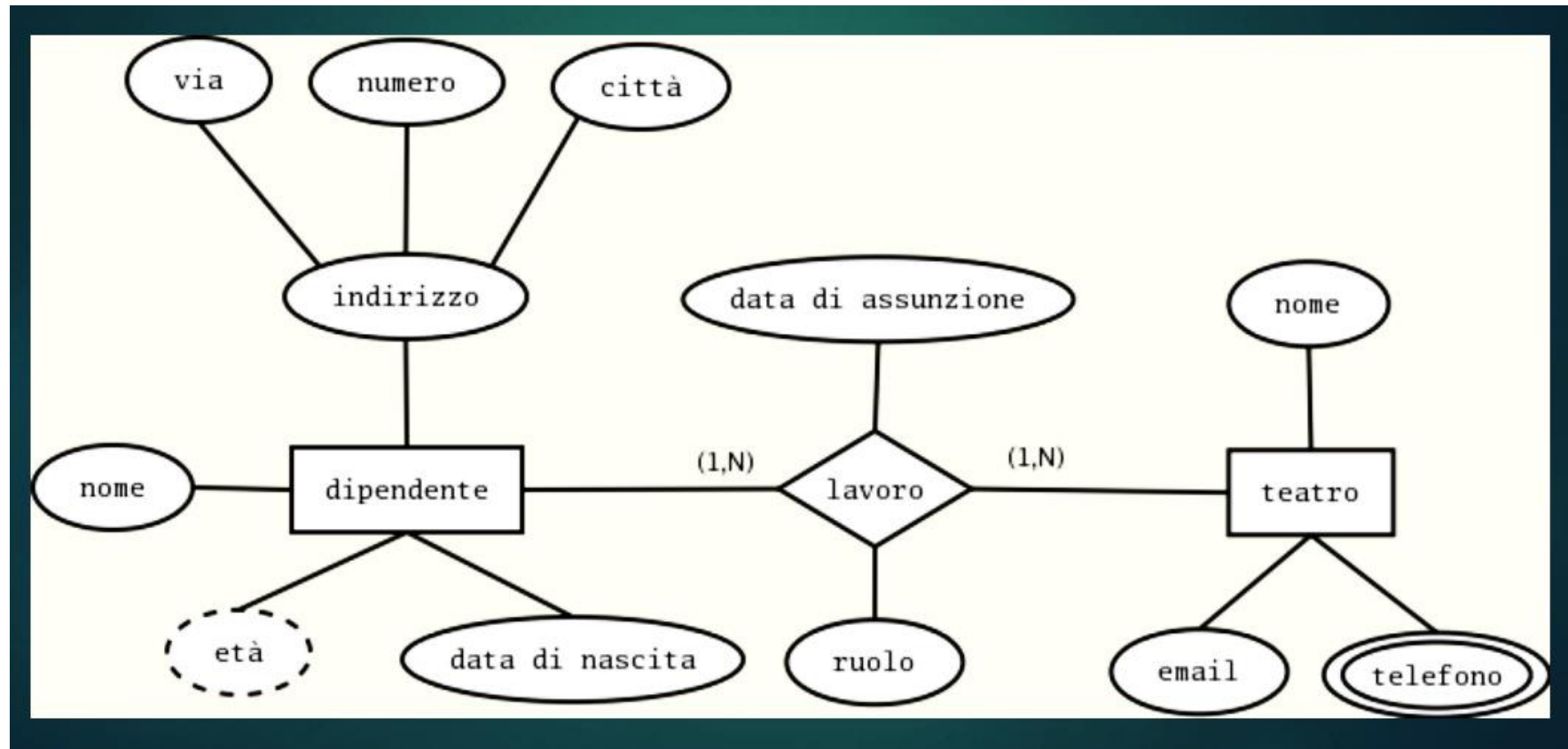
Diagramma entità-relazioni

Un **attributo** non ha esistenza autonoma ma è **associato ad una entità o ad una relazione**.

Un attributo può essere **obbligatorio o opzionale**, e può essere **semplice** (ha, per ogni istanza dell'entità associata, un unico valore, ad esempio un nome di persona), **multivalore** (più valori per la stessa istanza, ad esempio diversi numeri di telefono), **composto** (più attributi correlate, ad esempio indirizzo formato dagli attributi via, numero, e città) o **calcolato** (il valore è ottenuto calcolandolo da quello di altri attributi, ad esempio numero visitatori è la somma dei biglietti pagati e degli omaggi).

Un attributo si rappresenta graficamente con **una ellisse** (una doppia ellisse se multivalore, una ellisse tratteggiata se calcolato) con il nome all'interno, collegata all'entità, alla relazione o all'attributo composto di appartenenza.

Diagramma E-R: un esempio



Qualità di un modello concettuale

E' necessaria la **verifica di qualità** dei modelli concettuali prodotti, che deve concentrarsi almeno sulle seguenti **proprietà**:

- **Correttezza** - Uno schema è corretto quando utilizza nel giusto modo i costrutti messi a disposizione del modello.
- **Completezza** - Uno schema è completo quando rappresenta tutti i requisiti sui dati e quando tutte le transazioni e le regole organizzative possono essere eseguite a partire dallo schema.
- **Minimalità** - Uno schema è minimale quando non è possibile eliminare alcun concetto dallo schema senza per questo rinunciare alla sua completezza. Dunque uno schema non minimale contiene concetti superflui oppure ridondanti.

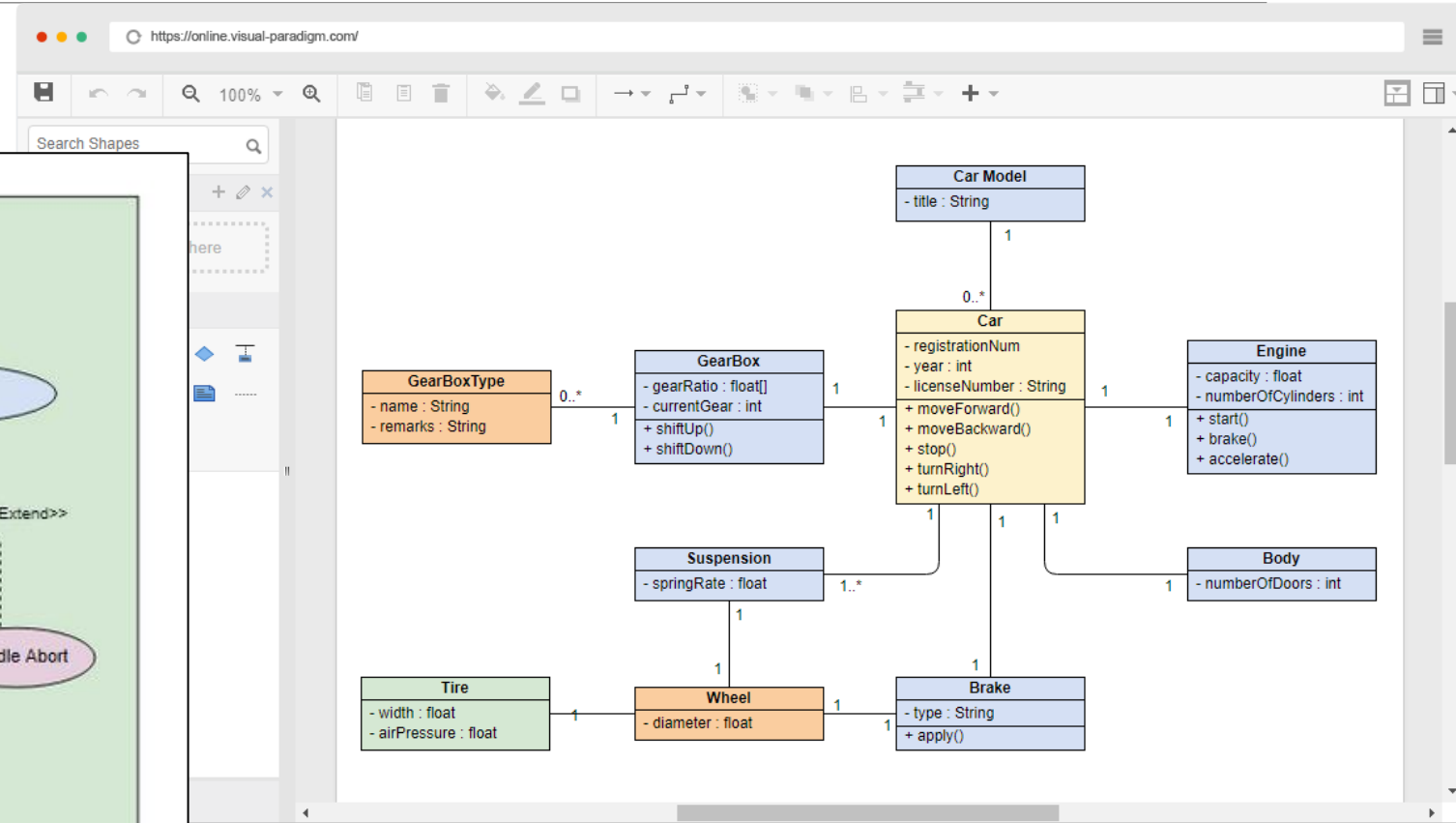
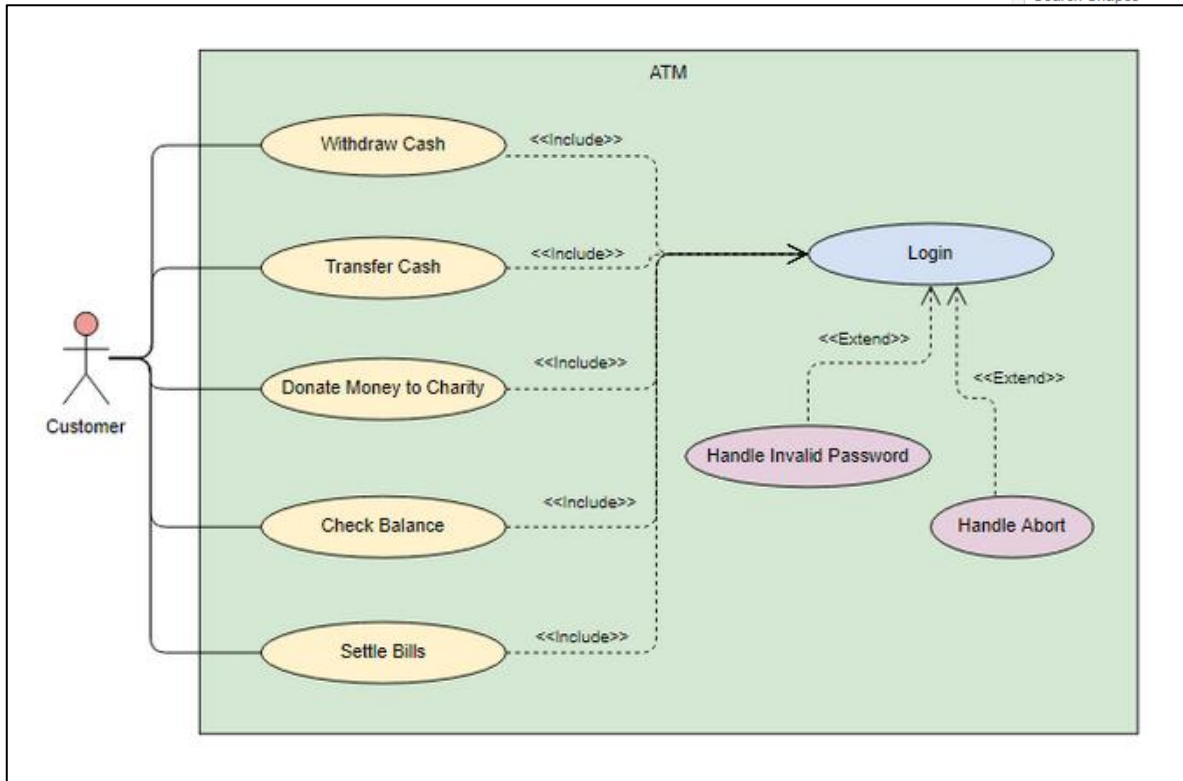
Il linguaggio di modellazione UML

UML (Unified Modeling Language, "linguaggio di modellizzazione unificato") è un meta-linguaggio di modellazione e di specifica generalista. L'ultima versione è la 2.5, rilasciata nel 2013.

Il linguaggio – che compone **elementi grafici** (con significato formalmente definito), **elementi testuali formali** ed elementi di **testo libero** - nasce con l'obiettivo di facilitare il supporto software alla costruzione di modelli e l'integrazione di questo supporto con gli ambienti integrati di sviluppo.

Esistono quindi molte applicazioni, commerciali e libere, che permettono di progettare in UML. Vi segnalo tra i tanti: [ArgoUML](#), [Eclipse](#) (Free); [Microsoft Visio](#), [Smartdraw](#)

Il linguaggio di modellazione UML



Architetture digitali

Le rappresentazioni digitali possono essere **incluse nell'oggetto digitale (metadati)**

Più frequentemente esse sono **esterne (metadati)**, cioè sono altre risorse/oggetti informativi, indipendenti ma collegate, che fungono da frontespizio (o da copertina) o da guida.

Infine possono essere solo **rinvii (link) ad un oggetto in rete** non gestito direttamente, la cui stabilità e le cui caratteristiche possiamo "schedare" ma non controllare nel tempo.

Architetture digitali

Database

Qualunque sia il tipo di database che adottiamo, siamo costretti a fissare delle **regole molto severe** sui dati, dando agli oggetti informativi una struttura rigida: vincoli sulla loro **forma** (testo, numeri, date, etc.), sulle loro **relazioni** (attributi di entità, gerarchie, relazioni tra attributi e tra entità, etc.), sulla loro obbligatorietà o meno, sulla protezione rispetto all'accesso, etc.

Esistono database in cui i dati sono separati rispetto alle regole, altri in cui regole e dati sono riuniti in formati proprietari di tipo applicazione, per noi quasi invisibili.

Architetture di oggetti digitali

Database

La scelta della tecnica del database è ormai quella più ovvia e sicura, tecnicamente, sul lungo termine.

Però, dal punto di vista della progettazione della struttura dei dati, è la più delicata, perché:

- Le regole e le architetture logiche stabilite sono **difficili da cambiare, quindi vanno progettate con molta attenzione**
- La **rigidità delle regole** potrebbe rendere difficile lo scambio di dati, e non sempre esistono degli standard

Architetture di oggetti digitali

Metadati strutturali

E' una tecnica molto "leggera", nel senso che **le entità informative possono avere qualunque forma** (testo, testo con marcatori, testo+multimedia, applicazioni, una somma di questi...) e **le loro relazioni sono documentate in file separati**, di solito di formato text/XML.

I visualizzatori (browser plug-in) e gli elaboratori dei dati leggono i file di metadati e si comportano di conseguenza, permettendoci di "navigare" le strutture.

Possono consentire diverse **logiche di architettura** degli oggetti e sono **più facilmente aggiornabili** rispetto ai database.

Architetture di oggetti digitali

Ipertesto

Un ipertesto è un documento (o un insieme di documenti) composto da **più unità informative** (dette **nodi**), connesse fra loro tramite **collegamenti** (detti **hyper-link** o **link**) appositamente predisposti:

- sia **a priori da chi produce il documento** stesso (che li seleziona fra tutti quelli logicamente possibili e «suggerisce» un percorso)
- che **a posteriori da chi invece legge l'ipertesto**, decidendo autonomamente di percorrere i link seguendo ogni volta un particolare tragitto o percorso (detto **path**) fra i molti che sono stati resi possibili dai creatori o, in certi casi, creandone anche di nuovi.

Questa caratteristica può definirsi **multilinearità**

Architetture di oggetti digitali

Iper testo - storia

Come molti altri fenomeni, gli ipertesti esistevano già molto tempo prima che qualcuno ne definisse il concetto e desse loro un nome. I rinvii, gli indici, le note e le citazioni bibliografiche presenti nei testi enciclopedici, giuridici e scientifici erano già sostanzialmente ipertestuali anche nei secoli precedenti al XX.

«Mi sia consentito di introdurre la parola “ipertesto” per far riferimento a un corpus di materiali scritti o iconografici, interconnessi in una maniera così complessa che non potrebbe essere adeguatamente presentata o rappresentata su carta. Esso può contenere indici o mappe dei propri contenuti e delle loro interrelazioni; può contenere annotazioni, aggiunte o note a piè di pagina inserite dagli studiosi che lo hanno esaminato.» (Ted Nelson, 1962)

Dagli anni '60 si contano diverse sperimentazioni della logica ipertestuale, legate anche alla diffusione domestica dei primi computer.

Architetture di oggetti digitali

Iper testo - storia

Alla fine degli anni '80 del 900 sia il concetto che il termine e la tecnologia dell'ipertesto erano 'di moda' sia fra i produttori che fra gli utilizzatori di prodotti informatici (esisteva un SW Apple che si chiamava HyperCard).

Tim Berners-Lee già nel 1980 aveva lavorato come consulente al CERN a Ginevra, dove sviluppò un sistema ipertestuale basato su schede denominato *Enquire*, un database di persone e di software che però non venne utilizzato. Berners-Lee tornò poi al CERN nel 1984 per lavorare su un sistema per l'archiviazione e la diffusione in tempo reale di documenti scientifici, riprendendo in mano Enquire per cercare di renderlo **compatibile con internet** e con una **molteplicità di database e di software**.

Da qui, nasce il World Wide Web, il Web, il www.

Architetture di oggetti digitali

Ipertesto

Torniamo alla teoria.

Il presupposto fondamentale dell'ipertestualità è la **granularità**, che è la caratteristica posseduta dai documenti scomponibili in parti più piccole, ma ancora dotate di senso e utilizzabili autonomamente, come ad esempio le voci di una enciclopedia.

Solo se un documento è scomponibile in molti nodi essi potranno poi essere collegati fra loro in tanti modi diversi.

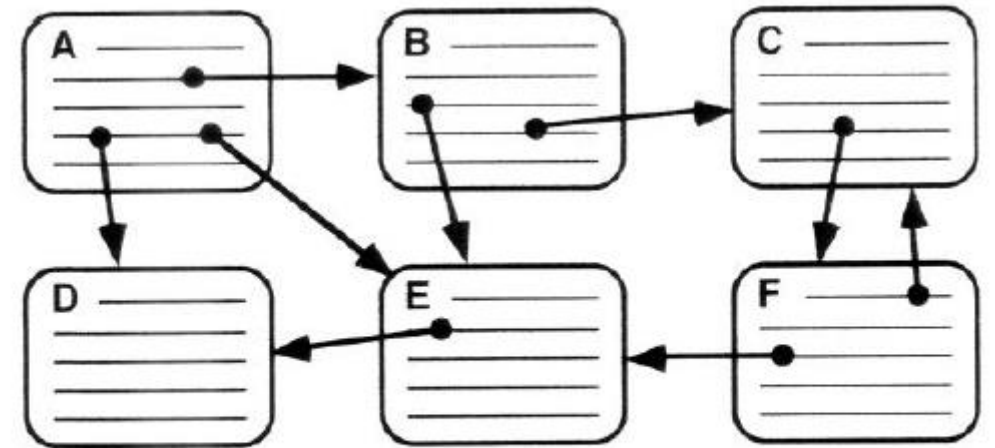


Figura 1 - Rappresentazione schematica di un ipertesto formato da sei nodi e nove link (da Nielsen 1995, p. 1).

Architetture di oggetti digitali

Iper testo

Altre due importanti caratteristiche dell'ipertestualità, oltre alla **multilinearità** e alla **granularità**, sono **l'interattività** e **l'integrabilità**.

- **L'interattività** è la possibilità, da parte del lettore, di intervenire creativamente sul documento, aggiungendo materiale (cioè nodi) o disegnando nuovi percorsi (cioè link), entrambi non previsti dall'autore.
- **Integrabilità** significa indefinita **estensibilità**, ossia la caratteristica per cui, seguendo i link di un ipertesto **si può arrivare ovunque**, procedendo all'infinito, senza mai raggiungere una fine (o un inizio).

Architetture di oggetti digitali

Iper testo

L'ultima caratteristica dell'ipertestualità è la **multimedialità**, che può essere una proprietà:

- sia **dei nodi** (che possono quindi essere dei testi in senso stretto, oppure anche immagini, suoni, filmati o loro miscele)
- che della **struttura dei link**, che può basarsi anche su schemi, diagrammi, grafici, immagini o altre forme di organizzazione dell'informazione di natura non testuale. Privilegiando una disposizione spaziale anziché crono/logica dei contenuti informativi, si rende l'architettura di un ipertesto più simile a una mappa che a un indice.

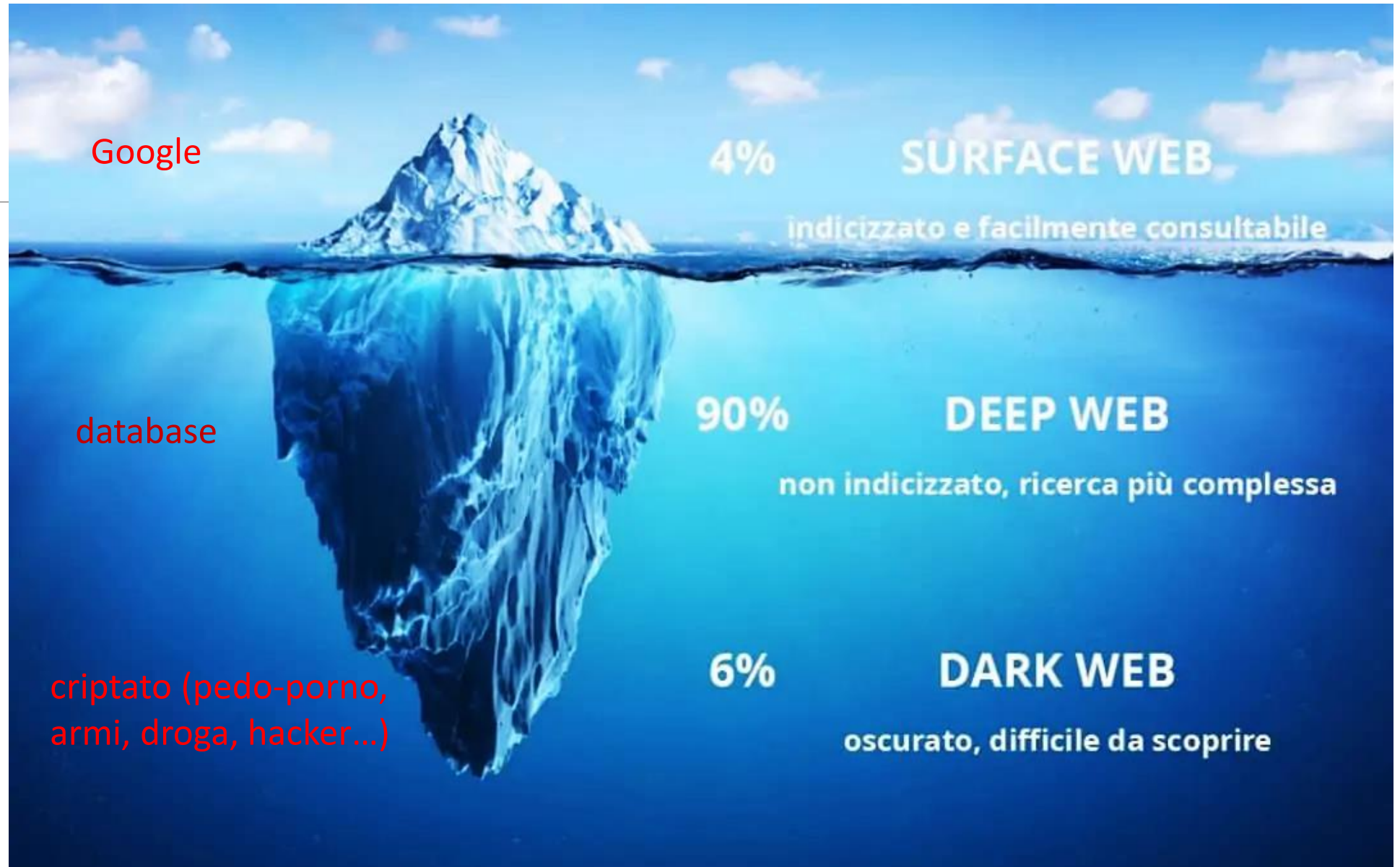
Oggetti, percezione e rappresentazioni

Un **catalogo**, in fondo, è una *collezione di rappresentazioni*

Tipicamente, nel Web, assume la forma di una **banca dati**, *nascondendo* le rappresentazioni dietro alla mediazione forzata di una *query* (*deep web*) e alle scelte grafiche per rendere i risultati usabili e gradevoli (riusabili?)

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

Web
di superficie,
profondo,
oscuro



Oggetti, percezione e rappresentazioni

Cosa è una **biblioteca digitale (digital library)**?

Uno spazio in cui mettere insieme collezione, servizi e persone a supporto dell'intero ciclo di vita della creazione, uso, preservazione di dati, informazione e conoscenza (Wikipedia, tratta dal *Santa Fe Planning Workshop on Distributed Knowledge Work Environments*: Digital Libraries - 1997)

Le biblioteche digitali sono conversazioni, fatte di risorse, di servizi e di persone (Manifesto AIB, <http://www.aib.it/aib/cg/gbdigd05a.htm3>)

Vediamo quella istituzionale dell'Università di Macerata, U-PAD (Pubblicazioni Aperte Digitali) e ragioniamoci sopra: <https://u-pad.unimc.it/>

Biblioteca digitale e catalogo

Se assumiamo che una **Digital Library (biblioteca digitale)** è fatta di ***servizio + utenti + risorse***, dovremmo includere anche gli OPAC, le bibliografie sul Web, i cataloghi di BBCC, i sistemi archivistici, paradossalmente **tutti i buoni siti web**...

Una distinzione necessaria, quindi, è da fare se la DL:

- ❑ è **composta di sole rappresentazioni/schede** (come un OPAC o una bibliografia, che rinvia a risorse esterne, analogiche o digitali)
- ❑ **include gli oggetti digitali** (riproduzioni o no) autoconsistenti, oltre alle loro rappresentazioni

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 di riprodurre tutto ciò che esiste e neanche solo i beni culturali, specie in un paese come l'Italia. Intanto, per la quantità, poi per lo sforzo che richiederebbe creare copie che siano:

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

La riproduzione digitale

Inoltre, la **riproduzione di beni culturali a scopo di valorizzazione** deve essere **coordinata** con la funzione essenziale della **tutela**:

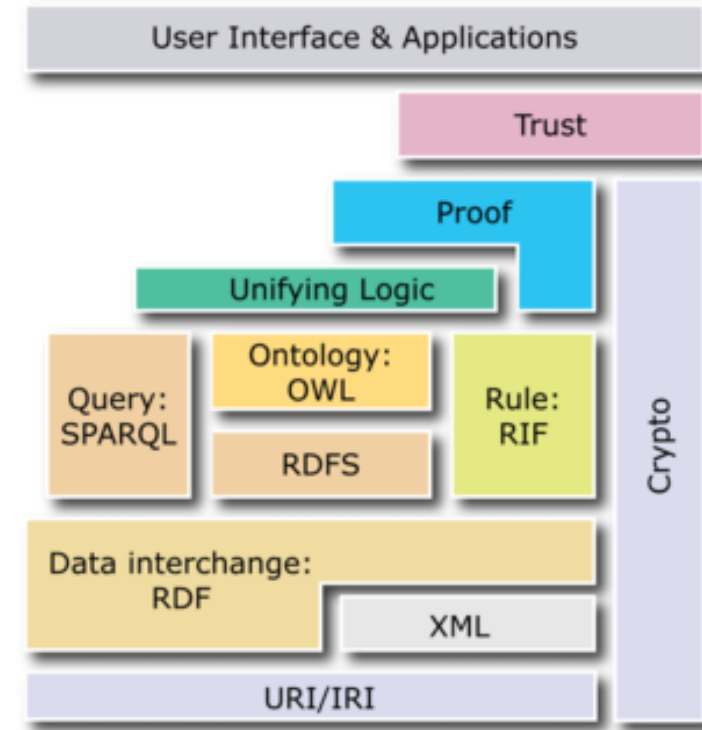
- Garantendo che gli oggetti fisici riprodotti continuino ad essere **accessibili, descritti e tutelati**, ovvero siano **beni comuni**
- Evitando l'effetto detto delle **hidden collections**: esiste solo ciò che è digitale.
- Garantendo insomma che gli oggetti digitali **non sostituiscano in toto gli oggetti reali**, creando su questi ultimi disinteresse da parte dei responsabili.

Il web da arcipelago a galassia

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, se ripensiamo alle sue origini, il Web è tuttora 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 dei dati**, creando uno strato di **Web of data**, che strutturi le informazioni in modo da renderle **comprensibili anche ai software**, facilitandone il ritrovamento.

Oltre il sito web: il Web 3.0

Il **Semantic Web** o **Web 3.0** o **Web dei dati** si basa su una architettura a livelli e la comunità scientifica sta investendo molte energie per implementarlo <http://www.semanticweb.org/>

La sfida è fornire **un linguaggio e degli strumenti** per esprimere dati e regole per far ragionare le machine sui dati. 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

Un altro componente necessario per il web semantico, oltre a XML e RDF è l'**Ontology Vocabulary** (*vocabolario 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 governare solo **un dominio di conoscenza definito** (es. gli archivi, la chimica inorganica, la zoologia...) e di gfa in modo che le ontologie di dominio siano connesse tra di loro tramite **core ontologies**, di livello più generale (concetti più generali: spazio, tempo, agenti, oggetti, materia...).

Oltre il sito web: il web 3.0: ontologie

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

È stato prodotto dal 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 **nel dominio culturale** **identifica elementi con lo stesso significato**

Nuova forma alle risorse: i Linked Open data

I dati, se isolati, però, hanno poco valore.

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 tra di loro e resi interrogabili, 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.

LOD - Linked Open data

«I Linked Open Data sono 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**, dati il cui significato è definito da una stringa di parole e marcatori (una *tripla*) che contribuisce al reticolo di dati collegati appartenenti a un dominio, ma al tempo stesso collegabili automaticamente ad altri *dataset* di altri domini nel Web.

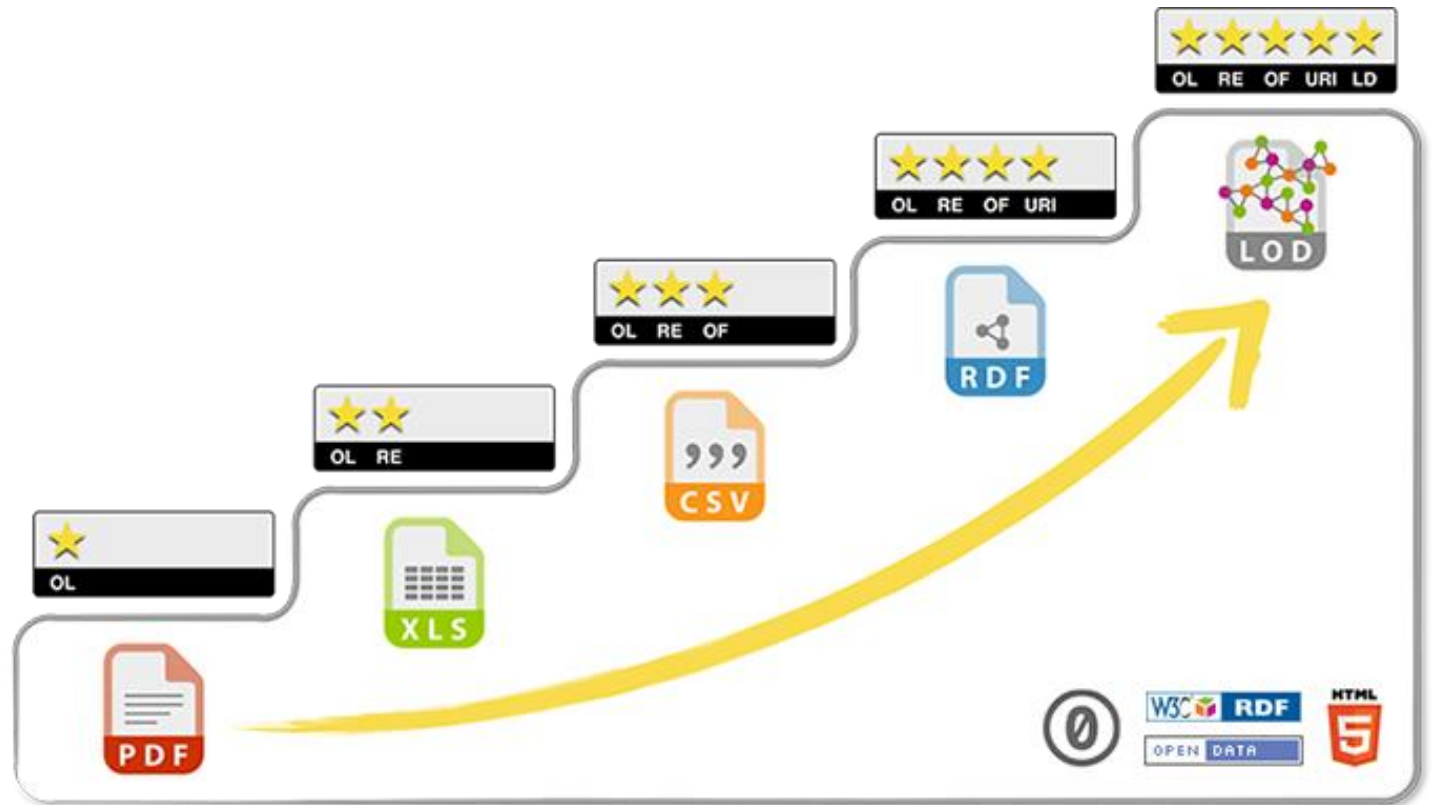
Si costruisce così una **rete di dati globale**, i cui contenuti possono essere **scambiati e interpretati dalle macchine**, la base del Web semantico

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

LOD - Linked Open data

Tim Berners-Lee aveva definito 5 livelli di efficacia per i dati in rete:

- 1 Stella – non dati (tipo PDF o immagini)
- 2 Stelle – tabelle
- 3 Stelle – tabelle in formato aperto
- 4 Stelle – asserzioni XML in RDF
- 5 Stelle – Linked Open Data



il paesaggio dell'*info-sfera* semantica (la galassia dei LOD)

