

Métadonnées

Jean-Marc LECARPENTIER
Université de Caen Normandie

Métadonnées

- Data about data
- Machine-understandable information about Web resources or other things
Tim Berners-Lee, W3C (1997)
- Structured data that describes the attributes of an "information package"
Taylor, 1994

Métadonnées

- Structured data about resources that can be used to help support a wide range of operations
Michael Day, 2001
- Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use or manage”
information objects
NISO, 2004
- Data associated with objects which relieves their potential users of having to have full advance knowledge of their existence or characteristics
Dempsey & Heery, 1998

Types de métadonnées

- Descriptive
- Structure
- Administrative
- Technique



Types de métadonnées

- Descriptive : sujet, titre, description ou résumé, présentation, type de document, public visé, etc
- Administrative : date de création, droits d'auteurs, droits d'accès ou de diffusion, disponibilité, etc
- Structure : composants (par ex. œuvre en plusieurs volumes), chapitres, etc
- Techniques : format de fichier, compression, type d'encodage, codec audio ou video, etc

Pour quoi faire ?

- Publication et découverte de ressources
- Utilisation de ressources
- Gestion de la propriété intellectuelle
- Commerce
- Localisation

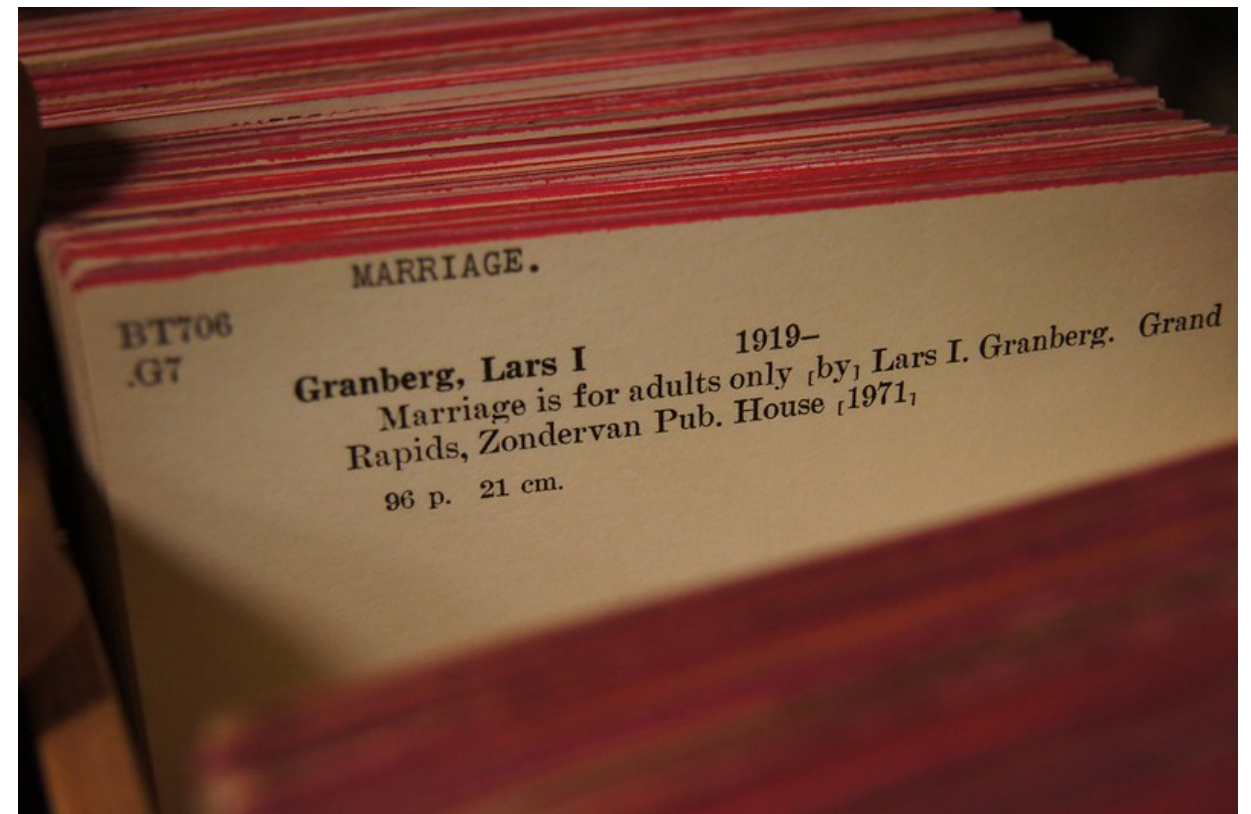


Image par Ted Eytan, CC-by-sa, <https://flic.kr/p/9k2EsX>

À quoi les appliquer ?

- Tim Berners-Lee : "Web resources or other things"
- sites Web, pages Web, images, bases de données, objets de musées, localisations géographiques, événements, concepts, archives, etc.
- Aux métadonnées elles-mêmes

Métadonnées et numérique

- Informations traitées par des machines
- Tim Berners Lee : Machine-understandable information
- Données structurées
- Associer des objets
- Quel(s) format(s) utiliser ?

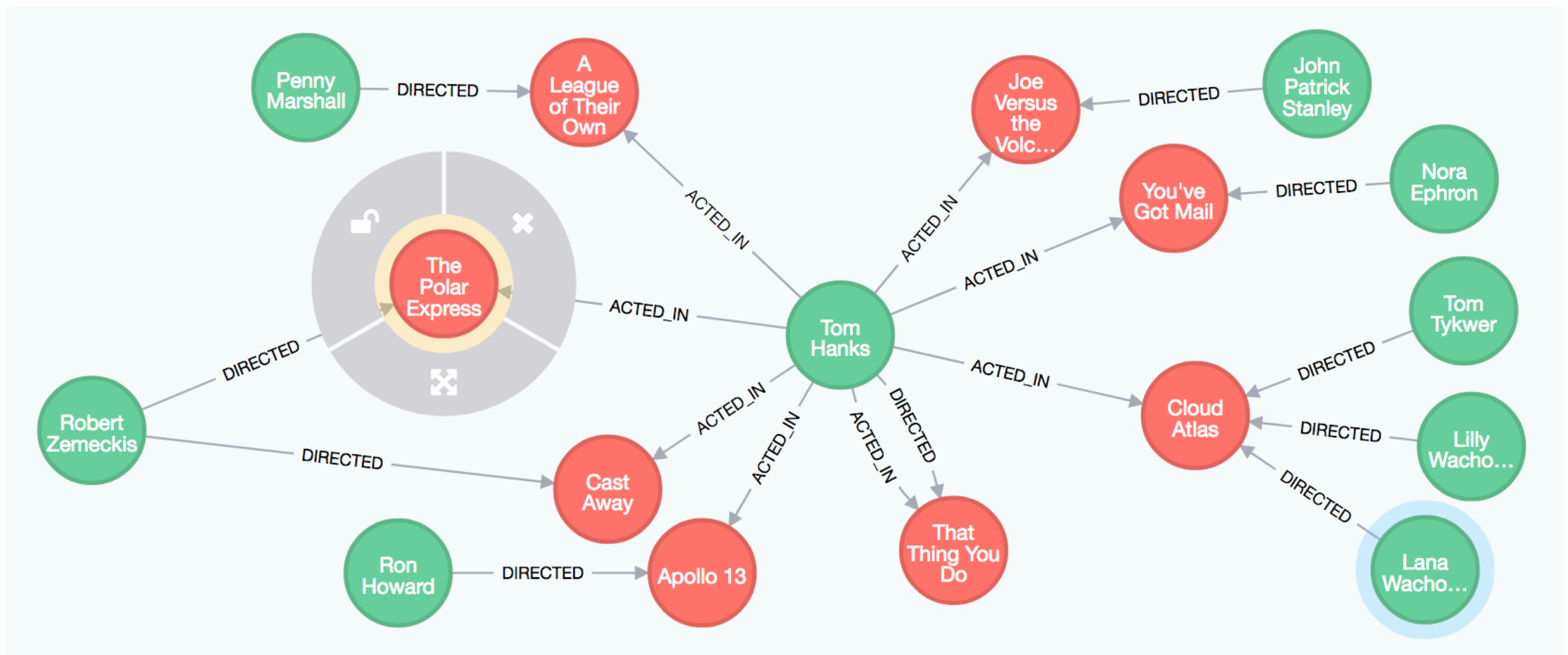
RDF

- Resource Description Framework
- Spécification W3C, 1999
- Représenter des informations à propos de ressources sur le Web
- Adapté à la description des métadonnées
- Adapté au traitement de l'information par des machines



Graphe RDF

Décomposer toute structure en un graphe qui « atomise » les informations

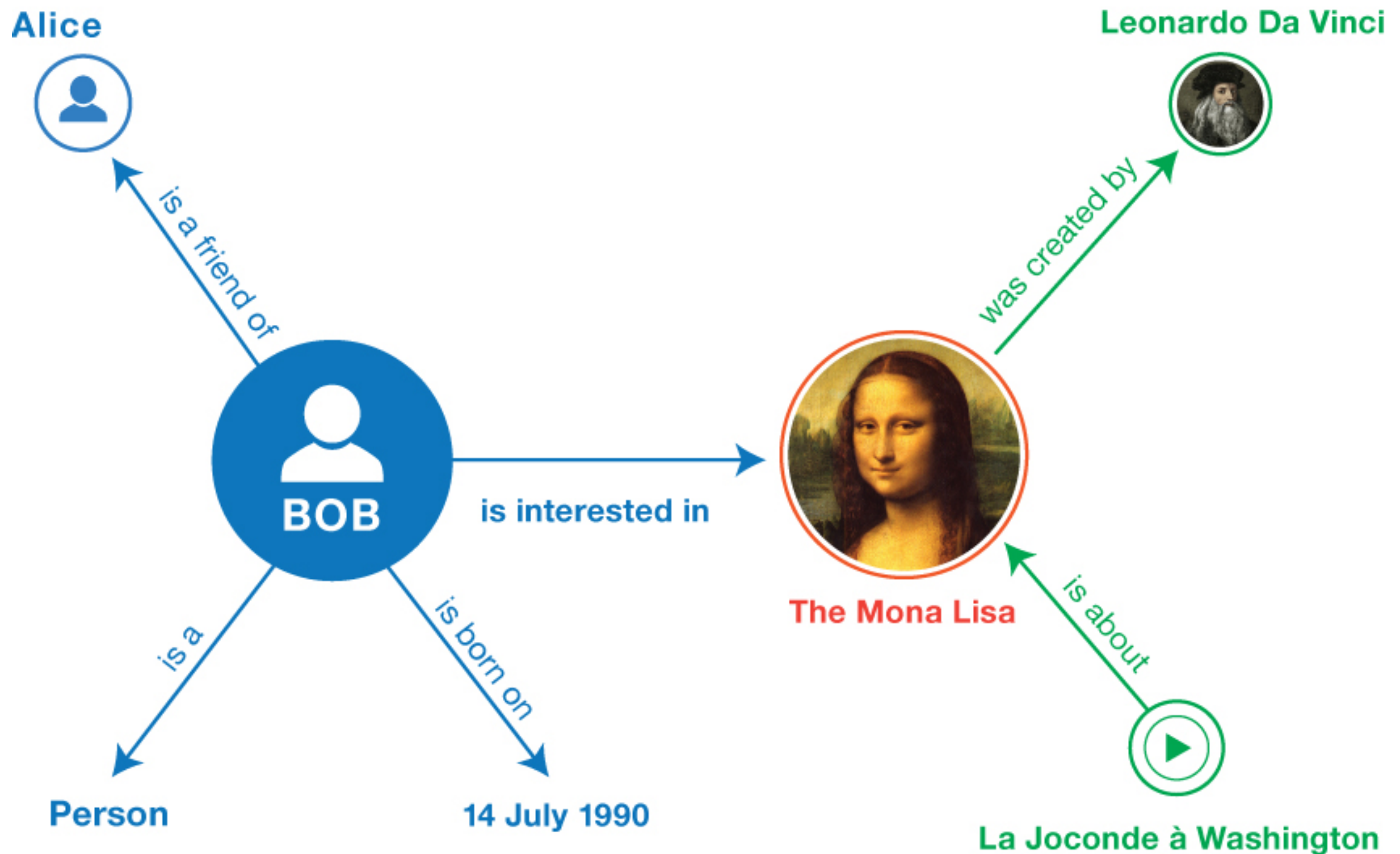


Source image <https://wilsonmar.github.io/graph-databases/>

Triplets RDF

- Source des exemples et illustrations : W3C RDF Primer
<https://www.w3.org/TR/rdf11-primer/>
- Triplet RDF : représentation d'une information à propos d'un « sujet » : **<subject> <predicate> <object>**
- Exemple :
 - <Bob> <is a> <person>.
 - <Bob> <is a friend of> <Alice>.
 - <Bob> <is born on> <the 4th of July 1990>.
 - <Bob> <is interested in> <the Mona Lisa>.
 - <the Mona Lisa> <was created by> <Leonardo da Vinci>.
 - <the video 'La Joconde à Washington'> <is about> <the Mona Lisa>

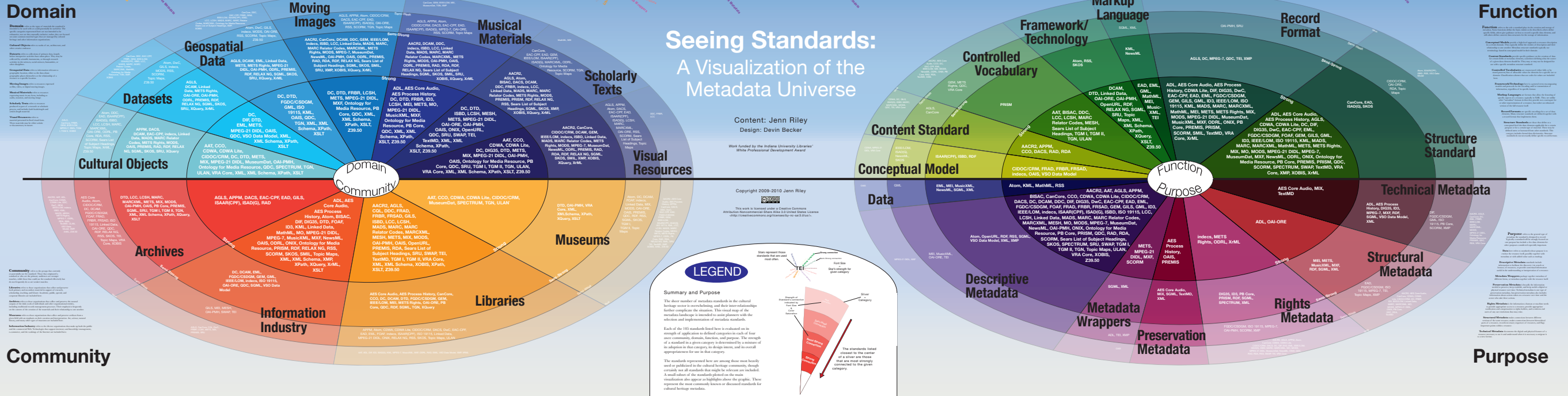
Triplets RDF



Vocabulaire

- Problématique : dans la relation
<Bob> <is born on> <the 4th of July 1990>
comment représenter le prédicat « est né le » pour que ce soit utilisable par les machines ?
- Nécessité pour tous d'utiliser le même formalisme
⇒ définition d'un vocabulaire RDF
⇒ géré/édité/certifié par un organisme « officiel »
⇒ communauté d'utilisateurs
- Utilisation de XML et des *namespace XML* (basé sur URIs)

Jenn Riley, CC-by-nc-sa, <http://jennriley.com/metadatamap/>



Notation N-triplets

<http://example.org/bob#me> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://xmlns.com/foaf/0.1/Person>

<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/knows> <http://example.org/alice#me>

<http://example.org/bob#me> <http://schema.org/birthDate> « 1990-07-04"^^<http://www.w3.org/2001/XMLSchema#date>

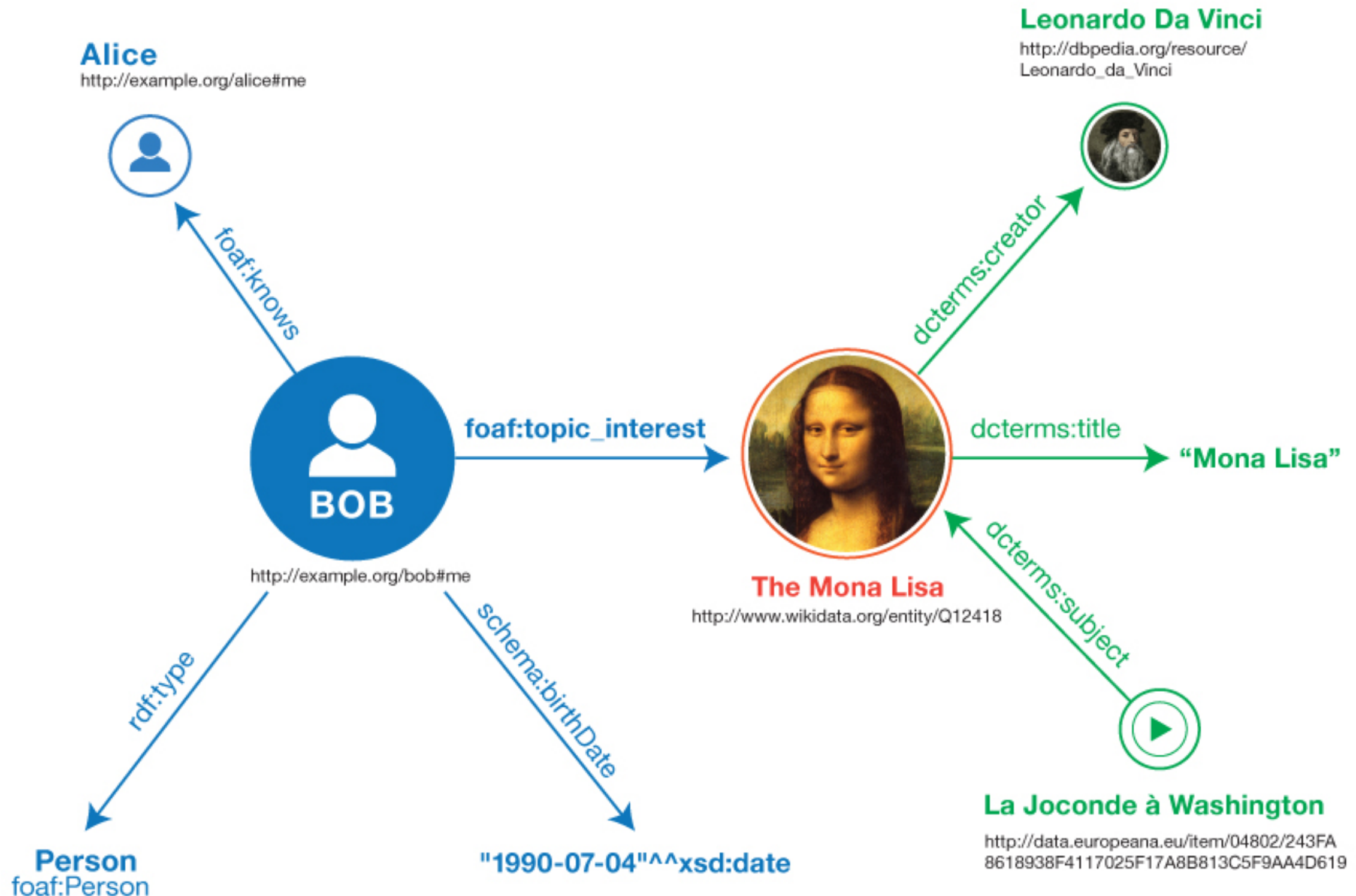
<http://example.org/bob#me> <http://xmlns.com/foaf/0.1/topic_interest> <http://www.wikidata.org/entity/Q12418>

<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/title> "Mona Lisa"

<http://www.wikidata.org/entity/Q12418> <http://purl.org/dc/terms/creator> <http://dbpedia.org/resource/Leonardo_da_Vinci>

<http://data.europeana.eu/item/04802/243FA8618938F4117025F17A8B813C5F9AA4D619> <http://purl.org/dc/terms/subject> <http://www.wikidata.org/entity/Q12418> .

Préfixes XML

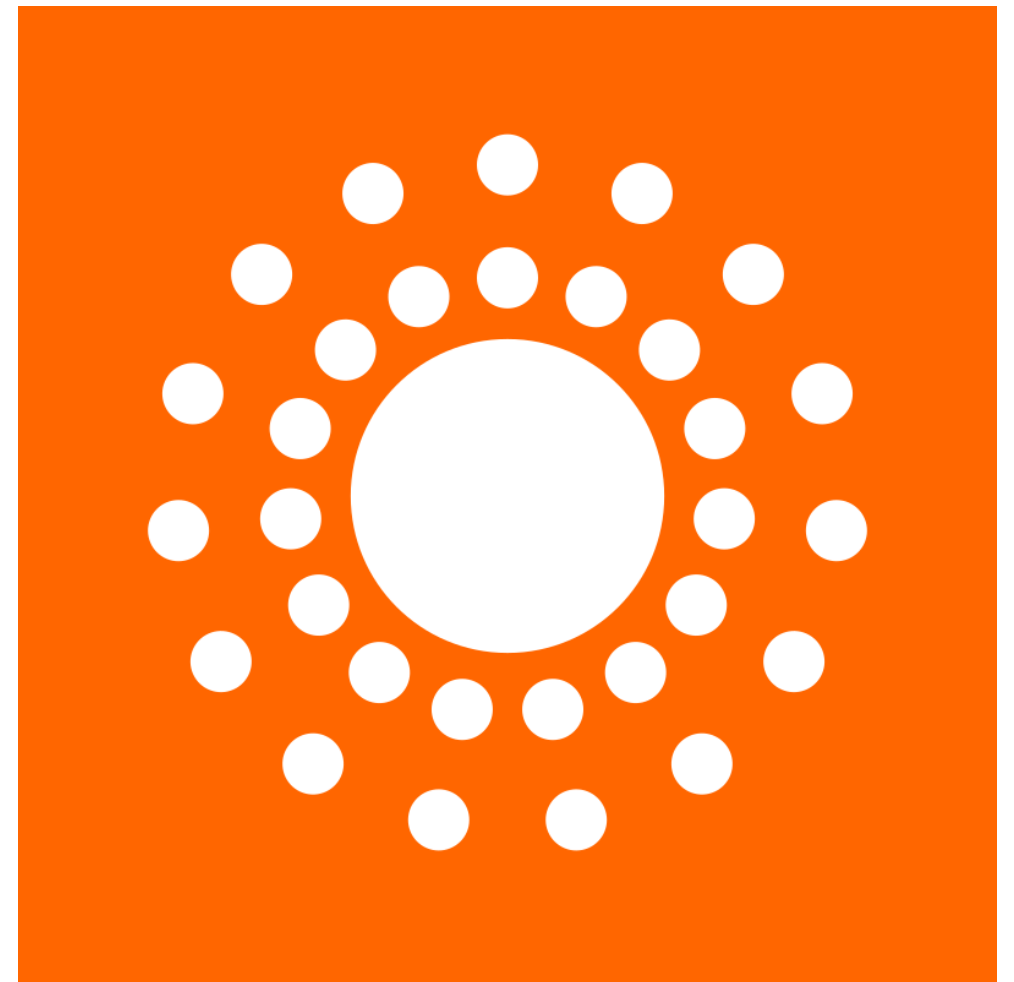


XML/RDF

```
01  <?xml version="1.0" encoding="utf-8"?>
02  <rdf:RDF
03      xmlns:dcterms="http://purl.org/dc/terms/"
04      xmlns:foaf="http://xmlns.com/foaf/0.1/"
05      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
06      xmlns:schema="http://schema.org/">
07      <rdf:Description rdf:about="http://example.org/bob#me">
08          <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
09          <schema:birthDate rdf:datatype="http://www.w3.org/2001/
XMLSchema#date">1990-07-04</schema:birthDate>
10          <foaf:knows rdf:resource="http://example.org/alice#me"/>
11          <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/
Q12418"/>
12      </rdf:Description>
13      <rdf:Description rdf:about="http://www.wikidata.org/entity/Q12418">
14          <dcterms:title>Mona Lisa</dcterms:title>
15          <dcterms:creator rdf:resource="http://dbpedia.org/resource/
Leonardo_da_Vinci"/>
16      </rdf:Description>
17      <rdf:Description rdf:about="http://data.europeana.eu/item/
04802/243FA8618938F4117025F17A8B813C5F9AA4D619">
18          <dcterms:subject rdf:resource="http://www.wikidata.org/entity/Q12418"/>
19      </rdf:Description>
20  </rdf:RDF>
```

Dublin Core

- Vocabulaires multiples : besoin d'uniformiser
- Dublin Core Metadata Initiative (DCMI)
- Création en 1995 suite à un workshop à Dublin, Ohio
- Schéma généraliste
- Renseigner de façon universelle des informations de base
- Mission : fournir des standards simples pour faciliter la découverte, le partage et la gestion de l'information
- Consensus interdisciplinaire sur 15 éléments



Dublin Core

- Consensus sur 15 éléments pour la découverte d'informations
- Éléments optionnels, répétables
- Axé sur la simplicité → ne convient pas pour des descriptions complexes
- Pas de règles de catalogage
- Fournit une interopérabilité sémantique de base au travers des domaines et des communautés

Éléments Dublin Core

- Title
- Subject
- Description
- Creator
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

Applications

- Catalogues de bibliothèques, musées, archives
- Machine-understandable : automatisation des traitements
- Interopérabilité entre catalogues : par ex. la BNF et la Library of Congress peuvent échanger des informations sans avoir à modifier leurs formats de données
- Logiciels et traitement des données : affichage de votre bibliothèque de musique ou ebooks dans divers logiciels

Métadonnées embarquées

- Métadonnées inscrites dans le document/fichier
→ informations transférées avec le fichier
- Exemples :
 - pages web
 - images
 - fichiers musicaux (MP3)
 - documents PDF
 - vidéos
 - livres numériques

Pages web

- Balises `<meta property="..." content="..." />`
- Microdata
- Dublin Core :
 - 1- déclarer le préfixe
`<head prefix="dcterms: http://purl.org/dc/terms/#">`
 - 2- ajouter les propriétés
`<meta property="dcterms:title" content="Ma page" />`
`<meta property="dcterms:description" content="..." />`

Microdata

HTML

- Apparu avec HTML5
- Inclure des informations complémentaires au fil du texte de la page
- Contenu enrichi pour les moteurs de recherche
- Initiative schema.org



Microdata

- Microdata extraites de la page
- Affichées dans les résultats de recherche
- Augmente la visibilité de la page dans les résultats

[Little Water Cantina - Eastlake - Seattle, WA](#)
[www.yelp.com](#) › Restaurants › Mexican
★★★★★ 90 reviews - Price range: \$\$
90 Reviews of **Little Water Cantina** "Three things are on my list when I eat out: great food, atmosphere, and

[Vegetarian Vegan Pizza No Cheese\) Recipe - Food.com - 248865](#)
[www.food.com/recipe/vegetarian-vegan-pizza-no-c...](#)
★★★★★ 2 reviews - 1 hr 32 mins - 242.9 cal
Aug 26, 2007 – This is from my dad, who developed some **vegan recipes** esn't have any cheese, and you

[Leonard Cohen – Free listening, videos, concerts, stats, & pictures at ...](#)
[www.last.fm/music/Leonard+Cohen](#)
Watch videos & listen to **Leonard Cohen**: Suzanne, Hallelujah & more, plus 132 pictures. **Leonard Cohen**, (born September 21, 1934 in Montréal, Quebec, ...

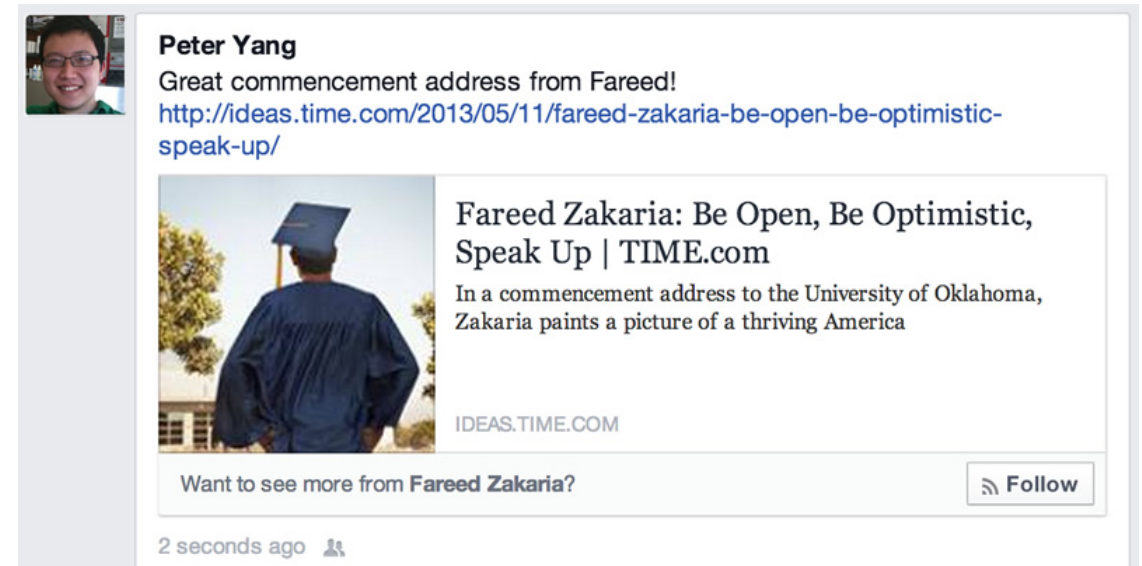
| Track | Duration |
|------------------------------|----------|
| Suzanne | ♫ 3:48 |
| The Darkness | ♫ 4:29 |
| Going Home | ♫ 3:51 |
| Hallelujah | ♫ 6:12 |

Microdata

- Documentation et tutoriel :
<https://schema.org/docs/gs.html>
- Tester sa page :
<https://search.google.com/structured-data/testing-tool/>
- Indispensable dans les sites modernes

Métadonnées et médias sociaux

- Facebook Open Graph Protocol
- Twitter Cards
Balises meta spécifiques
- Partage de liens affiché de façon plus complète
- OGP <https://ogp.me>
- Twitter cards
<https://developer.twitter.com/>



Fonctionnement

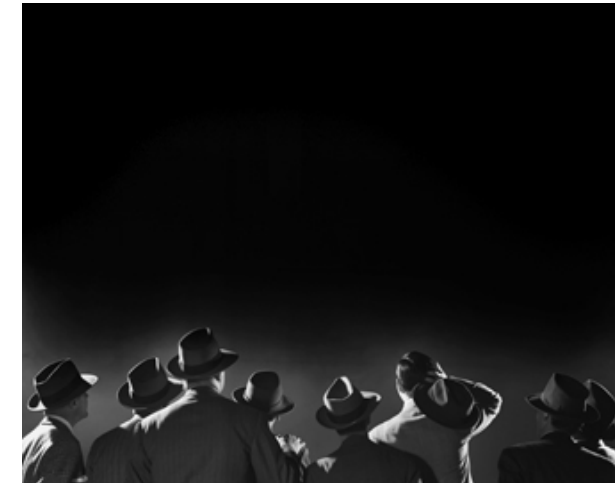
- 1.Utilisateur poste un lien avec son URL
- 2.Facebook/Twitter lance un robot qui analyse la page web
- 3.Extrait les données OGP/Cards et les stocke en mémoire
- 4.Affiche le partage du lien en utilisant les infos extraites
- 5.Si d'autres utilisateurs postent le même URL, inutile d'extraire à nouveau les données

Métadonnées et pages Web

- Importance stratégique pour un site
- Moteurs de recherche
- Réseaux sociaux
- Structure de l'information
- Destinées aux humains et aux machines

Métadonnées et images

- Enjeux importants
- Droits d'utilisation, de reproduction
- Utilisation par les logiciels
- Interopérabilité



| | |
|---------------------------------|---|
| File Name | BE023697.jpg |
| Headline | Men in Hats Looking Upward |
| Copyright Notice | © Corbis. All Rights Reserved. |
| Credit | © Bettmann/CORBIS |
| Caption/Abstract | ca. 1945 — Group of behatted men looking into the night. Image by © Bettmann/CORBIS |
| Keywords | 1940s style, 1950s style, Adults, Anticipation, Awe, Back view, Clothing, Confusion, Copy space, Fedora, Felt hat, Few, Hat, Headgear, Looking, Males, Men, Mid-adult, Mid-adult man, Mystery, Night, Old-fashioned, Outdoors, Outfit, People, Retro, Searching, Spectators, Suits, Waiting, Whites |
| Object Name | BE023697 |
| Original Transmission Reference | 10354723 |
| Special Instructions | For latest restrictions check www.corbis.com |

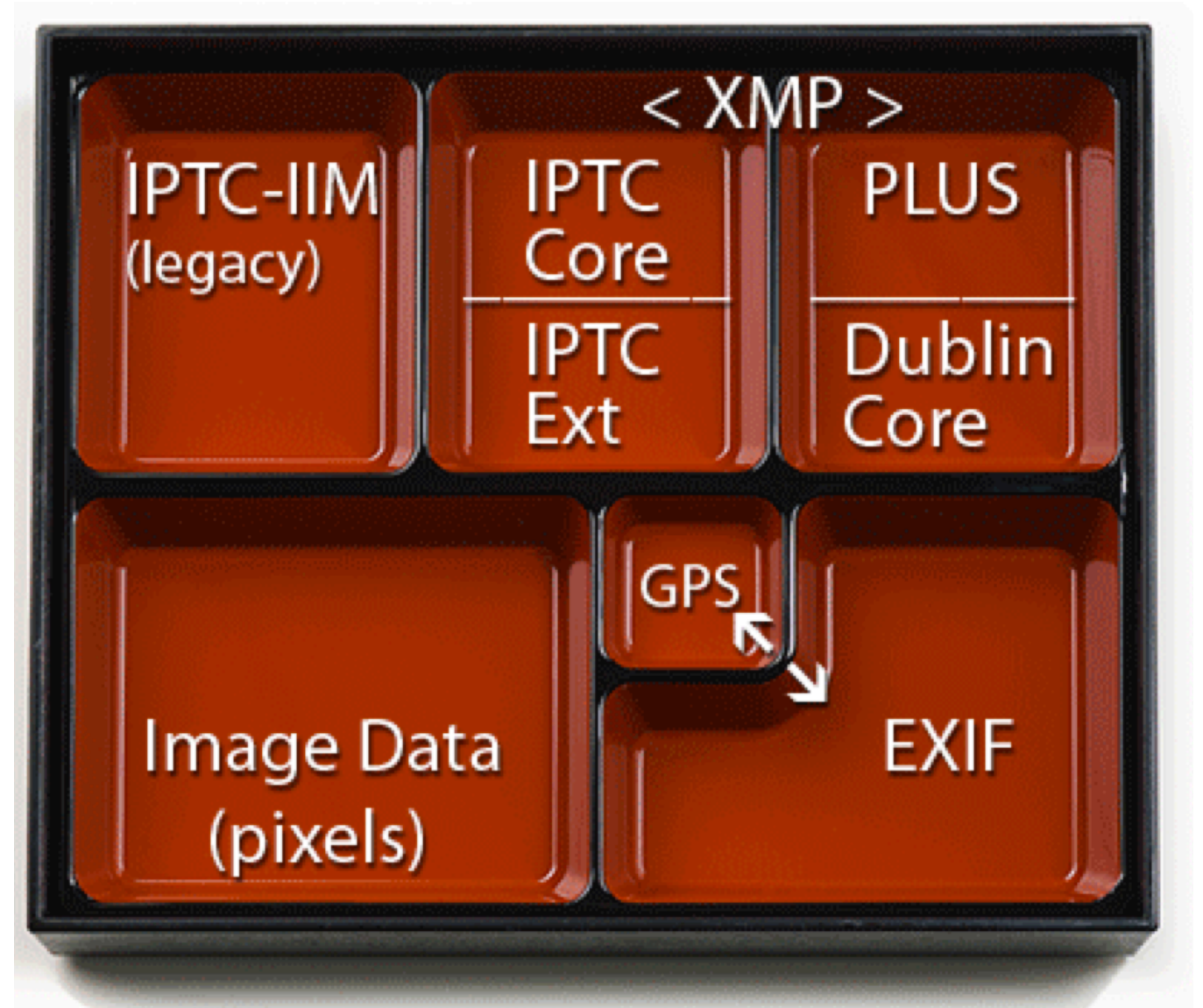


Métadonnées et images

- Plusieurs schémas principaux
 - Exchangeable Image Format (EXIF) : données techniques fournies par l'appareil
 - International Press Telecommunications Council (IPTC) : données sur le contenu de l'image, ses ayants-droits, etc
- XMP (Extensible Metadata Platform) : mécanisme générique d'inclusion de métadonnées dans les fichiers
⇒ très complet

Images et métadonnées embarquées

- Données de l'image (binaires)
- Métadonnées (texte)
- Inclus dans un seul fichier



EXIF

- Métadonnées techniques
- Créées par l'appareil photo
- Marque et modèle de l'appareil
- Données sur la prise de vue (ouverture, focale, vitesse d'obturation, flash, etc)
- Données sur le fichier (résolution, couleur, compression, etc)
- Données temporelles (date de prise de vue)
- Données GPS (selon l'appareil)

IPTC

- Métadonnées descriptives
- IPTC Core : métadonnées de base
- Title, Headline, Description, Creator, Right Usage Terms
- IPTC Extended : métadonnées plus précises
- Event, Image Supplier, Person Shown in The Image, Model Name, etc

Autres schémas

- File : métadonnées du fichier
- ICC : color profile
- PrintIM : Epson Print Image Matching pour les imprimantes Epson
- Photoshop : métadonnées générées par le logiciel

XMP

- Extensible Metadata Platform créé par Adobe
- Mécanisme générique d'inclusion de métadonnées
- Format XML/RDF
- Nombreux formats : images (gif, jpeg, png, tiff), PDF, son (mp3, wav, swf, wma), video (asf, mov, avi, wmv, flc)
- Format XML et RDF permettant d'intégrer tout schéma de métadonnées
- Possibilité de spécifier les données en plusieurs langues
- Interopérabilité maximale

Outils

- exiftool : le couteau suisse des métadonnées
- Exiftool créé par Phil Harvey (Queen's University, Ontario, Canada)
- Exécutable en PERL
- <https://exiftool.org>
- Viewer en ligne : <http://exif.regex.info/exif.cgi>
- Viewer avec carte : <https://www.pic2map.com>

Conclusion

- Sujet vaste et complexe
- Solutions dépendent du contexte et des communautés de pratique
- Problématique pour le développeur : quid des métadonnées ?
 - lesquelles ? quel schéma ? quel vocabulaire ?
 - métadonnées le plus souvent non gérées par les applications
- Comment apporter des solutions "génériques" qui peuvent être mises en application dans divers cas ?