

TEF en RDF

- PREMIER ESSAI -

Auteur : Yann NICOLAS, ABES ([nicolas chez abes point fr](mailto:nicolas@abes.fr))

Résumé : TEF (<http://www.abes.fr/abes/documents/tef/index.html>) est un vocabulaire pour les thèses électroniques françaises. C'est une recommandation AFNOR dont le but est de faciliter les échanges de métadonnées de thèses entre les établissements habilités français. Les métadonnées TEF sont encodées en XML (METS).

Ce texte est une première tentative pour exprimer TEF dans le langage RDF. TEF en RDF permettra de mieux refléter le modèle conceptuel de TEF, appuyé sur les FRBR, et d'explicitier les relations entre TEF et d'autres vocabulaires (Dublin Core, FOAF...). Les métadonnées de thèse seront alors plus faciles à diffuser, à réutiliser et à agréger à d'autres métadonnées (dans le cadre local d'un système d'information ou à l'échelle du Web).

Cet essai n'engage pas le groupe AFNOR, auteur de TEF.

Date : 4 Juillet 2007

RDF en quelques mots.....	2
Pourquoi RDF ?.....	5
De quoi on parle ?.....	8
Les personnes.....	13
Les autorités.....	16
Les ressources externes.....	20
Les Droits.....	22
A quoi sert le jury ? Comment faire confiance en RDF ?.....	25
Conclusion.....	27
Bibliographie.....	28

Dans la [recommandation](#)¹, on trouve ce passage qui pour justifier la priorité accordée à un encodage de TEF en XML (et non en RDF) :

(...) si l'échange de métadonnées complètes et valides à l'échelle nationale est l'un des besoins premiers auxquels TEF doit répondre, alors RDF n'est sans doute pas la meilleure solution. La logique RDF est de décomposer l'information en une série de petits faits élémentaires (" ceci a pour titre cela", "ceci a pour créateur untel"...), en principe indépendants les uns des autres. Ce qui importe en RDF n'est pas qu'un ensemble de métadonnées soit complet, autosuffisant et conforme à un schéma prescriptif, mais au contraire qu'il puisse être complété, enrichi par un autre ensemble de métadonnées, accessible sur le Web.

Dans les pages qui suivent, je vais présenter un premier stade de la réflexion sur l'expression de TEF en RDF. C'est un essai. Il n'engage pas le groupe AFNOR.

Exemple RDF : http://www.abes.fr/abes/documents/tef/rdf/tef_rdf_04072007.xml

Schéma OWL : http://www.abes.fr/abes/documents/tef/rdf/tef_owl_04072007.xml

Pour suivre les évolutions de ce travail « TEF en RDF », on peut consulter la section RDF du blog de TEF².

RDF en quelques mots

RDF n'est pas un format de données. C'est - littéralement - un cadre pour décrire des ressources (*Resource Description Framework*). C'est un modèle, un moule qui oblige à penser toute information sous la forme :

Sujet	Verbe	Complément
--------------	--------------	-------------------

ou plus exactement :

Sujet	Prédicat	Objet
--------------	-----------------	--------------

Ainsi l'expression XML suivante :

```
<these>  
  <auteur>Michelle Ciron</auteur>  
</these>
```

devient

Cette thèse	a-pour-auteur	Michelle Ciron
-------------	---------------	----------------

¹ http://www.abes.fr/abes/documents/tef/recommandation/TEF_en_XML.html

² <http://tefsav.canalblog.com/archives/rdf/index.html>



Il s'agit d'attribuer à un sujet (cette thèse) un prédicat (a-pour-auteur) ayant une valeur particulière. Grâce à cette décomposition en un triplet, on identifie bien de quoi on parle (le sujet, la ressource), quelle propriété on lui attribue (le prédicat) et quelle valeur a cet attribut (la valeur, l'objet). Il reste à donner un nom à ces différentes choses.

Pour nommer ces choses, on utilise des URI (*Uniform Resource Identifier*), identifiants Web par définition uniques – une manière d'éviter les homonymies.

On peut ainsi construire une URI à partir du numéro national de la thèse (NNT) :

<http://www.dom.fr/1998LY020073>

Cette URI ne correspond pas forcément à l'URL à laquelle on peut obtenir une édition de cette thèse. C'est un nom propre ; pas nécessairement une adresse.

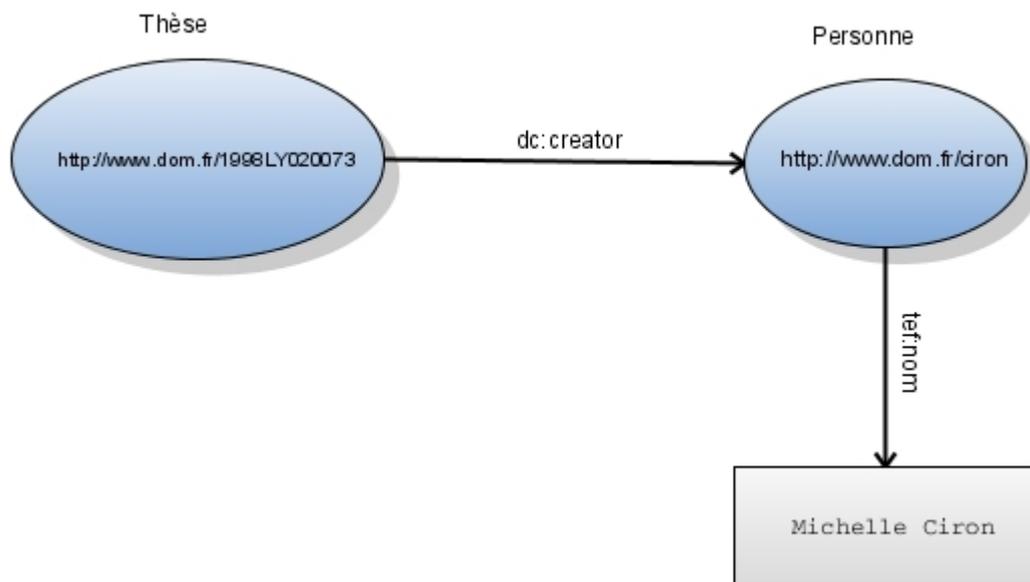
Pour nommer la propriété a-pour-auteur, inutile de chercher bien loin. Le Dublin Core possède <http://purl.org/dc/elements/1.1/creator>, qu'on peut abrégé en **dc:creator**.

Enfin, il reste le nom de l'auteur. L'effort de modélisation RDF nous aide à voir que l'auteur de la thèse n'est pas l'expression « Michelle Ciron », mais une personne réelle nommée Michelle Ciron. Notre expression XML devient en RDF :

<http://www.dom.fr/1998LY020073> **dc:creator** <http://www.dom.fr/ciron>
<http://www.dom.fr/ciron> **tef:nom** "Michelle Ciron"

La propriété **dc:creator** est une relation entre deux « choses », chacune étant désignée par une URI (on verra plus tard si on est vraiment condamné à affubler chaque humain d'une URI). La propriété **tef:nom** est une relation entre une chose (en l'occurrence une personne nommée <http://www.dom.fr/ciron>) et une valeur (en l'occurrence une chaîne de caractère, mais il pourrait s'agir d'autres types de valeurs comme un entier ou une date).

Les énoncés RDF sont fondamentalement des graphes (deux sommets reliés par un arc), qu'on peut représenter ainsi :



RDF possède plusieurs syntaxes, dont une syntaxe XML (qui n'est pas la plus limpide) :

```

<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:tef="info:tef#">
  <rdf:Description rdf:about="http://www.dom.fr/1998LY020073">
    <dc:creator rdf:resource="http://www.dom.fr/ciron"></dc:creator>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.dom.fr/ciron">
    <tef:nom>Michelle Ciron</tef:nom>
  </rdf:Description>
</rdf:RDF>

```

Voici le même graphe RDF sous une autre syntaxe (N-Triples) :

```

<http://www.dom.fr/1998LY020073> <http://purl.org/dc/elements/1.1/creator> <http://www.dom.fr/ciron> .
<http://www.dom.fr/ciron> <info:tef#nom> "Michelle Ciron"

```

Ce qui était implicite en XML devient explicite en RDF, et donc mieux compréhensible par des machines. L'objectif de RDF est précisément d'explicitier autant que possible le sens de l'information publiée sur le Web, pour faciliter l'interopérabilité, l'agrégation des données dispersées et enfin le raisonnement à partir de ces données. XML associe des mots tandis que RDF fait des phrases (vraies ou fausses). Or, raisonner, c'est générer de nouvelles phrases à partir de phrases données. Par exemple, imaginons que le Sudoc en RDF (un jour...) contienne l'énoncé suivant :

```

http://www.dom.fr/1998LY020073      dc:creator      http://www.dom.fr/ciron

```

et qu'on trouve ailleurs (un fichier d'autorité, une application administrative, le site Web de l'auteur...) :

<http://www.dom.fr/ciron>

owl:sameAs

<http://www.ailleurs.org/Lewis>

un raisonneur RDF pourra en déduire cet énoncé inédit :

<http://www.dom.fr/1998LY020073>

dc:creator

<http://www.ailleurs.org/Lewis>

En RDF, à côté de la description des faits (untel est l'auteur de cette thèse, cette thèse porte sur tel sujet...), il y a aussi des descriptions de niveau supérieur, qui – pour faire vite – porte sur les concepts qui servent à décrire les faits (tout créateur est une personne, une thèse n'a qu'un auteur, une thèse peut avoir plusieurs versions, telle valeur d'un thésaurus est équivalente à telle autre, etc.). Ce type d'information générale est exprimé dans des schémas. Ces schémas (RDF-S ou OWL) ne visent pas à valider la conformité des données comme c'est le cas des schémas XML, mais à **interpréter** les données RDF.

Pour rapide et approximative qu'elle soit, j'espère que cette présentation de RDF aidera à comprendre la suite de l'histoire. Dans les prochains articles, il va s'agir de réinterpréter en RDF les métadonnées TEF, en se posant à chaque fois ces questions : de quoi parle-t-on ? que lui attribue-t-on ? La valeur de cet attribut est-il un mot ou une autre entité ? Jusqu'à un certain point, ce travail a déjà été accompli par le [modèle de TEF](#)³.

Mais avant cela, on ne peut échapper à la question suivante : tout ça pour quoi faire ?

Pourquoi RDF ?

Pour répondre à cette question légitime, je propose une métaphore : XML, c'est de la pâte à modeler et RDF des légos.

Avec de la pâte à modeler, vous modelez exactement la forme que vous souhaitez. Vous êtes totalement libre. Vous pouvez aussi obtenir une forme qui réponde à des contraintes très précises, en utilisant un moule. Vous pouvez enfin la transformer facilement, la déformer pour remodeler un nouvel objet.

Avec les légos, vous êtes contraint de partir d'une forme imposée, minimale certes, mais rigide. Votre construction, aussi complexe soit elle, sera toujours composée de ces petites briques.

Avec XML, chacun peut imaginer sa propre structure de données, la faire épouser avec précision le corpus d'informations qu'on veut exprimer, et lui associer des outils de validation. D'où TEF en XML et ses schémas W3C XML Schema et Schematron. Le but est d'échanger

³ <http://www.abes.fr/abes/documents/tef/recommandation/modeleTEF.html>

avec certitude des données complètes et valides. Le langage XSLT permet de convertir sans peine TEF en d'autres formats XML (Dublin Core, UNIMARC en XML...).

Avec RDF, comme on le verra dans la suite de ce papier, la « notice » TEF éclate en une série d'énoncés élémentaires, qui « s'emboîtent » les uns dans les autres. L'avantage est que ces énoncés TEF de base peuvent non seulement s'emboîter entre eux, mais aussi avec d'autres énoncés qui se trouvent ailleurs sur le Web.

Pour filer la métaphore ludique : tu vas chez ton cousin pour Noël, en emportant tes légos et tes figurines en pâtes à modeler. Quelques jours plus tard, tu découvriras d'un côté une construction inédite composée de tes légos et de ceux de ton cousin, et de l'autre une bouillie de formes et de couleurs où chacun a du mal à trouver ses petits.

En publiant des métadonnées TEF en RDF, il devient possible d'articuler les informations qu'elles véhiculent avec d'autres informations portant sur l'auteur, ses autres publications, la description de son laboratoire, sa participation aux jurys et autres commissions d'évaluation, les auteurs cités... sans avoir à convertir TEF dans tous les formats qui encodent ces autres données. Grâce à RDF, TEF offre de multiples prises pour d'autres données. Le chemin est encore long qui permettra ces interconnexions de données, mais chacun doit en faire sa part, c'est-à-dire exposer ses données en RDF.

Je vois quatre justifications à cette RDFisation de TEF :

1. **Expérimenter** les technologies RDF pour voir si, grâce à elles, on peut faire de TEF des exploitations nouvelles et utiles (requêtes, navigation dans les données, visualisation graphique...).
2. **Exposer** les métadonnées des thèses françaises en RDF, en espérant qu'elles s'agrègent utilement à d'autres données RDF du Web.
3. Rendre plus **extensibles** ces métadonnées. Face à un besoin particulier, on n'a plus besoin de pester contre un élément manquant. Il suffit de le piocher dans un autre vocabulaire RDF existant (ou, au pire, de le créer *ad hoc*) et de le greffer aux énoncés TEF. Il suffit de parler de la même chose que TEF pour enrichir TEF. Il suffit que TEF parle de la même chose que d'autres énoncés pour enrichir les données des autres.
4. Fournir de la matière première à la **recherche** sur le web sémantique, qui doit s'appuyer sur de larges corpus de métadonnées riches pour progresser.

J'ajouterai la raison suivante : suite à la publication de son modèle abstrait (inspiré du RDF), le DCMI est en train de revoir les différentes syntaxes du Dublin Core. Or, la nouvelle syntaxe XML, censée refléter le modèle abstrait, sera probablement beaucoup plus sophistiquée que

l'ancienne et bien moins lisible. Ainsi, au moment où le Dublin Core en XML semble atteindre ses limites syntaxiques sans proposer de plus-value, il faut saisir cette occasion pour regarder de près ce que RDF peut apporter de plus. Le surcoût marginal du passage à RDF diminue tandis que sa plus-value marginale augmente avec la multiplication des données, services et outils RDF.

Nota Bene : on peut faire des légos en pâte à modeler, mais pas l'inverse. Pensez à la syntaxe XML de RDF.

De quoi on parle ?

RDF nous oblige d'abord à savoir de quoi on parle. Sur quoi portent les métadonnées ? Quel est le sujet de ces énoncés ? En Dublin Core comme en MARC, le sujet est implicite, ce qui peut masquer des confusions.

En moissonnant des notices DC, on trouve souvent ceci :

```
<description>
  <dc:creator>Michelle Ciron</dc:creator>
  <dc:format>PDF</dc:format>
  <dc:format>HTML</dc:format>
</description>
```

Sauf à supposer qu'une même ressource puisse être à la fois en PDF et en HTML, il faut en conclure que cette description a trois sujets implicites, et non un seul sujet. Elle dit d'une œuvre qu'elle a pour auteur Michelle Ciron et qu'elle possède deux éditions distinctes, une en PDF et une en HTML. Elle parle de trois choses. En RDF, on écrirait (en faisant vite) :

```
<rdf:RDF>

  <rdf:Description rdf:about="http://www.exemple/oeuvre">
    <creator>Michelle Ciron</creator>
    <edition rdf:resource="www.exemple.fr/edition1"/>
    <edition rdf:resource="www.exemple.fr/edition2"/>
  </rdf:Description>

  <rdf:Description rdf:about="www.exemple.fr/edition1">
    <format>PDF</format>
  </rdf:Description>

  <rdf:Description rdf:about="www.exemple.fr/edition2">
    <format>HTML</format>
  </rdf:Description>

</rdf:RDF>
```

La première description décrit l'œuvre. La seconde description décrit la première édition. La troisième description décrit la seconde édition.

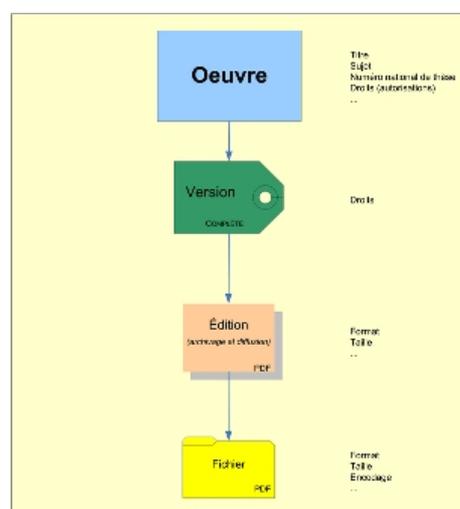
La propriété `edition` dit que l'œuvre a telle édition. Elle sert à relier différentes entités. La propriété `creator`, au contraire, semble relier une entité (l'œuvre) à une chaîne de caractère, un nom. On verra plus tard que c'est inexact.

En MARC aussi, le sujet est implicite. En général, une notice porte sur une édition (une manifestation dans le jargon FRBR). Mais une seule notice bibliographique peut porter sur plusieurs entités : une édition originale et son microfilm, une thèse et ses différentes éditions électroniques (PDF et HTML), une édition et chacun de ses volumes...

Voilà pourquoi TEF prend soin d'identifier les différentes entités bibliographiques sur lesquelles portent les métadonnées de thèse, en s'appuyant sur les FRBR :

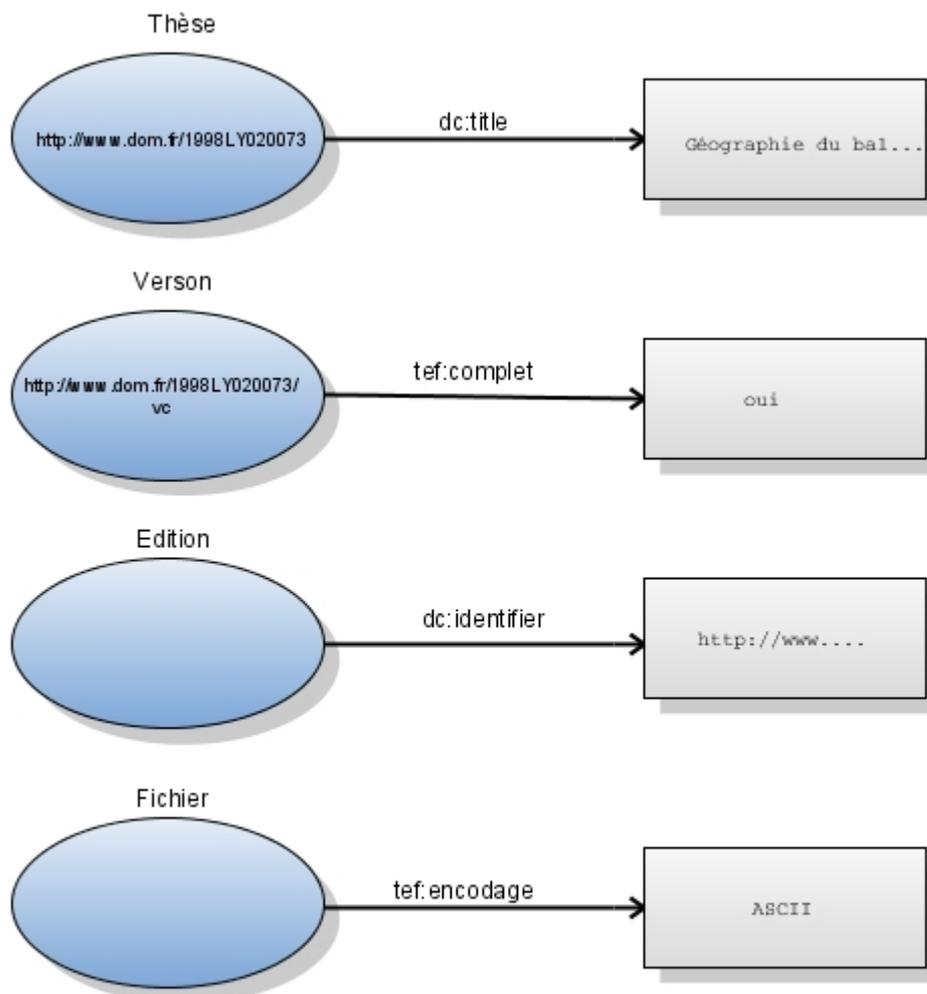
- La **thèse** en tant que telle – l'œuvre pour les **FRBR**.
- Les **versions** de la thèse – les **expressions** pour les FRBR. En fait, TEF ne s'intéresse qu'à la version validée par le jury (y compris les corrections post-soutenance) et les versions abrégées qu'on peut en faire pour des raisons techniques ou juridiques. Pour TEF, une version prés-soutenance ou remaniée pour un éditeur n'est pas une « version » de la thèse.
- Les **éditions** de la thèse – les **manifestations** pour les FRBR. Pour être précis, il faudrait dire : les éditions des versions de la thèse.
- Les **fichiers**. Les FRBR ne parlent pas explicitement de cette entité, mais on peut l'interpréter comme une **partie de manifestation**.

Chacune de ces entités a ses propres propriétés, ses propres métadonnées. C'est ce qu'illustre ce schéma du modèle TEF⁴ :



C'est aussi ce que montre ce graphe RDF :

⁴ <http://www.abes.fr/abes/documents/tef/recommandation/modeleTEF.html>



Les quatre principales entités du modèle TEF et leurs propriétés

Ce graphe appelle plusieurs remarques :

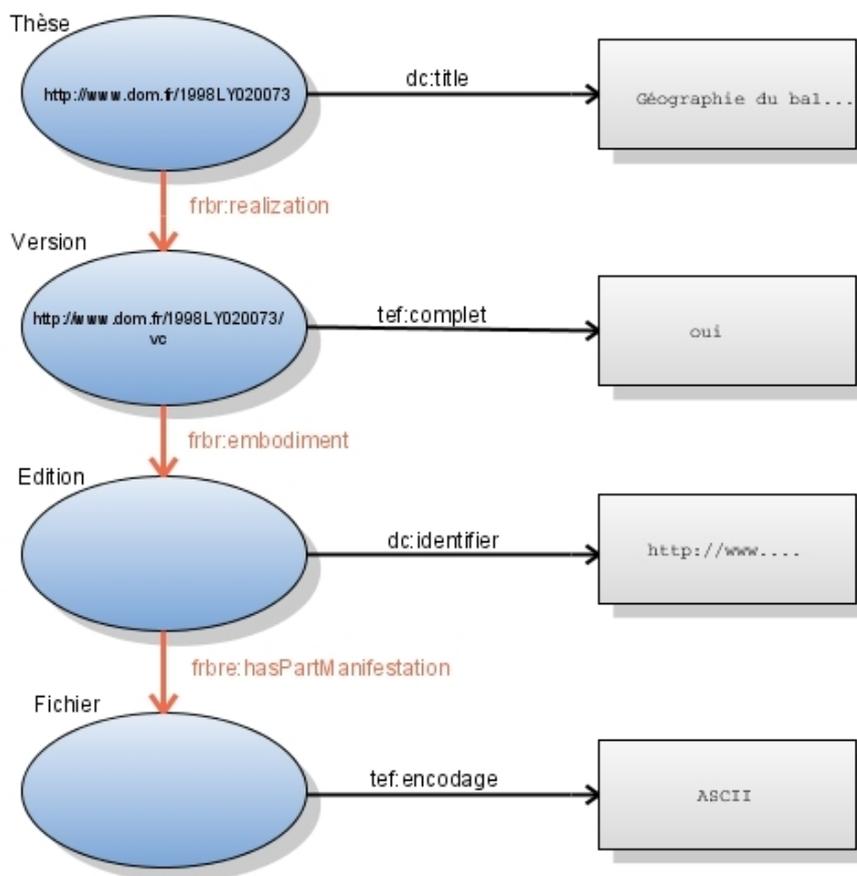
- Il ressemble beaucoup au schéma du modèle TEF. Moralité : quand on a fait un effort de modélisation, le passage à RDF est sans douleur (ou presque).
- La thèse et la version ont un nom (c'est-à-dire une URI), d'autres entités n'en ont pas. En RDF, les entités peuvent être nommées ou anonymes.
- Autant que possible, on utilise les propriétés Dublin Core ou celles d'un autre vocabulaire standard. C'est seulement en dernier recours qu'on crée une propriété TEF. C'est le cas de `tef:ecoleDoctorale`. C'est dans le schéma RDF-S (ou OWL) de TEF qu'on pourra dire que `tef:ecoleDoctorale` est une sous-propriété de `dc:contributor` ou que la classe `tef:These` est une sous-classe de la classe FRBR des œuvres. En d'autres termes, le schéma permet d'exprimer les relations entre le vocabulaire TEF et les vocabulaires standards. La relation à d'autres vocabulaires devient interne au vocabulaire décrit en

OWL ; elle n'est plus laissée à une annexe consacrée aux *mappings* ni à l'interprétation des utilisateurs.

- Les différentes entités ne sont pas encore reliées entre elles.

Il faut relier les entités TEF en explicitant leurs relations. Comme les entités TEF sont des cas particuliers d'entités FRBR, elles sont reliées entre elles comme n'importe quelles entités FRBR génériques. Il est donc inutile d'inventer une relation thèse-version quand la relation générique œuvre-expression suffit.

En 2005, Ian Davis et Richard Newman [ont modélisé](#)⁵ les FRBR en RDF. Nous pouvons réutiliser leurs relations FRBR ainsi :



**Les relations
entre les quatre principales entités
du modèle TEF**

C'est au moment où l'on souhaite relier les entités TEF qu'on voit l'intérêt de les nommer. Comme la version de la thèse est associée à un nom unique à l'échelle du Web

⁵ <http://vocab.org/frbr/core>

<http://www.dom.fr/1998LY020073> (URI), il importe peu que sa description soit dans le même fichier que la description de la thèse ou non. Certes, comme l'URI en question n'est qu'un nom propre, mais pas une adresse (ni pour la version elle-même, ni pour ses métadonnées), il resterait encore à découvrir l'endroit du Web où se trouve cette description RDF de la thèse, mais on voit toute la souplesse de la logique décentralisée de RDF.

Par contre, toutes les entités qu'on veut décrire n'ont pas d'URI (personnellement je n'en ai pas). Mais toutes n'en ont pas besoin. Est-il vraiment important de nommer fortement le fichier de notre exemple, au cas où quelqu'un d'autre voudrait le décrire ? RDF offre plusieurs mécanismes pour dire qu'une édition est composée d'un fichier sans lui attribuer une URI ?

Au lieu de distinguer la description de l'édition de la description de son fichier, en donnant une URI au fichier :

```
<Edition>
  <frbre:hasPartManifestation rdf:resource="http://www.dom.fr/1998LY020073/vc/ed1/fi1">
</Edition>
```

```
<Fichier rdf:about="http://www.dom.fr/1998LY020073/vc/ed1/fi1">
  <tef:encodage>ASCII</tef:encodage>
  <tef:formatFichier>XML 1.0</tef:formatFichier>
</Fichier>
```

on les emboîte :

```
<Edition>
  <frbre:hasPartManifestation>
    <Fichier>
      <tef:encodage>ASCII</tef:encodage>
      <tef:formatFichier>XML 1.0</tef:formatFichier>
    </Fichier>
  </frbre:hasPartManifestation>
</Edition>
```

Autre méthode :

```
<Edition>
  <frbre:hasPartManifestation rdf:nodeID="fi1"/>
</Edition>
```

```
<Fichier rdf:nodeID="fi1">
  <tef:encodage>ASCII</tef:encodage>
  <tef:formatFichier>XML 1.0</tef:formatFichier>
</Fichier>
```

Ici, on donne au fichier un nom provisoire (**rdf:nodeID**), qui ne vaut que dans le fichier RDF/XML. C'est ce qui distingue votre numéro INSEE du numéro qu'on vous donne pour

prendre place dans une file d'attente. Le premier est global et permanent ; le second local et éphémère.

Ces détails techniques sont importants car la question des URI fait couler beaucoup d'encre. C'est particulièrement vrai dès qu'il s'agit d'attribuer ce nom universel à des personnes. Nous aborderons cette question dans les deux prochaines sections.

Les personnes

Sauf dans la théorie littéraire française (pour la cellule de dégrisement, lire [Antoine Compagnon](#)⁶), ce ne sont pas des mots qui écrivent les livres, mais des personnes, des gens. On l'a vu plus haut. Si bien que, quand on écrit

```
<dc:creator>Michelle Ciron</dc:creator>
```

en fait, on parle à la fois d'un document qui a un auteur et d'un auteur qui a écrit un document.

Conclusion : les métadonnées ne parlent pas que des ressources documentaires. Entre autres entités, elles parlent aussi des personnes. Mais comment parler des personnes ?

Il existe plusieurs vocabulaires qui ont pour vocation de dire des choses sur les personnes, c'est-à-dire de leur associer des propriétés. Cette question est l'objet, en ce moment, d'une grande effervescence. Bien sûr, il y a une infinité de manières de parler d'une personne (comme de toute chose) et jamais aucun vocabulaire ne pourra couvrir cette diversité. Par contre, il serait utile de disposer d'un vocabulaire de base qui permette d'exprimer des notions comme le nom, la date de naissance, le domaine d'activité d'une personne... Libre aux autres vocabulaires d'y agréger d'autres informations (comme le [vocabulaire Bio](#)⁷ de Ian Davis ou [FRANAR](#)⁸).

Ce vocabulaire de base est [FOAF](#)⁹ (*Friend-of-a-Friend*). Comme son nom l'indique, la vocation première de FOAF est d'exprimer les relations interpersonnelles (X connaît Y), mais cela suppose d'en savoir un peu sur X et Y respectivement. Voici une description FOAF du créateur de FOAF :

```
<foaf:Person>
  <foaf:name>Dan Brickley</foaf:name>
  <foaf:homepage rdf:resource="http://rdfweb.org/people/danbri/" />
  <foaf:img rdf:resource="http://rdfweb.org/people/danbri/mugshot/danbri-small.jpeg" />
</foaf:Person>
```

⁶ <http://www.fabula.org/actualites/article2060.php>

⁷ <http://vocab.org/bio/0.1/>

⁸ <http://www.ifla.org/VII/d4/wg-franar.htm>

⁹ <http://www.foaf-project.org/>

Les propriétés que FOAF permet d'associer à une personne peuvent être regroupées en cinq catégories :

- Les propriétés basiques : nom, nom de famille, prénom, page Web...
- Les informations personnelles : travail, publication, connaissances, intérêts...
- Les comptes en ligne : compte MSN, Yahoo...
- Les projets et groupes auxquels la personne participe.
- Les documents et images liés à cette personne.

Voici l'URL complète du fichier FOAF de Dan Brickley : <http://danbri.org/foaf.rdf>

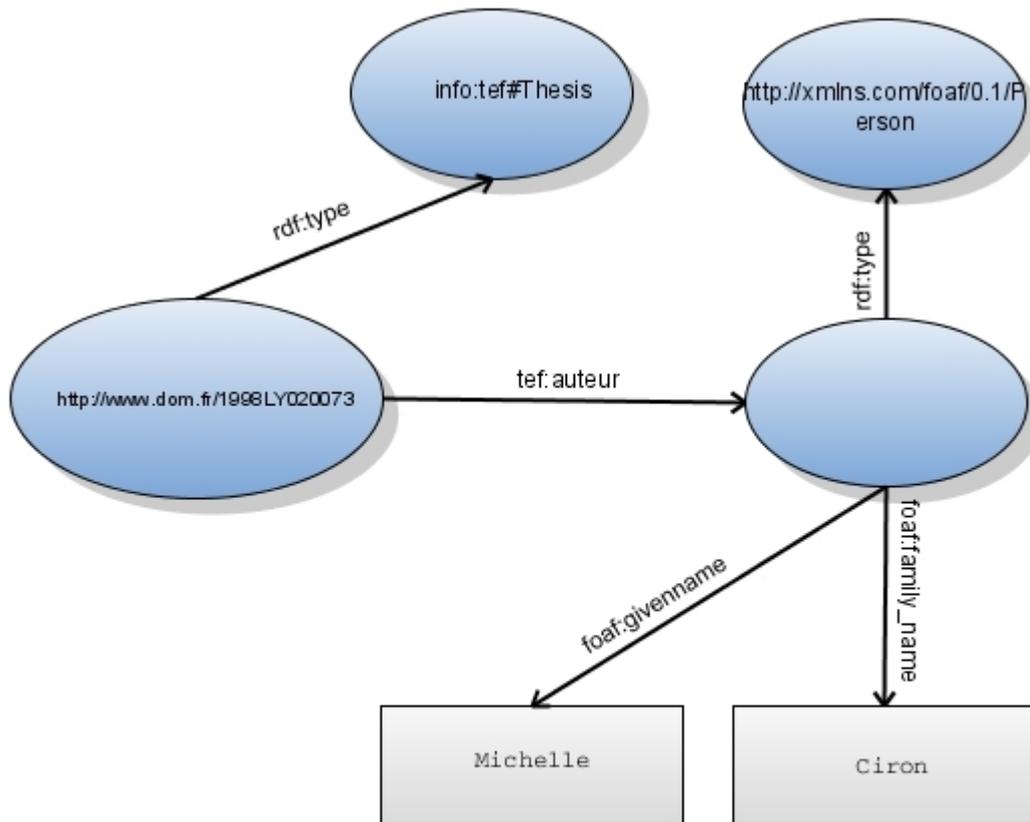
TEF parle de beaucoup de monde : l'auteur, son directeur, le jury, les rapporteurs, son université, son laboratoire...

FOAF est assez riche pour exprimer à peu près tout ce que TEF dit de ces personnes, physiques ou morales.

Pour écrire que Michelle Ciron est l'auteur, on écrira :

```
<tef:Thesis rdf:about=" http://www.dom.fr/1998LY020073 ">
  <tef:auteur rdf:nodeID="wyz"/>
</tef:Thesis>

<frbr:Person rdf:nodeID="wyz">
  <foaf:family_name>Ciron</foaf:family_name>
  <foaf:givenname>Michelle</foaf:givenname>
</frbr:Person>
```



**L'auteur de la thèse s'appelle "
"Michelle Ciron"**

Idem pour l'Université de soutenance ou le labo, sauf que ce ne sont pas des personnes physiques (instances de la classe **Person** de FOAF), mais des personnes morales (instances de la classe **Organization**). Les **Person** comme les **Organization** sont des **Agent**.

Pour le Jury, c'est un peu différent. Sans faire de haute scolastique, on conviendra que ce n'est pas chaque membre du jury qui approuve la thèse, mais le jury en tant que tel. Il se peut même que certains membres la désapprouvent ; un jury n'est pas toujours unanime. Pour autant, un jury n'est pas une organisation. Par chance, FOAF contient la notion de groupe (classe **Group**), sous-classe de **Agent**, ce qui signifie qu'un groupe peut *faire* quelque chose (comme valider une thèse).

On a désormais trois types d'entités en présence : une thèse, un jury, les membres du jury.

```
<tef:Thesis rdf:about=" http://www.dom.fr/1998LY020073 ">
  <tef:jury rdf:nodeID="jury"/>
</tef:Thesis>

<foaf:Group rdf:nodeID="jury">
  <foaf:member>
    <foaf:Person>
      <foaf:family_name>Simiand</foaf:family_name>
      <foaf:givenname>François</foaf:givenname>
      <tef:autorite rdf:nodeID="opujd"/>
    </foaf:Person>
  </foaf:member>
  <foaf:member>
    <foaf:Person>
      <foaf:family_name>Weber</foaf:family_name>
      <foaf:givenname>Max</foaf:givenname>
    </foaf:Person>
  </foaf:member>
</foaf:Group>
```

Les autorités

Quand on parle, on ne peut pas utiliser seulement des pronoms ou des démonstratifs, c'est-à-dire des noms qui ne valent que dans un contexte donné, ce qui est le sens de `rdf:nodeID`.

Ce qui permet d'agréger des descriptions RDF dispersées à travers le Web, c'est qu'elles parlent de la même chose. Mais encore faut-il qu'elles désignent cette chose de la même manière. D'où l'importance de donner un nom aux choses, un nom propre global qui permette d'identifier une chose de manière univoque. En RDF, on utilise les URI pour nommer les choses, mais cela ne veut pas dire que les URI ne nomment que des ressources documentaires accessibles sur le Web. Voici quelques exemples :

http://fr.wikipedia.org/wiki/Georg_Simmel désigne la page Web qui se trouve à cette page.

<info:inchi/InChI=1/H20/h1H2/p-1> désigne une substance chimique

<http://www.eionet.europa.eu/gemet/concept/9245> désigne le concept de transports maritimes dans un thésaurus environnemental.

<http://www.w3.org/People/Berners-Lee/card#i> désigne Tim Berners Lee.

Dans ce dernier exemple, une personne s'est donnée à elle-même une URI et l'a publiée pour inciter chacun à la nommer par ce moyen. Cette même personne (en passant, c'est l'inventeur

du Web et le promoteur de sa transformation en Web sémantique...) invite chacun à [se donner un nom de baptême sur le Web](#)¹⁰, en d'autres termes une URI.

C'est là que l'ambition du Web sémantique peut être mal interprétée. Son but n'est pas de construire un immense annuaire de toutes choses, y compris de toutes les personnes, vivantes ou non. Cet annuaire assignerait à chaque personne son URI, comme les pages blanches affichent votre numéro de téléphone. Une telle conception du Web risquerait de glisser vers une organisation centralisée, dans laquelle une agence serait habilitée à délivrer des URI.

Cette question d'un identifiant pour permettre de réidentifier sans équivoque une même personne à travers les différentes publications auxquelles elle a participé est une question bien connue des bibliothèques. C'est le principe du contrôle d'autorité. On relie toutes les notices bibliographiques des publications à une notice *ad hoc* qui identifie et décrit la personne en question. Chaque notice bibliographique mentionne bien le nom de la personne (et précisément ce nom peut varier d'une publication à l'autre !), mais surtout contient l'identifiant de la notice d'autorité de cette personne. Il est donc tentant d'exploiter ces millions de notices d'autorité pour identifier les personnes sur le Web, en construisant des URI sur la base de ces identifiants propriétaires, liés à une base (les [ppn](#) du Sudoc, pour les connaisseurs) ou par une autre méthode ([ISPI](#)¹¹). Par exemple, imaginons que les identifiants des notices d'autorité Sudoc deviennent des URI :

```
<foaf:Person rdf:nodeID="wyz">
  <foaf:family_name>Ciron</foaf:family_name>
  <foaf:givenname>Michelle</foaf:givenname>
  <tef:autorite rdf:resource="info:abes/aut/789578458"/>
</foaf:Person>
```

Mais qu'est-ce que ces URI désignent ? En fait, [info:abes/aut/789578458](#) ne désignerait pas un tel, mais sa notice d'autorité. Il faut distinguer les deux si on ne veut pas tomber dans l'absurdité selon laquelle une notice d'autorité parle d'elle-même. Il ne faut donc pas compter sur les notices d'autorité pour *tout simplement* baptiser les gens.

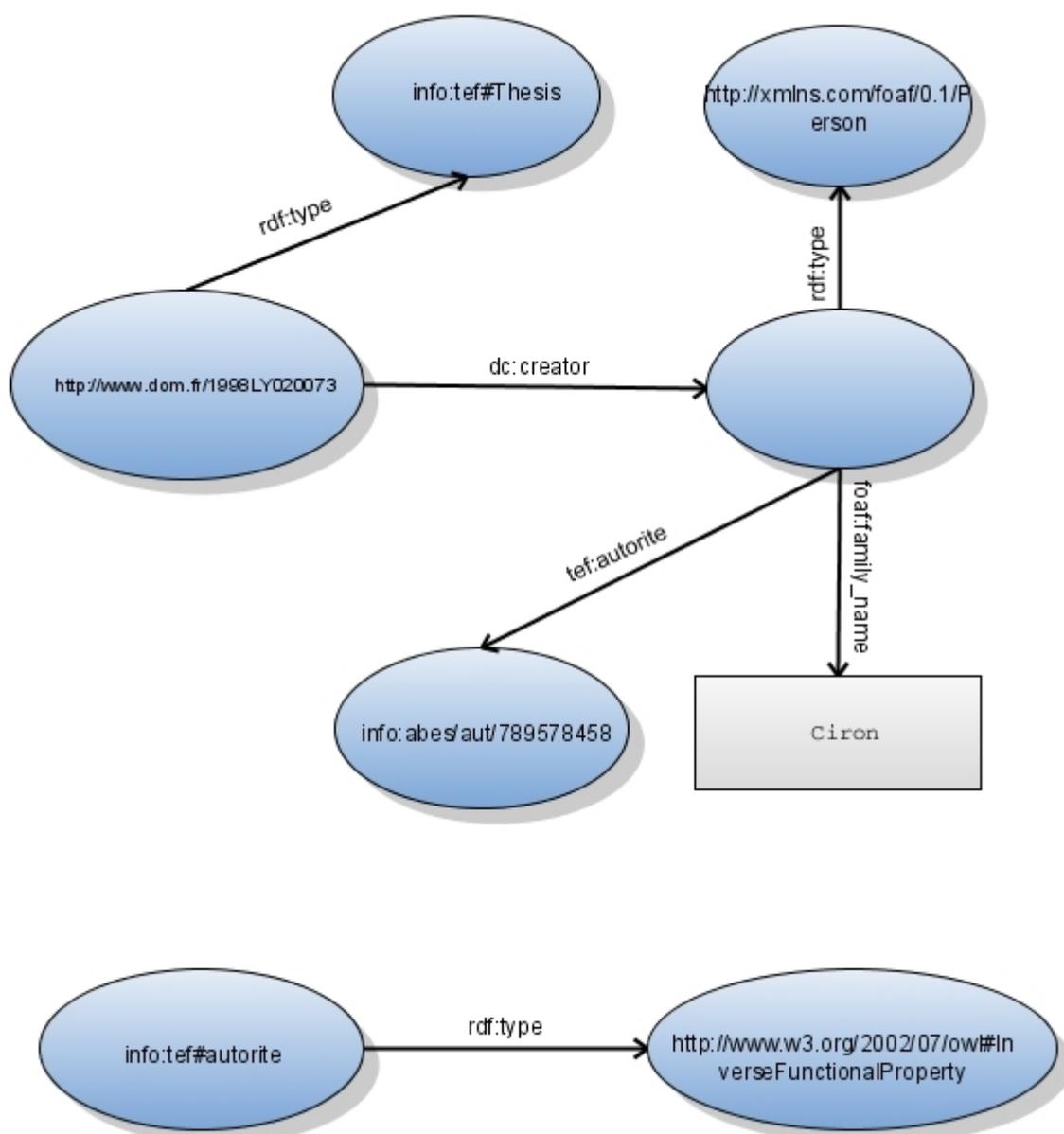
Par contre, ces mêmes notices peuvent exploiter certaines notions du web sémantique pour identifier *indirectement* les personnes. Ce qui est acquis, c'est qu'une notice d'autorité parle d'une et une seule personne (même si une personne peut avoir plusieurs notices d'autorité). "avoir telle autorité" est donc une propriété qui distingue une personne entre toutes les autres. C'est une propriété "*inverse-functional*" dans le vocabulaire d'OWL.

C'est une des notions les plus importantes du Web sémantique car elle permet d'identifier une même personne dans différents contextes descriptifs, même si elle n'est pas désignée de la même manière. Ainsi, quelle que soit la manière dont on la (les) nomme, si deux personnes

¹⁰ <http://dig.csail.mit.edu/breadcrumbs/node/71>

¹¹ <http://www.collectionscanada.ca/iso/tc46sc9/docs/sc9n429.pdf>

sont décrites en FOAF comme ayant la même adresse de messagerie (`foaf:mbox`), alors on peut en conclure qu'il s'agit de la même personne. La notice d'autorité remplit exactement le même rôle.



Il existe de nombreuses propriétés distinctives "*inverse functional*" : avoir tel numéro d'INSEE, avoir tel époux, être la mère biologique de... D'une manière générale, les technologies du Web sémantique autorisent une infinité de manières d'identifier une même chose. Ce n'est pas le cas de tous les systèmes de représentation de la connaissance, car certains postulent que deux noms différents désignent nécessairement deux choses différentes. Selon les contextes, les métiers, les données disponibles, il est bon qu'une même personne ou un même sujet puissent avoir des noms différents ou des propriétés distinctives différentes. Sinon, le Web sémantique exigerait une gestion centralisée de la connaissance universelle, ce qui serait à la fois dangereux, appauvrissant et impossible à réaliser. Pourtant, dans cet univers structuré mais ouvert et

décentralisé vers lequel évolue le Web, les notices d'autorité des bibliothécaires ont un rôle majeur à jouer, car s'y accroche une masse considérable d'information sur des personnes, des lieux, des concepts... Le défi est double : il faut mettre en ordre ces données d'autorité pour les rendre disponibles sur le marché de l'information RDF (et la modélisation FRANAR doit y contribuer) et réussir à mettre en relations ces données de bibliothécaire avec les données voisines venant d'autres horizons (sociétés de gestion de droits, encyclopédies comme Wikipedia¹², archives, musées, ...).

Il y aura toujours plusieurs manières d'identifier une personne, ne serait-ce que par différentes descriptions définies. Et même plusieurs identifiants, selon les métiers... Les identifiants de notices d'autorité ont leur rôle à jouer comme harpons d'information, même si ce n'est pas un identifiant direct pour les agents.

Il faut ajouter que la propriété **tef:auteur** possède elle-même une caractéristique discriminante. Elle n'est pas "*inverse functional*" car une même personne peut être l'auteur de plusieurs thèses. Par contre, elle est dite "*functional*" car une thèse ne peut avoir qu'un auteur. Cela signifie que le numéro national d'une thèse est un identifiant *indirect* de son auteur. En d'autres termes, si deux personnes ont écrit la même thèse, c'est qu'il s'agit de la même personne.

C'est pour cette seule raison que **tef:auteur** a été préférée à **dc:creator** (dont elle est une sous-propriété). **tef:auteur** est "*functional*", mais pas **dc:creator**. Dans la version XML de TEF, on a dû renoncer à réutiliser **dc:creator** et se résigner à créer **tef:auteur** pour des raisons de pure forme, de syntaxe : la syntaxe XML de Dublin Core ne permettait pas de décrire le créateur de manière structurée (nom, prénom, nationalité...).¹³ L'élément XML **dc:creator** ne pouvait contenir que du texte ("nom, prénom" par exemple), et non d'autres éléments XML. Désormais, en RDF, c'est pour des raisons de fond, de sémantique que l'on conserve **tef:auteur**.

De surcroît, le schéma OWL permet d'explicitier la relation entre ces deux propriétés, issues de vocabulaires différents :

```
<owl:ObjectProperty rdf:about="info:tef#auteur">
  <rdfs:domain rdf:resource="info:tef#Thesis"/>
  <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#FunctionalProperty" />
</owl:ObjectProperty>
```

¹² Voir le rapprochement entre la version allemande de Wikipedia et les notices d'autorité de la bibliothèque nationale allemande : <http://meta.wikimedia.org/wiki/Transwiki:Wikimania05/Paper-JV2>.

¹³ Ces limitations disparaîtront avec la nouvelle syntaxe XML du Dublin Core : <http://dublincore.org/documents/dc-xml/>. A quel prix ?

Dans l'ordre, cette description OWL de la propriété `tef:auteur` dit que :

- seules des thèses (au sens de TEF) peuvent avoir la propriété `tef:auteur` ;
- L'auteur d'une thèse est une personne (au sens de FOAF) ;
- `tef:auteur` est un cas particulier de `dc:creator`. En d'autres termes, tout auteur de thèse (au sens de TEF) est par définition un créateur (au sens de Dublin Core).
- `tef:auteur` est une propriété "*functional*". En d'autres termes, une thèse ne peut avoir qu'un auteur.

Les ressources externes

Si une thèse est, en tant que telle, un texte inédit, ce dernier peut contenir des parties qui ont déjà fait l'objet de publications. Toute thèse emprunte. Non seulement des idées, mais aussi des textes, des images, des tableaux de données, des schémas... Cette présence des "ressources externes" (dixit TEF) est d'autant plus naturel que la thèse sert à synthétiser l'état de l'art sur un sujet.

Certaines de ces ressources sont empruntées à d'autres auteurs, d'autres sont des autocitations. En général, l'auteur emprunte des extraits, même de "longs extraits", non prévus par le droit de la propriété intellectuelle ([sauf](#)¹⁴ sans les accords sectoriels et par la DADVSI à partir de 2009). Parfois, c'est une œuvre intégrale qui est incorporée à la thèse. Sauf cas de fraude ou de traduction, ces emprunts exhaustifs concernent la plupart du temps les thèses sur travaux, qui sont essentiellement des anthologies.

En METS, exprimer ce genre de relations est possible, mais un peu lourd. En RDF, il suffit de puiser dans le Dublin Core pour disposer de tous les outils nécessaires, à savoir la propriété `dcterms:hasPart` et son inverse `dcterms:isPartOf`.

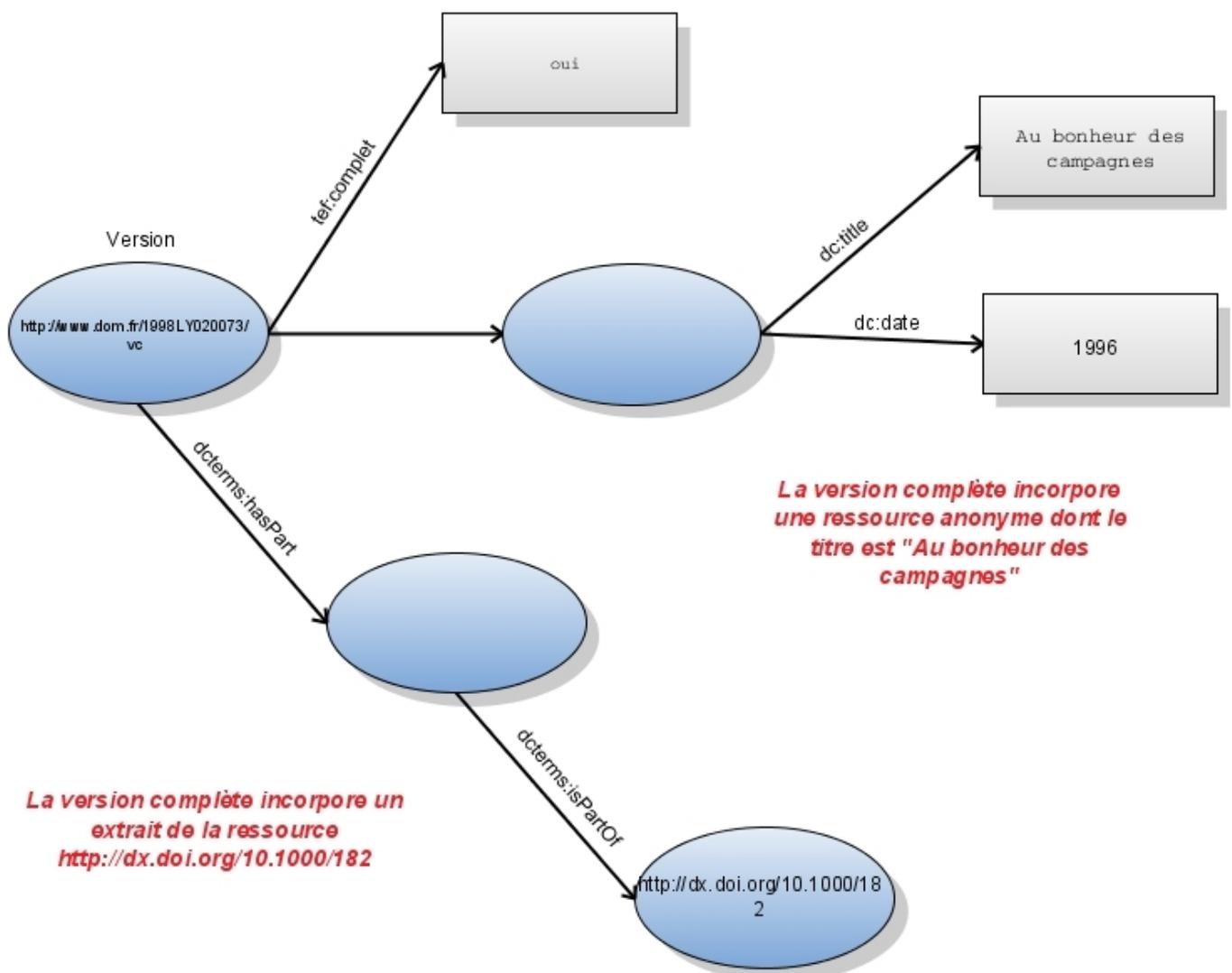
```
<Version rdf:about="http://www.dom.fr/1998LY020073/vc">  
  <dcterms:hasPart rdf:nodeID="rext"/>  
</Version>
```

Certes, ce vocabulaire générique du tout et des parties ne suffit pas à exprimer la spécificité des thèses sur travaux et des articles qui les composent. Une solution serait de créer une relation spécifique `tef:travauxContenus`, qu'on déclare comme étant une sous-propriété de `dcterms:hasPart`. A aucun moment on n'aura eu besoin de créer les classes d'objets *ad hoc* que sont dans TEF les "ressources tiers" et les "travaux". Une image empruntée n'est pas une "ressource tiers" en soi, mais une entité de type Image qui se trouve entretenir une relation

¹⁴ <http://tefsav.canalblog.com/archives/2007/02/25/4131295.html>

partie-tout avec une entité de type Thèse, comme elle peut aussi en avoir une avec une Ressource pédagogique, par exemple.

Enfin, comment décrire les ressources externes elles-mêmes ? TEF est très ouvert sur ce point et se contente de recommander [trois méthodes](#)¹⁵ : décrire la ressource en DC structuré ; la décrire avec la propriété `dcterms:bibliographicCitation` (bloc de texte idéalement extrait du document lui-même) ; identifier la ressource avec `dc:identifier`. Dans le contexte décentralisé de RDF, qui assume sans vergogne cette logique qui est déjà au cœur du Web, c'est évidemment cette dernière méthode qui est préférable. Pour mentionner une ressource externe, il devrait suffire de mentionner son URI pour permettre à une application exploitant les métadonnées de thèse d'en savoir plus sur la ressource externe en question, c'est-à-dire d'en découvrir quelque part des descriptions RDF. Dans notre exemple, le DOI d'un article sert de base pour lui associer une URI :



¹⁵ http://www.abes.fr/abes/documents/tef/recommandation/intro_desc_externes.html

Les Droits

La gestion des droits afférents à une thèse est une affaire complexe, pour plusieurs raisons :

- Comme l'a [montré](#)¹⁶ la SDBIS, la thèse est au centre d'une multitude de contraintes juridiques, que le système d'information de l'établissement de soutenance finira bien par exprimer un jour (on en est loin). Comme ces différentes contraintes juridiques n'ont pas les mêmes sources, elles peuvent être dispersées, soit à travers le SI, soit plus largement.
- Une thèse peut être une œuvre complexe, composite, au sens où elle peut inclure d'autres œuvres, en tout ou partie, on vient de le voir.

Si un système d'information doit prendre au sérieux jusqu'au bout les dimensions juridiques d'une thèse, il doit traiter toute une diversité d'informations. Il ne suffit pas de savoir que l'auteur autorise la diffusion de sa thèse, il faut encore s'assurer que celle-ci ne contient pas des ressources externes interdites à la diffusion, qu'un contrat industriel ne diffère pas la diffusion, il faut expliciter ce qu'on entend par "diffusion"... RDF est un bon moyen pour **exprimer** ces diverses informations dans un langage commun, alors qu'originellement elles peuvent être stockées sous des formes hétérogènes voire non structurées et donc difficiles à exploiter (un joli contrat signé sur un coin de table, par exemple). De tels langages d'expression des droits existent¹⁷, et sont en train d'effectuer leur mue vers les standards du Web sémantique, mais rien de très mûr n'existe à ce jour¹⁸.

Une fois que ces informations dispersées et hétérogènes auront été harmonisées, elles pourront être **agrégées**. On pourra lire des triplets RDF disant en substance :

L'auteur X autorise la diffusion de la thèse 1998LY020073

Le partenaire industriel Y refuse la diffusion de la thèse 1998LY020073

L'agrégation des triplets RDF permet de "contempler" toutes les informations se rapportant à la même thèse, voire de constater qu'elles peuvent entrer en conflit, mais elle ne nous dit pas ce qu'il faut faire de cette thèse.

¹⁶ <http://tefsav.canalblog.com/archives/2007/03/06/4221613.html>

¹⁷ Rapport de K. Coyle (2004), <http://www.loc.gov/standards/relreport.pdf>.

¹⁸ <http://rhizomik.net/semdrms/>

C'est là qu'à terme (quand ?) les technologies du Web sémantique devraient apporter une vraie plus-value. Elles permettent d'**exprimer** des informations sous un modèle commun, de les **agréger**, mais surtout d'en **déduire** de nouvelles informations.

Si l'auteur autorise la diffusion (contrainte) et que le partenaire l'autorise aussi (contrainte)..., **alors** la thèse est diffusable (c'est son état).

Si l'auteur autorise la diffusion de sa thèse (contrainte) et que celle-ci contient des ressources externes non diffusables (contrainte), **alors** la version complète est non diffusable (état) et la version expurgée est diffusable (état).

L'état juridique final de la thèse se déduit des différentes contraintes, un peu comme la direction d'un corps physique se compose à partir des différentes forces qui agissent sur lui.

Des métadonnées se déduisent d'autres métadonnées. Il serait dommage de devoir **saisir** des informations qui peuvent **se déduire**.

Pour qu'un jour les métadonnées de thèse puissent profiter de ces mécanismes de déduction, il faut que les ontologies du domaine juridique (en particulier de la propriété intellectuelle) se développent et que les métadonnées de thèse puissent s'exprimer dans le cadre de ces ontologies. En s'appuyant sur les métadonnées Dublin Core et sur la modélisation FRBR, les métadonnées TEF font un pari qui semble peu hasardeux, si on observe que des langages comme [ODRL](http://odrl.net/)¹⁹ ou des modèles comme [Indecs](http://cordis.europa.eu/econtent/mmrcs/indecs.htm)²⁰ ont reconnu dans les FRBR un fondement conceptuel incontournable.

Pourtant, cette perspective est lointaine. On peut aussi se demander si investir dans cette voie a quelque chance d'être rentable. Il est vrai qu'il n'est pas nécessaire d'attendre le Web sémantique pour écrire les programmes qui déduisent l'état juridique d'une thèse à partir des différentes contraintes juridiques qui s'appliquent à elle. C'est d'ailleurs sur la liste des développements programmés de [STAR](http://www.abes.fr/abes/page,428,star.html)²¹. Mais, au lieu d'effectuer ces développements sur la structure XML propre à TEF, il serait plus efficient de compter sur les règles logiques génériques incorporées dans OWL, l'ontologie FRBR et les ontologies juridiques. Pourtant, il serait naïf d'imaginer (même dans cinq ans) qu'une université investisse dans un système de gestion des droits appuyé sur RDF dans le seul objectif de rationaliser la gestion des thèses. Si un tel système s'avère rentable, ce sera pour gérer la documentation administrative ou pédagogique (e-Learning payant). Si la gestion des thèses en profite, ce sera par ricochet, en passager clandestin...

L'exemple de TEF en RDF ne va pas bien loin sur ce point. On imagine que l'auteur accepte de diffuser sa thèse sous une licence [Creative Commons](http://creativecommons.org/licenses/by/4.0/)²² (CC) très libérale, mais qu'il a incorporé dans son travail des images soumis à une licence CC plus restrictive. Alors, on en

¹⁹ <http://odrl.net/>

²⁰ <http://cordis.europa.eu/econtent/mmrcs/indecs.htm>

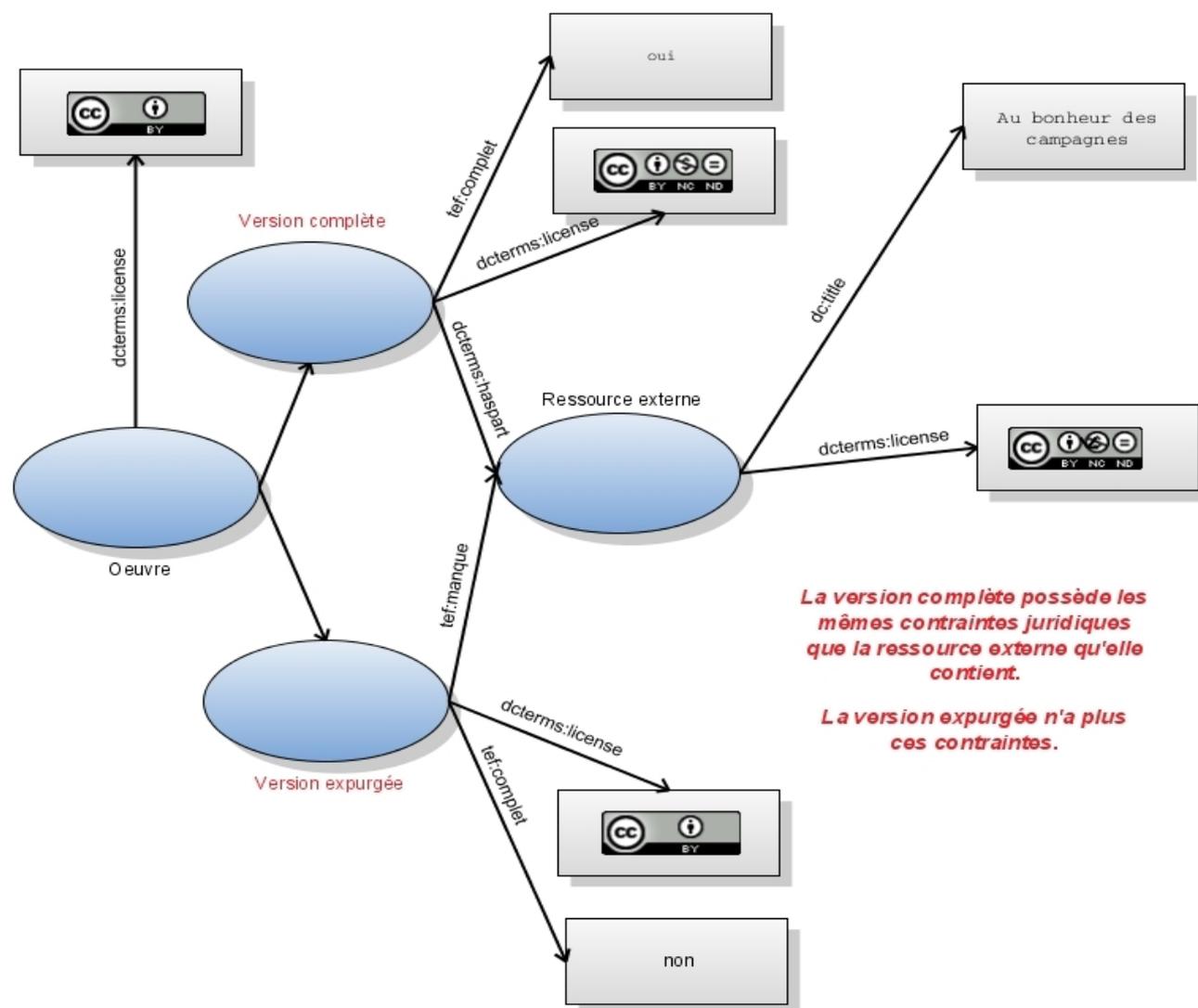
²¹ <http://www.abes.fr/abes/page,428,star.html>

conclut que la version complète doit se conformer à la licence plus restrictive, tandis que la version expurgée de la ressource externe peut bénéficier de la licence libérale. A noter que si les licences sont attribuées au niveau œuvre, c'est au niveau de la version (complète ou pas) que se décide la diffusion. Dans TEF, une thèse n'est pas diffusable ou non en soi, elle est diffusable ou non en fonction de son contenu, c'est-à-dire en tant que version.

Dans notre exemple, le code RDF est simplifié, car on ne décrit pas en détail les différentes clauses de la licence. On se contente d'écrire :

```
<dcterms:license rdf:resource="http://creativecommons.org/licenses/by/3.0/" />
```

On renvoie à une description RDF plus détaillée qui porterait sur la licence nommée par l'URI <http://creativecommons.org/licenses/by/3.0/>.

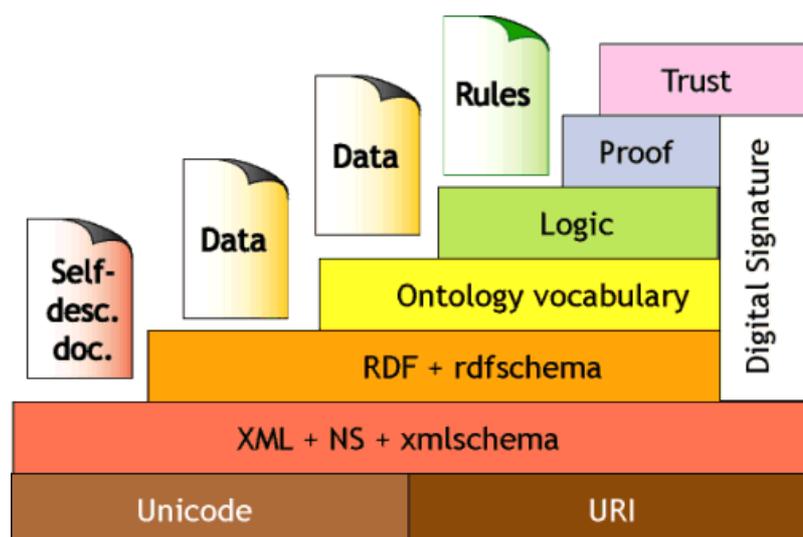


²² <http://creativecommons.org/>

A quoi sert le jury ? Comment faire confiance en RDF ?

Une thèse n'est pas un texte comme un autre. Son statut implique une notion d'évaluation, de certification, de validation voire de recommandation. Ces notions sont proches mais distinctes. Leur définition ne va pas de soi et peut varier selon le contexte. Néanmoins, elles ont le mérite de rapprocher les thèses d'autres types de documents, au premier chef les documents scientifiques en général. Pourtant, au lieu d'essayer de débrouiller ces notions (ce qui demanderait quelques thèses, précisément), je les rassemble sous une notion plus large : la notion de confiance. Outre la première raison évoquée à l'instant (un flou salutaire), je le fais pour deux autres raisons :

1. alors que la notion précise de validation scientifique permet de situer la thèse dans le contexte particulier de la communauté scientifique, une notion plus générale permet d'envisager les usages sociaux d'une thèse dans le contexte plus large du Web ;
2. la notion de confiance fait partie du paysage officiel du web sémantique : elle en occupe le dernier étage :



Quand le jury reconnaît à un étudiant le titre de docteur sur la base (sur la foi ?) du texte soutenu, à quoi, à qui accorde-t-il sa confiance ?

Ce n'est pas à la thèse que le jury déclare sa confiance, car la confiance suppose un élément d'incertitude... or chaque membre est censé l'avoir lue.

Ce n'est pas non plus la personne du thésard en elle-même qui s'avère digne de confiance - ça se saurait...

C'est plutôt le *thésard en tant que chercheur*, et plus précisément, *chercheur dans telle discipline*. La thèse, c'est un peu le rite de passage dans une discipline. Le thésard devient un pair – pas un ami.

Il existe une ontologie OWL nommée [Trust](#)²³ qui exprime exactement cela :

```
<foaf:Group rdf:ID="jury">
  <trust:trustsRegarding>
    <trust:TopicalTrust>
      <trust:trustSubject rdf:resource="info:ddc/22/fr/390"/>
      <trust:trustedPerson rdf:resource="#wyz"/>
      <trust:trustValue>8</trust:trustValue>
    </trust:TopicalTrust>
  </trust:trustsRegarding>
</foaf:Group>
```

Ce qui signifie : **le jury fait confiance à l'auteur de la thèse en tant que chercheur dans une discipline correspondant à la classe Dewey 390.**

En prime, la propriété `trust:trustValue` permet de graduer la confiance, certes à la louche (de 1 à 10). En théorie, voilà de quoi exprimer les mentions et les félicitations du jury. Dans les faits, la valeur de la mention est assez sujette à caution pour qu'on se dispense de l'exprimer.

La science est un réseau de réseaux basés sur la confiance : *peer-reviewing* des articles, fiabilité des résultats d'une expérience et des moyens offerts pour la reproduire, recensions, effets de réputation des revues, des labos, des directeurs... La soutenance de thèse est un épisode dans cette vie de réseaux. L'ontologie Trust est loin d'en refléter toute la subtilité, mais c'en est une première approximation.

Notez pour finir que, dans notre modélisation RDF, l'**événement** de la soutenance n'a pas été explicité comme entité à part. Ce qui nous intéresse, c'est la thèse et ses acteurs. La soutenance n'est qu'un moyen, un épisode, bien souvent une formalité.

Certaines ontologies, au contraire, voient dans les événements et autres entités temporelles (états, processus) des entités fondamentales. C'est le cas de [CRM-CIDOC](#)²⁴ qui sert à modéliser l'information muséale, mais promet d'aller plus loin, notamment par son [rapprochement avec le modèle bibliographique des FRBR](#)²⁵. Dans CRM, l'événement de création est au moins aussi important que le créateur ou la chose créée.

Dans le cas des thèses, c'est plutôt leur statut juridique qui pourrait inciter à faire de la soutenance une entité reconnue dans le modèle. En droit, les événements ont une place fondamentale : les actes juridiques (consignés dans le procès-verbal de soutenance) créent d'autres actes juridiques (délivrance du diplôme) et des états juridiques (titre de docteur). Dans le cas des métadonnées de thèse, on pourrait défendre ce détour conceptuel par la notion d'événement car, à l'échelle d'un établissement d'enseignement supérieur et de recherche, le

²³ <http://trust.mindswap.org/trustOnt.shtml>

²⁴ <http://cidoc.ics.forth.gr/>

²⁵ http://cidoc.ics.forth.gr/frbr_inro.html

système d'information est global en ceci qu'il agrège notamment gestion documentaire et gestion administrative. Et précisément, les thèses sont à cheval sur les deux. De plus, même d'un strict point de vue bibliographique, parler explicitement de l'événement Soutenance permettrait de mettre en lumière voire de dépasser la fiction traditionnelle en catalogage des thèses qui fait passer la date de soutenance pour la date de création ou de publication du document.

De ce qui précède, j'en conclus qu'il n'y a pas qu'une seule manière de décrire et modéliser la réalité, en l'occurrence la réalité *Thèse de doctorat*. Toute ontologie est appliquée, appliquée à un domaine mais aussi inscrite dans un contexte, soumise à des objectifs. Un anthropologue des sciences ne modélisera pas la Thèse comme l'exige la construction d'un système d'information global d'université. Pourtant, on peut maintenir le postulat qu'il n'existe qu'une seule réalité à décrire. Ce postulat n'est pas d'ordre philosophique ; il est méthodologique. Il donne un sens à l'effort d'agréger les triplets RDF divers et variés qui portent sur les thèses et les entités qui gravitent autour.

Conclusion

Officiellement, TEF demeure un format de métadonnées en XML. Mais, l'essentiel de TEF n'est pas dans sa peau XML ; il est dans son modèle. Et il apparaît que la forme RDF est le langage qui épouse le mieux les contours de ce modèle. Plus exactement, RDF permet de **préciser** le modèle TEF. Au bas de la page Web sur le modèle TEF, on trouve cette note :

Ce modèle conceptuel sera complété, affiné et formalisé à l'occasion de la conception du schéma RDF-S/OWL pour TEF. En l'état, il suffit à expliciter les entités principales de TEF et à justifier la forme qu'il prend en XML.

Le travail est en cours. Le passage de TEF en RDF est loin d'être achevé, d'un point de vue technique – *a fortiori* d'un point de vue normatif – je rappelle que les pages qui précèdent n'engagent nullement le groupe AFNOR auteur de TEF.

Ce passage de TEF en RDF dépend aussi d'autres avancées : expression du Dublin Core en RDF, expression des FRBR en RDF, réutilisation des autorités à la sauce Web sémantique, modélisation des droits, travaux internationaux sur les normes de catalogage RDA et leurs [relations](#)²⁶ avec le Dublin Core et le Web sémantique...

Le passage en RDF n'en est pas un nouvel avatar exotique, mais la banalisation de TEF, son inscription dans un écosystème international de métadonnées structurées (« Pléonasmes ! »), réutilisables, détachées du seul domaine documentaire. On peut aussi prononcer l'expression

²⁶ <http://dublincore.org/librarieswiki/DataModelMeeting>

"Web sémantique", pour aller plus vite... mais on risque aussi bien de perdre alors son temps en querelles de mots.

Bibliographie

TEF

Recommandation : <http://www.abes.fr/abes/documents/tef/recommandation/>

Exemples XML : <http://www.abes.fr/abes/documents/tef/recommandation/exemples.html>

RDF

Introduction officielle : <http://www.w3.org/TR/rdf-primer/>

Introduction par Tim Bray : <http://www.xml.com/pub/a/2001/01/24/rdf.html>

Dublin Core

Site officiel : <http://dublincore.org/>

Modèle : <http://dublincore.org/documents/abstract-model/>

Encodage RDF : <http://dublincore.org/documents/2007/06/04/dc-rdf/>

FRBR

Site officiel : <http://www.ifla.org/VII/s13/wgfrbr/index.htm>

Texte officiel (fr) : <http://www.ifla.org/VII/s13/frbr/frbr-fr.pdf>

FRBRoo, nouvelle modélisation des FRBR (version de travail) :

http://www.ifla.org/VII/s13/wgfrbr/FRBR-CRMdialogue_wg.htm

FRBR en OWL (Ian Davis, Richard Newman) : <http://vocab.org/frbr/core>