

# TEF, métadonnées pour les thèses françaises

**L**a recommandation TEF (Thèses électroniques françaises) normalise les métadonnées des thèses dans le cadre du circuit national des thèses. Elle est le fruit de plusieurs années de travail au sein du groupe d'experts AFNOR CG46/CN357/GE5, qui regroupe des universités, des grandes écoles, l'INIST, l'ABES... Après une phase d'appel à commentaires qui s'est achevée le 12 mai 2006, TEF sera bientôt officiellement en vigueur.

## TEF, un des outils du nouveau circuit des thèses numériques

Le dispositif national de promotion des thèses électroniques distingue le niveau local du niveau national. Au niveau local, les établissements sont principalement responsables de l'encadrement du travail de recherche et de rédaction de la thèse, de la préparation de la soutenance, de la validation de la version officielle de la thèse et enfin de la délivrance du diplôme. Naturellement, à côté de ces missions qui sont de leur responsabilité exclusive, les établissements de soutenance peuvent veiller aussi à la diffusion de la thèse, à son signalement ou encore à sa conservation. Toutefois, ces dernières opérations ne relèvent *pas exclusivement* du niveau local. Elles peuvent être prises en charge par d'autres acteurs, à vocation nationale ou non. En particulier, le signalement exhaustif des thèses françaises est une des missions confiées au catalogue collectif SUDOC, qui fait office de bibliographie nationale des thèses. De même, le CINES a une mission nationale d'archivage pérenne des thèses électroniques. D'une manière générale, diffusion, signalement et archivage des thèses sont des opérations qui peuvent être prises en charge par une pluralité d'acteurs.

C'est précisément cette diversité des acteurs qui donne à TEF sa raison d'être et qui en fixe le périmètre. Pour mener à bien leurs missions, tous ces acteurs ont besoin de disposer de métadonnées structurées. TEF est le format d'échange par lequel l'établissement de soutenance communique à ses partenaires les

métadonnées dont ils ont besoin, et seulement celles-ci. Au cours du traitement de la thèse, l'établissement de soutenance peut avoir recueilli ou généré des informations qui n'ont d'intérêt que local (en relation avec la soutenance, par exemple). TEF n'est pas concerné par ce genre de métadonnées (même si sa structure est suffisamment extensible pour les intégrer, en cas de besoin).

## Quelles métadonnées ?

Parmi les métadonnées TEF, on peut distinguer les métadonnées descriptives et les métadonnées de gestion.

Les métadonnées descriptives sont de nature bibliographique. Elles sont compatibles avec les normes de catalogage et le format UNIMARC, si bien que STAR pourra convertir automatiquement et sans perte les « notices » TEF en notices UNIMARC versées au Sudoc.

Parmi les métadonnées de gestion, on distingue les métadonnées administratives, les métadonnées de droit et les métadonnées de conservation.

Les métadonnées administratives concernent la thèse en tant que diplôme – et non en tant que simple document. Elles mentionnent le jury et ses membres (facultatif), la discipline, la date de soutenance, l'autorisation de diffusion...

Les métadonnées de droits doivent informer le diffuseur, le conservateur ou l'utilisateur final de la thèse sur les actions qu'il peut effectuer (lecture, reproduction, impression, réutilisation...). Ces métadonnées mettent en jeu les autorisations de l'auteur ou du chef d'établissement, mais aussi, éventuellement, celles des ayants droit sur les ressources réutilisées dans une thèse (images, son, cartes, textes...).<sup>1</sup> Ces enjeux de propriété intellectuelle ne peuvent se résumer à une simple alternative oui/non. TEF doit pouvoir exprimer des informations assez fines, qui reflètent certaines des clauses du droit ou des contrats en vigueur et qui permettent d'automatiser certaines opérations (la levée de la confidentialité, par exemple).

Enfin, la conservation à long terme des documents numériques nécessite des

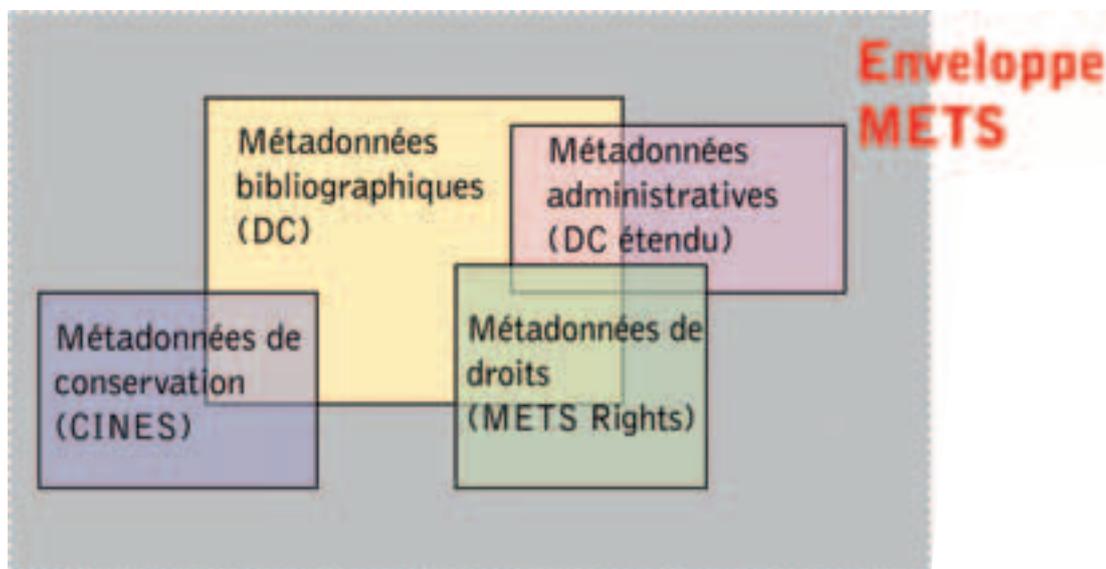
métadonnées spécifiques, qui sont en cours de normalisation au niveau international.<sup>2</sup> TEF a pris le parti de ne rendre obligatoires que les métadonnées que le CINES demande à ses partenaires, même si la structure TEF permet d'y ajouter d'autres métadonnées techniques de conservation si nécessaire.

## TEF en XML

C'est seulement après avoir défini une modélisation conceptuelle de la thèse et de ses métadonnées, en s'appuyant sur les FRBR, et avoir analysé les besoins en métadonnées dans le contexte national qu'il a été possible de spécifier un format. Ce format doit permettre aux métadonnées TEF de voyager entre les applications informatiques et d'être exploitées par celles-ci.

Dans le format TEF, les métadonnées sont structurées en XML. XML est aujourd'hui le langage incontournable pour structurer des données de manière souple, rigoureuse et interopérable. Par sa souplesse, XML laisse chacun définir son propre format, conforme à ses besoins. Par la rigueur des schémas XML, TEF est aussi bien un format d'échange qu'un outil de validation : la recommandation livre les outils qui permettent de contrôler très finement qu'une notice est conforme ou non. Enfin, si XML permet d'inventer son propre langage, au risque d'une babélisation des formats, il permet aussi à une communauté de réutiliser les formats des autres, quitte à les enrichir ou les amender. Le groupe TEF s'est efforcé de concilier ces trois principes : imaginer un format XML qui reflètent le cadre français de gestion des thèses numériques, garantir la validité des métadonnées échangées et réutiliser autant que possible les standards actuels.

Pour ce faire, le format TEF s'appuie sur les standards Dublin Core et METS. Dublin Core sert de noyau et METS d'enveloppe. A chaque fois qu'il est possible, les métadonnées TEF sont exprimées par des éléments DC, en respectant leur sémantique et leur syntaxe XML. C'est seulement quand le vocabulaire Dublin Core manque de précision ou de complétude qu'on réutilise d'autres vocabulaires (comme METS



Rights pour exprimer les droits ou MADS pour les autorités). En dernier recours, on crée des éléments inédits, propres à TEF. METS est un schéma XML qui sert à rassembler et organiser toutes sortes de métadonnées relatives à des objets numériques complexes (périodiques électroniques, sites web). Une «notice» TEF est en fait un document METS obéissant à certaines règles particulières. TEF est donc moins un nouveau format de métadonnées qu'une convention particulière pour utiliser la structure METS – ce qu'on appelle un profil METS.

## Comment produire du TEF ?

On ne produit pas du TEF comme on produit une notice MARC. Dans l'esprit de ses concepteurs, TEF est un format d'échange qui permet de rassembler des métadonnées diverses, aux origines et aux fonctions variées. Au sein du système d'information global d'un établissement, les métadonnées de thèses peuvent être extraites du document lui-même (notamment de la page de titre), d'une application administrative comme APOGEE, d'une application dédiée à la gestion des

thèses, d'un formulaire en ligne rempli par l'auteur... Il ne faut pas imaginer le catalogueur seul devant un formulaire TEF – et encore moins devant son éditeur XML. Le recueil des métadonnées de thèse, pour être efficace, nécessite de coordonner différentes applications et différents métiers au sein d'un établissement. TEF peut contribuer à rationaliser ce processus interne à l'établissement. On peut aussi ne l'utiliser qu'à titre de format d'échange, pour l'export vers STAR par exemple.

### Conclusion

Même si elle répond à un cahier des charges très précis, fixé par les textes réglementaires nationaux qui encadrent les thèses, la recommandation TEF obéit à un souci d'interopérabilité (réutilisation de METS, de Dublin Core..., lien à des fichiers d'autorité). Pour les établissements de soutenance, cette interopérabilité doit faciliter la conversion de leur format interne vers TEF. Mais, au-delà des échanges prévus par le dispositif national, TEF est un bon moyen pour mettre à disposition sur le Web des métadonnées riches, normalisées et réutilisables sur les thèses françaises. Ces métadonnées pourront être exploitées par des outils info-

métriques ou à des fins de pilotage de la recherche. Elles pourront rencontrer d'autres métadonnées, ce qui rapprochera les thèses françaises d'autres ressources : thèses étrangères, articles, données scientifiques, brevets, rapports, projets scientifiques. Les métadonnées sont faites pour voyager et faire des rencontres. Et plus si affinités...

Yann Nicolas

 nicolas@abes.fr

Site web de la recommandation TEF : <http://www.abes.fr/abes/documents/tef/index.html>  
 Blog TEF : <http://tefsav.canalblog.com/>

<sup>1</sup>Cf. les accords récemment signés par le MENESR sur [l'utilisation des œuvres et objets protégés par la propriété littéraire et artistique à des fins d'illustration des activités d'enseignement et de recherche](http://www.abes.fr/abes/documents/tef/index.html). Ces accords sont disponibles à cette adresse : [http://droिताuteur.levil-lage.org/spip/article.php3?id\\_article=44](http://droिताuteur.levil-lage.org/spip/article.php3?id_article=44).

<sup>2</sup>Cf. les travaux du groupe international PREMIS (<http://www.oclc.org/research/projects/pmwg/>).