

# Formuler la connaissance

Par Dominique VAUQUIER, Praxademia S.A.S.

Relecteurs et contributeurs : Bruno SEZNEC, Thierry BIARD (Ambesas)

## Table des matières

1° Des textes : leur classement .....	1
2° Des termes : leur élucidation par la terminologie .....	2
3° Des concepts et des relations : les ontologies et le Web.....	3
4° Des connaissances : leur modélisation et l'automatisation.....	4
Conclusion .....	6

La gestion de la connaissance revêt des formes multiples, autant que la connaissance elle-même. Étant donné les enjeux qu'elle représente, il vaut la peine de recenser les différentes possibilités de capturer, exprimer, conserver et faire fructifier la connaissance. Ce papier n'a pas l'ambition d'épuiser le sujet, ni même de présenter toutes les solutions envisageables. Il s'intéresse à l'expression – plus ou moins formelle – des connaissances, laissant de côté les questions liées à leur acquisition, leur gestion et leur transmission. Le but est d'indiquer les solutions qui sont à la portée des entreprises pour exprimer leurs connaissances. Nous montrerons qu'elles s'échelonnent dans un continuum : textes, termes, concepts, modèles. Ce continuum permet de fonder une démarche progressive de gestion des connaissances, mettant en regard des possibilités d'automatisation croissante.

## 1° Des textes : leur classement

Tout d'abord, la connaissance apparaît sous forme de textes. Quand elle n'est pas encore formulée, les techniques d'enquête et d'entretien finiront par la passer à l'état de texte. Une gestion des connaissances qui s'arrête à cette étape rend des services, mais se heurte vite à une limite : il s'agit d'ordonner ces textes, de les rentrer dans des classements qui permettront leur exploitation efficace. À cette fin, nous devons bâtir une **classification** telle que celles, bien connues, en usage dans les bibliothèques<sup>1</sup>. Presque toujours, notre culture et – dans le cas des bibliothèques – les contraintes physiques de rangement conduisent à une classification unique et hiérarchique. Comme nous en faisons l'expérience tous les jours, ce n'est pas la façon la plus naturelle pour organiser des connaissances et préparer leur exploitation. Leur ensemble bouillonnant et sans cesse renouvelé ne se coule pas aisément dans un moule figé et définitif. Au minimum, nous avons intérêt à adopter non pas une, mais plusieurs **taxinomies**<sup>2</sup>, chacune se plaçant dans une perspective particulière.

---

<sup>1</sup> Par exemple : la classification Dewey, la classification décimale universelle, la classification de la Bibliothèque du Congrès américain, le classement Freinet.

<sup>2</sup> Taxinomie : classification d'éléments.

## 2° Des termes : leur élucidation par la terminologie

À partir de là, on le sent bien, les choses se compliquent : Comment identifier ces taxinomies ? Comment faire en sorte d'en limiter le nombre ? Comment vérifier qu'elles n'entrent pas en contradiction ? Etc. De plus, les textes, à commencer par les titres et les résumés, devront être interprétés pour que l'on puisse leur associer les taxons<sup>3</sup> appropriés. La clarification du vocabulaire est donc un préalable à l'établissement d'une classification. Ici, la **terminologie**, en tant que discipline<sup>4</sup>, vient à notre secours. L'analyse terminologique est un passage obligé pour les étapes qui suivent et qui vont vers plus de formalisation. En elle-même, elle produit déjà des résultats dont les parties prenantes perçoivent instantanément la valeur. Notamment, dans la discipline de l'architecture « métier »<sup>5</sup>, la terminologie d'entreprise, en tant que livrable clef, est reconnue comme le pilier des approches de la connaissance et de la convergence entre plusieurs entités.

La terminologie, pratiquée dans l'entreprise, apporte les bénéfices suivants :

1. Elle démontre une « écoute active », à travers ses actions de recueil et d'analyse de corpus de textes.
2. Elle dépassionne les querelles de termes ou de définitions<sup>6</sup> en étudiant les usages et en apportant un appareillage suffisamment précis pour objectiver la langue.
3. Elle donne les règles pour arriver à de bonnes définitions, c'est-à-dire des définitions aussi simples que possible, qui saisissent l'essence d'un concept.
4. Elle nous guide dans l'analyse des relations entre les termes, qui forment la trame de la connaissance<sup>7</sup>.

On pense, pour la terminologie d'entreprise, à la forme du dictionnaire, éventuellement « de référence ». Si l'on y ajoute la multiplicité des sources, l'étude des différents usages et, surtout, les relations entre les termes, on passe à la forme du **thesaurus**. C'est sous cette forme que la terminologie d'entreprise sera la plus utile et qu'elle vivra le mieux, sur le long terme. Nous disposons, aujourd'hui, d'outils efficaces pour développer et administrer des thesaurus que nous pouvons publier, par exemple sur l'intranet de l'entreprise.

La terminologie nous aide à établir de meilleures classifications, notamment pour étiqueter nos documents. Elle nous aide à aller plus loin aussi, dans l'expression de la connaissance. En effet, elle organise le matériau qui servira d'entrée aux efforts de modélisation.

---

<sup>3</sup> Taxon : entrée dans une taxinomie.

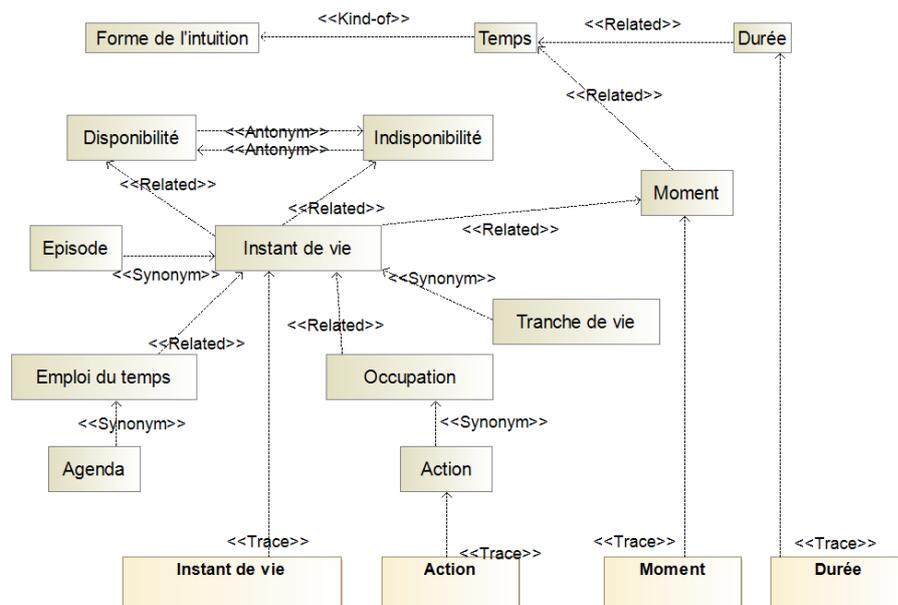
<sup>4</sup> Au sens 3 donné par le Grand Robert : « Ensemble des activités théoriques et pratiques portant sur les systèmes de notions et leurs désignations au moyen de noms organisés en système (termes) ».

<sup>5</sup> Ou plutôt la *Business Architecture* puisque c'est dans le monde anglo-saxon que cette discipline s'est autonomisée et développée au sein des entreprises.

<sup>6</sup> La langue est un capital symbolique. Tout capital se transforme en pouvoir. D'où, parfois, des débats enflammés qui surgissent autour des termes ou de leur normalisation.

<sup>7</sup> Les relations entre les termes ont été analysées et codifiées depuis longtemps.

Figure 1. Exemple de diagramme terminologique



Ce diagramme a été réalisé avec un outil de modélisation qui offre une fonctionnalité de « dictionnaire ». Il présente des termes du champ sémantique de la temporalité, reliés par des relations types. Les rectangles qui apparaissent en bas désignent des classes, éléments de modélisation formalisant les concepts.

### 3° Des concepts et des relations : les ontologies et le Web

Avant cela, il convient d'aborder une autre étape possible, celle de la constitution des **ontologies**. Une ontologie (au sens informatique) est un modèle qui relie des concepts, exprimés par des termes. Terminologie et ontologie sont des approches jumelles, nourries de la même tradition théorique. Un trait les distingue, au moins dans les pratiques : l'importance de l'outillage dans le cas des ontologies<sup>8</sup>. D'ailleurs, les ontologies peuvent s'exprimer dans des langages standardisés comme RDFS<sup>9</sup> et OWL<sup>10</sup>. Ces langages forment le socle de ce qu'il est convenu d'appeler le Web sémantique. L'avantage réside ici dans le partage des informations, lesquelles peuvent se présenter sous la forme de propositions complexes. Un langage comme SPARQL<sup>11</sup> permet de formuler des questions et d'interroger une ou plusieurs bases de triplets RDF (on dit « *triplestore* »).

On voit ainsi ce qui distingue les deux courants :

- Une terminologie apparaît pour elle-même, avec ses termes, ses réseaux de termes (éventuellement dessinés) et ses définitions. Publiée, elle sert comme sert un dictionnaire, soit en élucidant des usages, soit en normalisant un vocabulaire.
- Une ontologie se construit de la même façon, mais ses représentations, peut-être plus complexes, sont réservées aux concepteurs et non destinées au public. Son contenu, au contraire, est mis à disposition et accessible sur Internet ou une forme apparentée, par le truchement de solutions informatiques. La technique des ontologies se range parmi les solutions de communication de machine à machine, avec l'ensemble des solutions constituant le Web sémantique.

<sup>8</sup> Un des outils les plus connus est Protégé (<http://protege.stanford.edu>).

<sup>9</sup> RDF : *Resource Description Framework*. RDF permet de définir des triplets de la forme : sujet, verbe, prédicat (objet). RDFS : *RDF Schema*. RDFS permet de représenter des ontologies.

<sup>10</sup> OWL : *Web Ontology Language* (standard de l'OMG, *Object Management Group*). Appuyé sur RDF, OWL introduit la notion de classe, à la fois concept et ensemble d'instances.

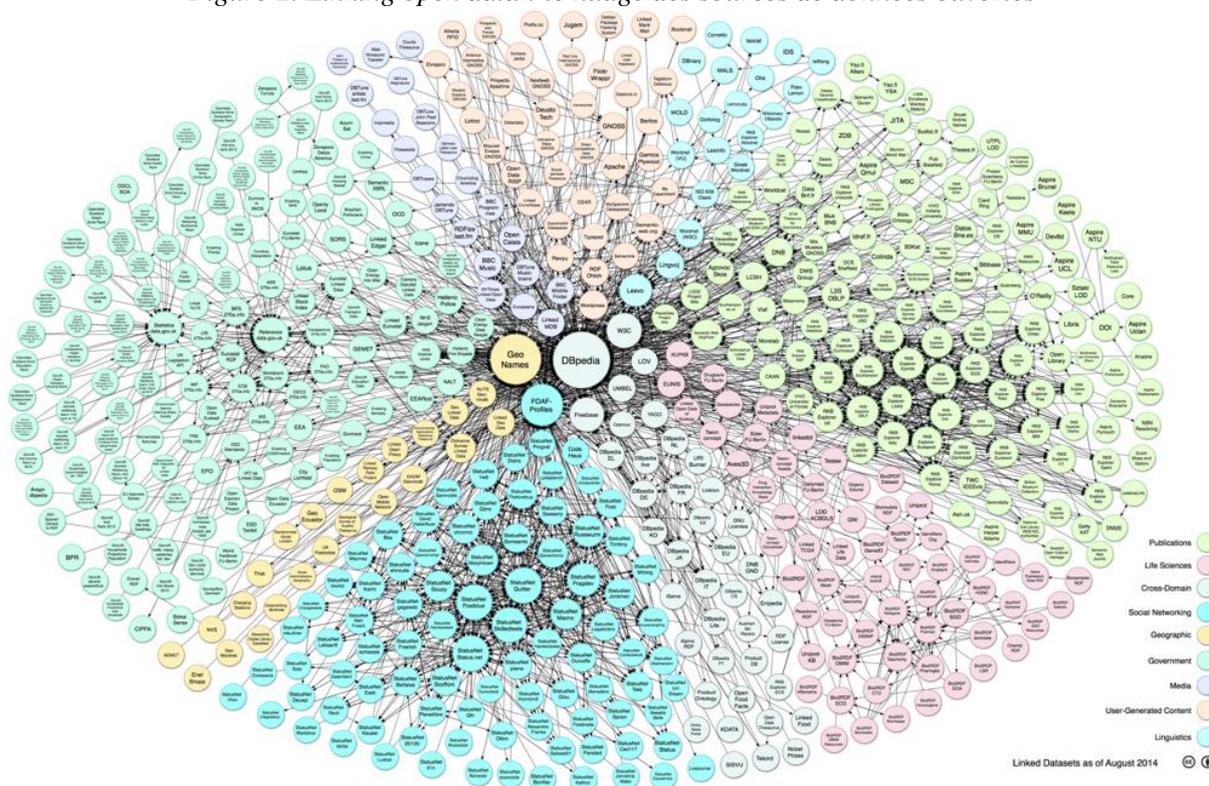
<sup>11</sup> SPARQL : *SPARQL Protocol and RDF Query Language*. Langage de requête adapté aux expressions en RDF et utilisé pour interroger les bases RDF.

D'un point de vue méthodologique, on a envie de conseiller les deux approches, puisqu'elles sont complémentaires, l'une pour présenter la connaissance à la manière d'une encyclopédie, l'autre pour lui donner une forme exploitable informatiquement. Ceci entraîne tout de même un ajustement entre les outils utilisés, si l'on veut optimiser l'investissement.

La technique des ontologies offre, de plus, les avantages suivants :

- la réutilisation des ontologies existantes ainsi que des bases de triplets disponibles publiquement<sup>12</sup>,
- l'interopérabilité, grâce à la standardisation (le *semantic Web cake* qui ordonne les standards formant le socle du Web sémantique),
- la capacité de validation formelle d'une ontologie<sup>13</sup>.

Figure 2. Linking open data : le nuage des sources de données ouvertes



Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. <http://lod-cloud.net/>

#### 4° Des connaissances : leur modélisation et l'automatisation

Enfin, il reste une étape sur notre chemin vers la **formalisation de la connaissance**. Dans cette étape ultime, non contents de nous assurer que la connaissance est parfaitement formulée, sans manque et sans contradiction, nous souhaitons qu'elle le soit *formellement* au point qu'elle puisse être comprise et reproduite par une machine. Cette étape repose sur des modèles, tout particulièrement ceux

<sup>12</sup> Le *Linking Open Data cloud diagram* (<http://lod-cloud.net/>) recense les sources disponibles de données ouvertes.

<sup>13</sup> Voir, par exemple, la fonction *reasoner* dans Protégé qui déduit les triplets en suivant la logique des prédicats « Tous les hommes sont mortels, Socrate est un homme, donc Socrate est mortel. ».

produits par la modélisation sémantique. Pour les introduire, revenons un instant sur les types de connaissances que nous cherchons à saisir dans la vie professionnelle.

La connaissance « métier », à l'œuvre dans la vie professionnelle et sans laquelle rien ne se ferait, prend les formes suivantes :

- le savoir, la connaissance conceptuelle, « engrammée » sous la forme de concepts et de relations entre les concepts ;
- le savoir-faire, la connaissance pratique, qui peut se traduire par des gestes (dans les activités manuelles) ou des modes opératoires ;
- le devoir-faire, la connaissance impérative, imposée à travers les procédures et les règles ;
- le devoir, la connaissance déontique (l'éthique, les valeurs), qui constitue une forme de connaissance profondément enfouie mais, néanmoins, nécessaire à diffuser.

La nature d'un élément de connaissance, c'est-à-dire son appartenance à l'une de ces catégories, détermine sa détection et sa formulation. La Topologie du Système Entreprise<sup>14</sup> met bon ordre dans ces éléments. Le tableau ci-après associe les catégories de connaissance à l'aspect sous lequel elles se rangent.

*Figure 3. Association des catégories de connaissances aux aspects du Système Entreprise*

Catégorie	Exemples	Aspect du Système Entreprise
<b>Savoir</b>	Le calcul d'une couverture de sinistre. Les règles d'attribution d'un compte à un client. Le comportement d'un engin. La réglementation portant sur une activité.	<b>Aspect sémantique</b> : aspect de l'entreprise constitué de la connaissance des fondamentaux de son métier.  Représenté en termes de classes (« objets métier »), de relations et de cycles de vie.
<b>Savoir-faire</b>	Savoir analyser la situation d'un client. Savoir monter un mur parfaitement vertical. Savoir réparer une machine.	<b>Aspect pragmatique</b> : aspect de l'entreprise portant sur les activités et la façon de les mener. Représenté en termes de processus, de situations de travail, d'organisation...
<b>Devoir-faire</b>	Procédure qualité. Règle d'organisation définissant un pouvoir, un rôle. Règles de sécurité...	Entre savoir-faire et devoir-faire, la nuance est la même qu'entre la description et la prescription. Du point de vue de la représentation des connaissances, les mêmes outils sont utilisés.
<b>Devoir</b>	Se montrer disponible auprès des clients. Intervenir dans des délais minimaux. Ne rien faire qui mette en danger la planète.	<b>Aspect intentionnel</b> , facette « Valeurs ». On distinguera, ici aussi, la réalité morale de l'entreprise (observée) et la prescription de comportement (affirmée).

Ce sont donc plusieurs modèles, de différents types, qui nous aideront à formuler toute la connaissance. Auparavant, nous pouvons nous demander s'il nous faut développer tous ces modèles et couvrir toutes ces catégories de connaissances. La réponse dépend du contexte et du type de transformation visé. La plupart du temps, le modèle sémantique sera indispensable puisqu'il fournit l'expression des concepts et que ceux-ci sont la matière de tout ce qui se passe dans l'entreprise.

<sup>14</sup> La Topologie du Système Entreprise est le cadre de référence proposé par la méthode publique Praxeme. Elle identifie et articule sept aspects appartenant à tout système et qu'il convient de décrire en respectant leur logique propre.

Le travail sur la terminologie et, éventuellement, sur les ontologies a préparé la modélisation sémantique. Qu'apporte cette dernière ? Les exigences de formalisme et d'économie propres à cette discipline débouchent sur plusieurs bénéfiques :

- Par rapport à une terminologie ou une ontologie<sup>15</sup>, le modèle sémantique est plus compact : il ne conserve que les concepts qui portent des propriétés caractéristiques utiles ; il ne s'oblige pas à reprendre tous les termes<sup>16</sup>.
- Le modèle sémantique débusque systématiquement les concepts génériques, souvent cachés sous des concepts spécialisés et inaperçus des acteurs « métier ». La généralité d'un modèle est un facteur d'économie et de partage. Dans certains cas, elle assure l'interopérabilité avec d'autres organisations.
- Le modèle cherche à exprimer toute la sémantique des concepts, sous trois axes : information, transformation, action. Il se donne des catégories de représentation dans le but d'exprimer cette sémantique de façon à ce qu'elle puisse être ensuite exploitée mécaniquement (opérations avec une signature documentée, automate à états...). Différents types de relations permettent de formaliser les déterminations entre concepts (héritage, associations : simples, réifiées, qualifiées, n-aires). Les cardinalités des associations y contribuent grandement.
- Ce faisant, le modèle sémantique assure une forme de vérification et de bouclage dans l'expression de la connaissance.
- Le modèle sémantique fournit le point de départ pour une révision des processus « métier » ainsi que pour une conception correcte de la solution informatique, par application de « règles de dérivation »<sup>17</sup>.

## Conclusion

Nous disposons de plusieurs approches pour l'expression des connaissances. Elles ont toutes leur intérêt et nous pouvons les adopter simultanément, soit dans une approche progressive, soit pour leur complémentarité. Notre message central, dans cet essai, est que ces techniques s'ordonnent sur une échelle qui va de l'acte le plus intuitif (les classifications de documents) à l'effort le plus exigeant (la modélisation sémantique). En s'élevant sur cette échelle, les bénéfices augmentent, d'un double point de vue :

- l'assurance d'avoir exprimé *toute* la connaissance, de façon correcte, vérifiable, voire « activable » ;
- la capacité à automatiser tout ou partie de cette connaissance.

Le schéma ci-après résume cette position. La méthodologie de transformation d'entreprise, Praxeme, cherche à tirer parti des différentes disciplines, certaines confinées dans la sphère universitaire. Sa vocation est de les amener dans les entreprises pour les mettre au service de leur transformation. À cette fin, elle en tire des procédés qui constituent autant de modes d'emploi guidant les pratiques.

---

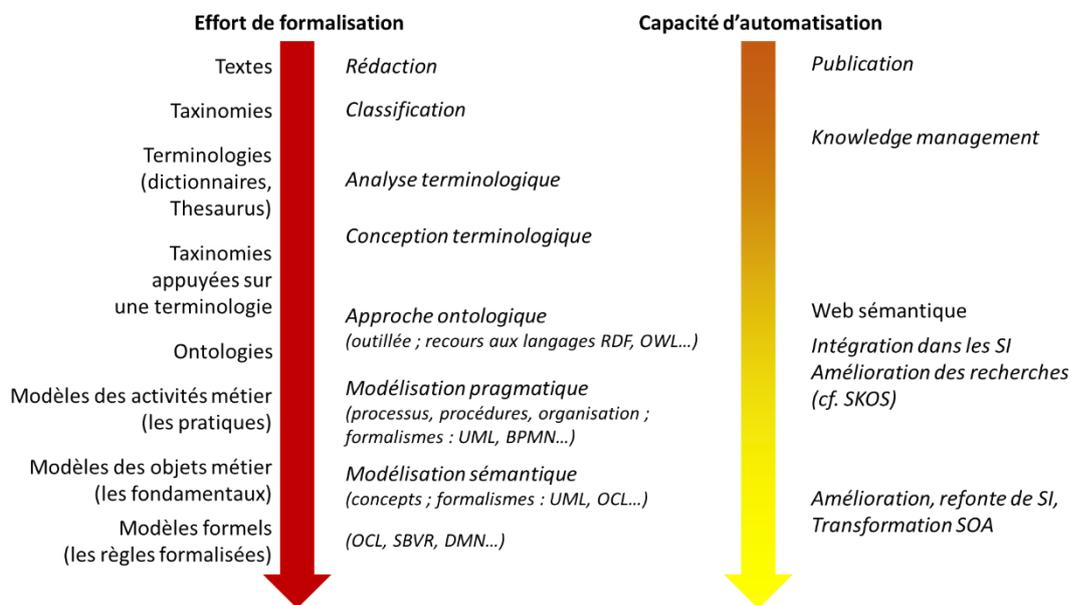
<sup>15</sup> Une ontologie peut très bien mêler des classes et des instances, sur un même plan – ce qui ferait frémir un modélisateur sémantique. Ce dernier utilise les classes et les instances, mais en les séparant rigoureusement (impossible, par exemple, de les représenter sur un même diagramme).

<sup>16</sup> Le modélisateur applique le critère du rasoir d'Ockham, éliminant systématiquement les éléments dont on peut se passer dans le modèle, sans réduire son pouvoir d'expression.

<sup>17</sup> Cette exploitation mécanique de la connaissance est rendue possible par le périmètre de la méthode qui embrasse tous les aspects de l'entreprise.

Puisque son ambition est de couvrir tous les aspects de l'entreprise, elle peut indiquer comment ces procédés se positionnent et s'enchaînent dans un effort maîtrisé de transformation.

Figure 4. Gradation de l'effort dans la formalisation de la connaissance et niveau des retombées attendues



D'autres solutions ou approches concernent la formulation de la connaissance. On peut citer, notamment :

- le traitement automatique de la langue (TAL) qui permet d'exploiter des corpus de textes ou des messages pour en extraire des informations (jusqu'aux sentiments exprimés dans des courriers)<sup>18</sup>,
- les *big data* grâce auxquelles une nouvelle connaissance émerge à travers des techniques d'analyse avancée (y compris l'apprentissage artificiel).

Ainsi, on constate que les entreprises qui en perçoivent les enjeux disposent d'une large gamme de solutions pour exprimer leurs connaissances. Elles ont à décider de leur stratégie en combinant ces solutions et en pondérant l'effort de formalisation, rapporté aux retombées en termes de maîtrise et d'automatisation.

<sup>18</sup> Voir l'offre de la société Proxeme et, plus généralement, les solutions de traitement automatique de la langue (TAL ; en anglais : NPL pour *natural language processing*).