

Le projet européen DHYDRO : la normalisation à l'épreuve d'un forum terminologique

Blampain (D.), Descotte (S.), Husson (J.-L.), Rohde (H.), Romary (L.),
Van Campenhoudt (M.) et Viscogliosi (N.)¹

Résumé

Le projet de forum DHYDRO a pour objet de créer sur Internet un espace terminologique multilingue spécialisé dans le domaine de l'hydrographie, une science appliquée indispensable à la présence humaine sur ou sous les océans. Ce site sera conçu autour du *Dictionnaire hydrographique international (D.H.I.)*, qui décrit quelque 7 000 concepts, et de différentes publications de l'Organisation hydrographique internationale. L'objectif est d'illustrer comment un espace interactif sur Internet permet d'améliorer, de mettre à jour et de consolider des données terminologiques décrites dans un grand nombre de langues. Des solutions technologiques robustes, accessibles et reproductibles devront permettre de suivre rapidement l'évolution des terminologies, de manière à pouvoir définir adéquatement les nouveaux concepts et en normaliser la désignation.

Mots-clés : MLIS – hydrographie – forum – X.M.L. – normalisation.

Le domaine : l'hydrographie, mission internationale de service public

L'hydrographie débouche sur la création et la diffusion d'informations engageant directement la sécurité de la navigation, sous la forme de cartes marines, d'ouvrages nautiques et d'avertissements aux navigateurs. Assurer la sécurité de la navigation maritime est donc une responsabilité lourde, qui souvent ne peut être garantie qu'au niveau de l'État. Cette mission est remplie au profit de tous les navigateurs, professionnels ou plaisanciers, nationaux ou étrangers, conformément aux obligations internationales.

Les domaines d'application des données hydrographiques sont nombreux en dépit de l'apparente spécificité de la matière. Ces domaines recouvrent entre autres les transports maritimes, la pêche, les ressources énergétiques, les câbles sous-marins, le balisage côtier et le dragage. Les informations hydrographiques servent également à ceux qui se préoccupent de l'environnement marin, comme par exemple les ingénieurs en sciences océaniques, les océanographes, les spécialistes en biologie marine et les spécialistes de l'environnement. L'ensemble de ces activités conduit à la publication et à la traduction de nombreux ouvrages qui sont autant de références pour la communauté maritime.

1. Les auteurs remercient Charles Delapalme (SHOM) pour sa collaboration à l'écriture de cette communication.

Le contexte : l'enjeu du *Dictionnaire hydrographique international*

Le Bureau hydrographique international publie périodiquement un *Dictionnaire hydrographique international (D.H.I.)*, édité sous la forme d'un volume indépendant pour chacune des trois langues du dictionnaire (anglais, français et espagnol). Des versions allemande, chinoise et japonaises sont en cours de rédaction. Chacun des volumes est révisé tous les dix ans environ, donnant lieu à l'impression d'une nouvelle édition. Le *D.H.I.*, édité jusqu'à présent sous forme « papier », recouvre aujourd'hui entre 6 800 et 7 000 entrées selon les versions. Loin d'être restreint à une spécialité scientifique, le corpus couvre une très large part du vocabulaire maritime, particulièrement un très grand nombre de concepts abstraits peu traités dans les autres dictionnaires de marine.

On remarquera que le *D.H.I.* est d'une complexité similaire à différents projets de dictionnaires informatisés spécialisés que l'on peut trouver actuellement sur le web. Cependant, peu de projets reposent sur des standards internationaux pour représenter leur contenu et sur la définition d'une architecture générique et ouverte permettant de dépasser la simple consultation des entrées terminologiques correspondantes.

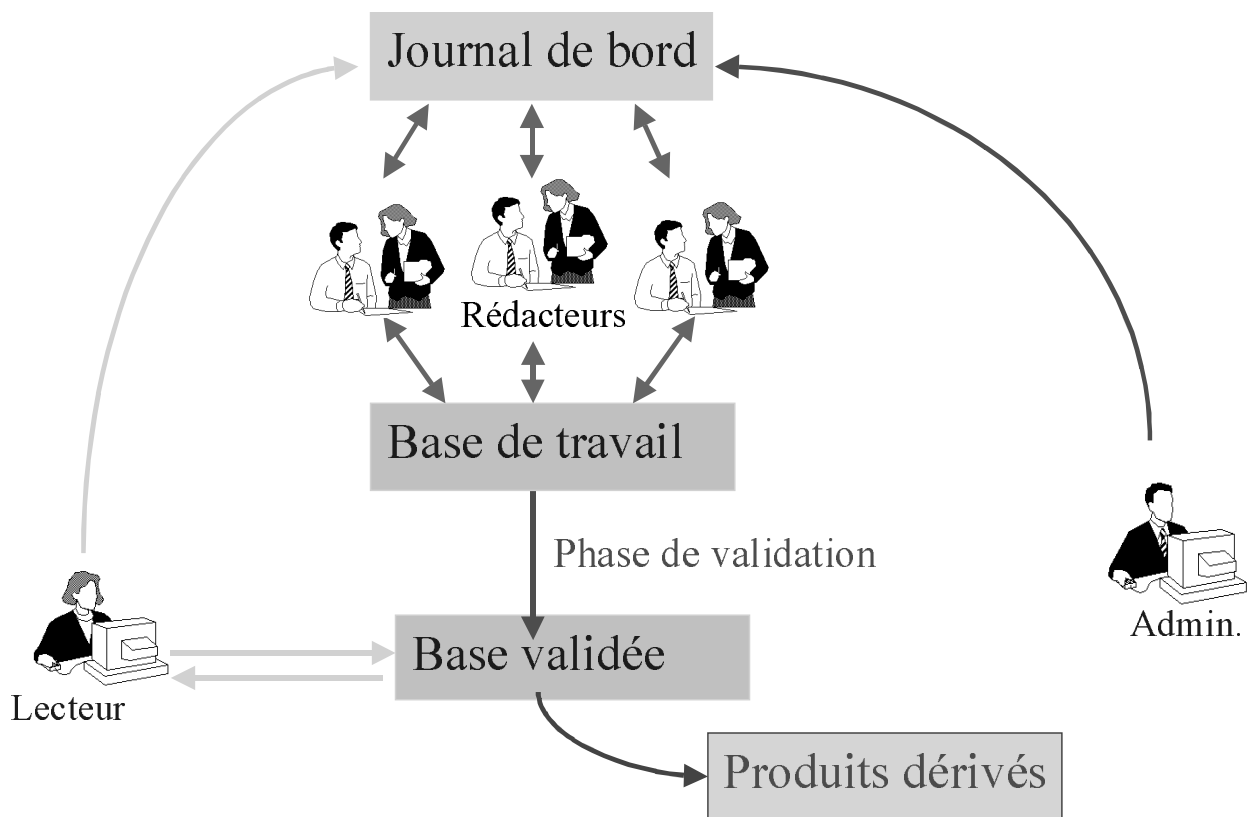
Par ailleurs, le B.H.I. diffuse, dans ses différentes langues de travail, diverses publications qui méritent d'être rassemblées dans une même base de données textuelles :

- la *Revue hydrographique internationale*, publiée deux fois par an, contenant des articles spécialisés se rapportant à l'hydrographie ou à des sujets connexes,
- le *Bulletin hydrographique international*, publié tous les mois, contenant des informations courantes intéressant l'hydrographie,
- l'annuaire de l'O.H.I., présentant l'ensemble des pays membres et la structure de l'O.H.I.,
- une série de publications spéciales dont chacune constitue un travail définitif, comme par exemple le *D.H.I.* dans ses trois versions.

Les objectifs poursuivis et les services envisagés

Le projet DHYDRO a pour objet de créer sur Internet un espace terminologique multilingue spécialisé dans le domaine de l'hydrographie générale. Le consortium ne cherche pas à concevoir des produits d'ingénierie linguistique expérimentaux, mais bien à rassembler sur un site web interactif des services utiles aux terminographes, aux traducteurs et aux spécialistes. Cet espace doit comprendre les fonctionnalités suivantes :

- consultation sur site d'une sélection des publications du B.H.I. au format X.M.L., incluant notamment le *D.H.I.* dans ses trois versions française, anglaise et espagnole,
- introduction en ligne des nouvelles publications du B.H.I. à partir de documents en format *Word*, à des fins de recherche d'occurrences,
- modification en ligne de chacune des versions monolingues du dictionnaire par les comités de rédacteurs spécialisés et concertation sur le découpage conceptuel commun,
- mise en place d'un espace interactif de discussion entre utilisateurs et créateurs de terminologie hydrographique dans le cadre de la fusion en une base de données terminologiques multilingue ayant une valeur de référence pour le secteur.



Les formats d'échange

Dans la mesure où le consortium a choisi de travailler dans l'univers de X.M.L., il semble logique qu'il opte pour un balisage des données qui soit conforme à une norme internationale d'échanges.

On attend d'une norme d'échange qu'elle réponde aux exigences énumérées ci-dessous :

- possibilité de rendre compte du mode de fonctionnement du comité de rédaction ;
- souplesse : le format ne doit pas interdire l'approche conceptuelle souhaitée par les partenaires ni imposer un modèle de fiche ;

- structuration : le format doit rendre fidèlement compte de la structure de la fiche et des relations entre les champs ;
- évolutivité : facilité d'adjonction de nouvelles langues, ouverture vers les progrès dans la représentation des connaissances (notamment les bases de connaissances terminologiques) ;
- capacité à prendre en compte des alphabets spéciaux (phonétique, japonais, chinois...).

Dans la mesure où l'annexe technique du projet DHYDRO prévoit un balisage des données dans la perspective de la T.E.I. (*Text Encoding Initiative*), il semble logique que le projet opte pour le format MARTIF², quitte à devoir proposer à ses concepteurs les aménagements qui seraient jugés indispensables.

Les moyens techniques mis en œuvre

Pour répondre aux différents objectifs décrits ci-dessus et garantir en même temps la généricité et l'évolutivité nécessaire à un projet tel que le notre, les choix technologiques suivant ont été effectués :

- la consultation en ligne du *D.H.I.*, qui doit pouvoir se faire à partir de n'importe quel navigateur internet est réalisé à l'aide d'une *applet* - c'est à dire un élément logiciel téléchargeable en ligne - qui permet d'effectuer des requêtes multiples en direction de la base ;
- l'édition des fiches du *D.H.I.* par le comité de rédaction repose quant à elle sur une application spécifique permettant à chaque rédacteur de travailler sans être nécessairement connecté à la base centrale, tout en effectuant régulièrement des opérations de synchronisation avec celle-ci ;
- le journal de bord repose sur un système classique de gestion de messagerie et contient, d'une part, les interactions spontanées des rédacteurs et, d'autre part, des messages générées automatiquement lors des phases de synchronisation entre les rédacteurs et la base de travail, de façon à signaler les modifications effectués par ceux-ci (par exemple la création d'un nouveau concept). Nous envisageons, au passage, de donner la possibilité à toute personne consultant la base de faire des propositions éditoriales par le biais de ce journal de bord.

L'ensemble des développements effectués repose sur l'utilisation du langage de programmation *Java*, qui présente, d'une part, l'avantage d'être parfaitement adapté aux mécanismes de communications propres à Internet et, d'autre part, est compatible avec la norme

2. Dans le cadre du programme MLIS, trois projets européens qui impliquaient la fusion de données terminologiques (INTERLEX, PUBLINET et INESTERM) ont adopté le format-pivot GENETER créé par l'Université de Rennes II. Ce standard *de facto* a l'avantage de permettre des échanges aveugles, pour lesquels l'ISO TC37 n'a pas encore déterminé s'il le retiendrait comme norme ou choisirait la toute récente proposition *Blind-MARTIF* de l'organisme T.T.T.

UNICODE pour la représentation de données multilingues telles qu'envisagée pour le *D.H.I.*. Par ailleurs, l'utilisation du format X.M.L. permet de rendre parfaitement flexible et modulaire les opérations de visualisation des fiches terminologiques par le biais de feuilles de styles décrites en format X.S.L. (*eXtended Style Language*). Ce langage permet de décrire des vues multiples (lexique bi- ou multilingue, fiches monolingues, etc.) associées à une même base terminologique et de redéfinir ces vues sans avoir à modifier les programmes informatiques sous-jacents.

Les progrès attendus

On peut raisonnablement penser que le projet DHYDRO permettra de diffuser une information plus souple d'utilisation, ouverte à davantage d'utilisateurs et affranchie des contraintes qui limitent traditionnellement le travail de description terminographique et celui de traduction. Le forum hydrographique sera, à ce titre, un outil précieux pour les utilisateurs de terminologie hydrographique, qu'ils soient chercheurs, ingénieurs, professionnels de la navigation, traducteurs ou terminologues. Il permettra également de mettre en contact ces utilisateurs avec les rédacteurs du *D.H.I.*, selon une structure habituelle de forum de discussion.

Le projet DHYDRO est susceptible d'entraîner des progrès conséquents dans l'utilisation et dans la gestion de la terminologie hydrographique selon les axes suivants :

- l'accès à l'information hydrographique est potentiellement beaucoup plus large dans le cas d'un accès en ligne que dans celui d'une diffusion sur support papier ;
- la convivialité et la maniabilité d'un dictionnaire au format X.M.L. accessible depuis Internet sont supérieures à celles d'un dictionnaire sur support papier ;
- la mise en place d'une structure de modification des documents en ligne permettra aux rédacteurs autorisés d'effectuer une mise à jour continue du *D.H.I.*, au lieu d'une édition « papier » tous les dix ans³ ;
- les possibilités d'interactions entre spécialistes, traducteurs et terminographes permettront une consultation mutuelle aisée, de nature à garantir le processus de certification des terminologies au sein de l'O.H.I. ;
- la possibilité de consulter des corpus textuels spécialisés et d'y rechercher des termes nouveaux, des termes existants ou des termes équivalents devrait permettre d'améliorer considérablement le travail de rédaction du dictionnaire. Son usage à des fins de traduction spécialisée devrait, dans le même temps, être profondément transformé.

3. Ceci n'empêchant bien sûr pas l'édition sous forme papier ou sous cédérom d'une version « figée » du *D.H.I.* à

Enfin, les formats et les outils définis dans le cadre du projet permettront d'une part d'étendre la méthodologie et l'infrastructure à d'autres projets de rédaction de dictionnaires mono ou multilingues et, d'autre part, d'envisager l'introduction d'outils spécifiques (p.ex. extracteurs semi-automatiques de termes à partir de textes de référence) au sein de l'architecture logicielle développée.

L'impact attendu

Prise en compte des objectifs du programme MLIS

Dans le cadre du programme MLIS, la Commission européenne souhaite mobiliser les acteurs du domaine de la terminologie et promouvoir la coordination et la coopération dans les divers secteurs concernés. Conformément à l'objet de l'appel à propositions, les promoteurs du projet DHYDRO entendent proposer des « techniques flexibles et décentralisées permettant un suivi en temps réel de l'évolution des termes, la consultation et l'intégration éventuelle des termes existants, la définition de nouveaux termes et la normalisation des propositions terminologiques ».

Plus particulièrement, le projet DHYDRO est en accord avec les perspectives du programme MLIS à différents niveaux :

- Il vise à faire progresser les activités de traduction dans le domaine de l'hydrographie. Les traducteurs bénéficient, en effet, de l'accès à des outils en ligne, d'usage rapide (trouver la traduction d'un terme), complets (trouver des documents de référence utilisant un terme), fiables (corrections régulières) et puissants (filtrage d'un texte par un moteur de recherche).
- L'information terminologique sur l'hydrographie gagnera en précision et sera mise régulièrement à jour. La modification continue du *D.H.I.* par les comités de rédaction, autant que la confrontation itérative de ces comités avec les utilisateurs de la terminologie grâce à l'espace terminologique interactif, garantissent une information de qualité.
- Les techniques envisagées devraient permettre de suivre rapidement l'évolution des terminologies, de manière à pouvoir définir adéquatement les nouveaux concepts et en normaliser la désignation.
- Le consortium est conçu autour d'un organisme international (l'O.H.I.) ayant un important rôle de coordination dans son domaine et ayant une activité de production terminologique de référence pour de nombreux secteurs. Les autres partenaires sont le LORIA (C.N.R.S.-INRIA, Nancy), le centre de recherche TERMISTI (ISTI, Bruxelles), L'Institut für deutsche Sprache

(I.D.S., Mannheim) et le Service hydrographique et océanographique de la Marine (SHOM, Paris). Ils apportent chacun des compétences particulières, liées à leur connaissance du domaine, à leur capacité à fédérer un réseau d'experts, à leur expérience de la gestion terminologique multilingue, à leur savoir-faire en matière de gestion informatique des données terminologiques et textuelles.

Mesure de l'impact et des performances

L'impact des fonctionnalités offertes sur le forum DHYDRO fera l'objet d'une étude attentive, de manière à les améliorer en cours de projet et à proposer au terme de celui-ci une méthodologie reproductible pour de nombreux autres domaines de spécialité. En effet, le projet n'est pas techniquement spécifique au cadre de l'hydrographie : sa méthodologie et ses fonctionnalités opérationnelles devront pouvoir être aisément transposables en vue du développement de forums similaires.

Par ailleurs, même si le projet DHYDRO est dans un premier temps limité à trois langues européennes (anglais, français, espagnol), il sera pensé de manière à pouvoir être étendu à d'autres langues propres aux États membres de l'O.H.I. On pense par exemple à la version allemande (achevée à 80 %) ou aux versions chinoise et japonaise.

En définitive, les progrès permis par le projet DHYDRO devraient être notamment perceptibles aux niveaux de :

- la qualité de l'information diffusée : le dictionnaire sera régulièrement mis à jour, consolidé et complété et ces améliorations seront dûment observées et quantifiées ;
- le volume accru des utilisateurs de cette information : un historique des connections au forum et du profil des utilisateurs sera en ce sens un bon indicateur d'efficacité du projet ;
- l'amélioration évidente des processus de traduction des documents officiels, des rapports d'activité et autres documents multilingues au sein des organismes membres de l'O.H.I. ;
- l'amélioration des services d'interprétation simultanée, par exemple lors des Commissions hydrographiques régionales et lors de l'organisation de séminaires internationaux.

Bibliographie

CENTRE DE RECHERCHE TERMISTI & LORIA (1999) : *Le modèle éditorial du forum*, rapport D1.1 du projet MLIS-2009 DHYDRO, Bruxelles-Nancy, mars 1999.

CENTRE DE RECHERCHE TERMISTI & LORIA (1999) : *Le modèle de données et la D.T.D.*, rapport D1.2 du projet MLIS-2009 DHYDRO, Bruxelles-Nancy, mars 1999.

INTERNATIONAL HYDROGRAPHIC ORGANIZATION (1994) : *Hydrographic Dictionary*, 5th ed., Monaco, Bureau hydrographique international.

ISO/FDIS 12200 (1998) : *Applications informatiques en terminologie - Format de transfert de données terminologiques exploitables par la machine (MARTIF) - Transfert négocié*, Genève, ISO.

ISO/FDIS 12620 (1998) : *Aides informatiques en terminologie - Catégories de données*, Genève, ISO.

LORIA & CENTRE DE RECHERCHE TERMISTI (1999) : *Choix techniques d'accès aux données, échanges et outils*, rapport D1.2 du projet MLIS-2009 DHYDRO, Nancy-Bruxelles, mars 1999.

ORGANISATION HYDROGRAPHIQUE INTERNATIONALE (1998) : *Dictionnaire hydrographique*, 5^e édit., Monaco, Bureau hydrographique international.

ORGANIZACIÓN HIDROGRÁFICA INTERNACIONAL (1996) : *Diccionario Hidrográfico*, 5^e édit., Monaco, Bureau hydrographique international.