



Multilingual management  
of conceptual and lexical resources.  
A comparison of Bioportal and HeTOP  
(**H**ealth **T**erminology/**O**ntology **P**ortal).

---

**Darmoni Stefan, Julien Grosjean**

Laboratoire d' informatique, de traitement de l' information et des systèmes.  
Département de Médecine, Centre Hospitalier Universitaire de Rouen, France

**Annex 3**

**to the SHN Work Package 3**

**Deliverable D3.3**

*Final version, March 31, 2015*

## Document description

Deliverable:	Annex 3 to SHN WP3 Deliverable D3.3
Publishable summary:	In this Annex, the internet portal HETOP for multilingual terminologies and ontologies is described, and compared to the portal BIOPORTAL, with regard to content, functionality, user interface, and technical aspects. Potential applications of the HETOP site for semantic interoperability and the provision of lexical and reference terminologies are highlighted.
Status:	Final draft
Version:	1.5
Public:	<input checked="" type="checkbox"/> Yes
Deadline:	March 31, 2015
Contact:	Elena Cardillo <a href="mailto:Stefan.Darmoni@chu-rouen.fr">Stefan.Darmoni@chu-rouen.fr</a> Robert Vander Stichele <a href="mailto:robert.vanderstichele@ugent.be">robert.vanderstichele@ugent.be</a>
Editors:	Stefan Darmoni, Julien Grosjean

## Table of Content

Introduction .....	3
Aim of this study .....	3
Description of the 2 portals under comparison .....	3
BioPortal - <a href="http://bioportal.bioontology.org">http://bioportal.bioontology.org</a> .....	3
HeTOP - <a href="http://www.hetop.org">http://www.hetop.org</a> .....	3
Methods and criteria of comparison .....	5
Results of the comparison .....	5
Content Comparison .....	5
Functions and Tools Comparison .....	5
User Interface and Usability Comparison .....	6
Technical Comparison .....	6
Objectives for the future of HeTOP .....	7
With regard to multilingual management of Reference terminologies .....	7
With regard to multilingual management of lexical resources .....	7
References .....	9

Note : This annex was commissioned by Prof. Dr. R. Vander Stichele, Workpackage Leader of SemanticHealth Net WP3, to Stefan Darmoni, Laboratoire d' informatique, de traitement de l' information et des systèmes. Département de Médecine, Centre Hospitalier Universitaire de Rouen, France, 1, rue de Germont, 76031 Rouen Cedex Tel. : +33 2 32 88 88 32

E-mail : [Stefan.Darmoni@chu-rouen.fr](mailto:Stefan.Darmoni@chu-rouen.fr)

## Introduction

Because of the constant growth of biomedical data it became mandatory to index or annotate it with controlled concept-oriented linguistic resources in order to store it and to retrieve it intelligently. A key aspect in addressing semantic interoperability for biomedical data is the use of terminologies or ontologies (T/O) as a common data structure [1] [2]. Many T/O have been created in the last decade for different purposes: indexing or annotating documents, organizing knowledge, inferring facts, teaching. Their differences justify using them simultaneously.

Several tools have been created to store, search and use:

- the Unified Medical Language System (UMLS) server [3],
- the European Bioinformatics Institute (EBI) Ontology Lookup Service [4],
- the National Center for Biomedical Ontology (NCBO) BioPortal [5] or
- the Health Terminology/Ontology Portal (HeTOP) [6].

## Aim of this study

Currently, BioPortal is the most used terminology server on the Web. Therefore, we want to compare BioPortal to HeTOP, a crosslingual multiterminology portal.

## Description of the 2 portals under comparison

### BioPortal - <http://bioportal.bioontology.org>

Developed by the US NCBO, BioPortal is a repository of biomedical ontologies which hosts more than 350 ontologies in different formats [7], [8]. These ontologies are regularly updated by users and accessible via a web site for the humans and web-based services for programs. BioPortal is a library of community ontologies [9] designed as a "one-stop shop" repository. Users have access to the ontologies with or without restriction (depending on the level of restriction set by the editor) and can perform editing, comment, rating and content adding operations.

### HeTOP - <http://www.hetop.org>

HeTOP [6] has been designed as a reference multi-terminology and cross-lingual portal [10] to help health professionals, students, patients, librarians and translators to retrieve resources and knowledge across a high variety of complex health fields. HeTOP is a T/O portal developed by the Department of Biomedical Informatics, Rouen University Hospital, Normandy, France. It hosts more than 56 T/O in several languages, mainly European ones, such as English or French, but also German, Italian and Dutch but also with no Latin alphabet (Greek, Russian) and more recently outside Europe (Japanese, Mandarin, Arabic & Hebrew) (**23 different languages**). Thanks to an academic cooperation

with several colleagues around the world, the HeTOP interface has also been (partially or fully) translated in English, French, German, Spanish, Italian, Portuguese, Turkish, Hebrew, Arabic.

The HeTOP data back-end has been designed as a generic knowledge resources repository and is compliant with the Tao's guidelines [12]. The system (back-end, HeTOP web site and HeTOP web service) has been mainly conceived and developed during the Julien Grosjean PhD thesis [13], but initiated during the InterSTIS project (2007-10, funded by the French National Research Agency, programme Technologies for Health).

Most of the T/O are international or French national references such as MeSH, ICD-10, or CCAM. These T/O are regularly updated and are accessible via a web site and a web service.

HeTOP provides the terminological data for each concept: preferred terms, original code, synonyms, definitions and other attributes, relations and hierarchies.

To develop HeTOP was a time consuming task (14 man-years. Maintenance requires 2 man-years per year (integration & maintenance of T/O + mappings).

Several services on demand were developed and integrated to HeTOP:

- access to other resources on the Internet (PubMed, CISMef, etc.) through a French InfoButton (InfoRoute; URL: [inforoute.chu-rouen.fr](http://inforoute.chu-rouen.fr))
- access to mappings tools
- access to automatic indexing tool (ECMT; URL: [ecmt.chu-rouen.fr](http://ecmt.chu-rouen.fr))

Twenty out of 56 T/O are freely available. Most of the rest need an ID/password.

More than 2,100 users are currently registered. 650 are connected to HeTOP each working day.

Overall, the number of distinct CUI with at least one French translation in HeTOP is **317,539** vs. **45,405** in UMLS (MeSH, MedDRA, WHOART, ICPC) vs. **142,933** UMLS+ (including the French version of SNOMED Int. & ICD10).

Table 1 shows the details of the terminologies and ontologies included in UMLS : number of terms included into HeTOP and the number of CUIs translated in French

The HeTOP curators have translated their set of Terminologies and Ontologies mainly in French. A similar effort has to be done for the other European languages.

Some information is already available in 21 other languages (e.g. MeSH is already available in 16 languages).

Table 2 shows the same information for the Terminologies and ontologies not included in UMLS.

## Methods and criteria of comparison

As far as we know, there is no existing method to compare such applications. The work of [4] has inspired us for establishing a criteria list for comparing BioPortal and HeTOP properties (technologies, methodologies, policies, target users, final purposes, features, etc.) but also from other similar portals [11]. We categorized all the criteria in four groups: (i) Content, (ii) Functions & Tools, (iii) User Interface & Usability, (iv) Methods & Technologies.

## Results of the comparison

### Content Comparison

The comparison of content reveals three important differences between BioPortal and HeTOP: (i) The volume of data is more important in BioPortal considering the T/O numbers or concept numbers (respectively 5,960,457 and 1,951,834). However, since T/O have various number of concepts and terms (terms are preferred terms plus synonyms, in any languages) and since HeTOP is dealing with multilingual content, T/O numbers are not suited comparison indicators: the more relevant figures are the number of terms and the total number of relations. Indeed, BioPortal and HeTOP have around the same number of terms (about 6,600,000). Unfortunately, it is not possible to easily calculate the total number of relations in BioPortal. (ii) T/O formats and update frequencies are quite different in BioPortal compared to HeTOP. While BioPortal is focused on ontologies and takes advantage of standard formats (e.g. OBO, OWL) and programs, HeTOP is also hosting heterogeneous representation formats such as Microsoft Excel files, XML files or database dumps. A special work has to be done for every new T/O source: this cannot be executed automatically and it implies expertise and development. (iii) One major difference between BioPortal and HeTOP is the expertise brought to every T/O. While BioPortal is automatically importing ontologies and does not change anything, except new automatic mappings and some manual users mappings, each HeTOP hosted T/O undergoes a series of process to leverage its content and meta-data (new translations, synonyms, definitions, relations, mappings, etc.).

### Functions and Tools Comparison

Despite the similarity of basic tools for both portals, some details and tools differ slightly but have a direct consequence for the human users. (i) The BioPortal search engine only searches for exact terms in English (among preferred terms and synonyms), while HeTOP search engine is able to add wildcards to search for terms containing the query, in two languages simultaneously. This has a direct impact on how users can search terms; for instance, if one searches "myopathy" (in English) in the NCIT, BioPortal retrieves 5 terms whereas HeTOP retrieves 25; because wildcards are handled, HeTOP search engine is actually querying "\*myopathy\*" and retrieves terms such as "Cardiomyopathy".

## User Interface and Usability Comparison

To compare the user experience using both portals, we compared two functions involving time responses. First, a comparison has been made between search engines performances in terms of response time and result numbers; we picked and performed 10 random queries on both search engines and we measured the user experience time

(with the FireBug Mozilla Firefox plug-in; URL: <http://getfirebug.com/>) and noted the results. No options have been selected in both portals and the searches have been made among all T/O, in English and in French in HeTOP and only in English in BioPortal (no possible multilingual search).

The average response time was 5.57 seconds vs. 3.17 seconds respectively for BioPortal and HeTOP. The average number of results was 19.7 vs. 359.4 respectively for BioPortal and HeTOP. About multilingualism, within BioPortal, T/O in other languages than English are mostly available as "views" of the corresponding English T/O (e.g., the French MeSH is a view of the MeSH) but it is impossible to get the French term while browsing the English one and vice-versa. Within HeTOP, T/O are not language specific (e.g., the MeSH exists only once with the available translated terms). Therefore it is easy from the English term to get to the French one and vice-versa. The whole user interface is internationalized and the search can be performed per language. Switching from one language to another is context sensitive. Therefore, HeTOP seems more adapted for Europe and beyond, where multilingualism is more than important. The details of the comparison among these two terminology portals are available at: [http://www.chu-rouen.fr/cismef/papers/comp\\_biportal\\_hetop\\_table.pdf](http://www.chu-rouen.fr/cismef/papers/comp_biportal_hetop_table.pdf).

## Technical Comparison

While BioPortal is based on a RDF triple store data model and RDF database backend, HeTOP has a metamodel for T/O which encapsulates specific T/O models into an Oracle 11g r2 relational database. Both portals are coupled to a web-based service:

### **Bioportal:**

[http://www.bioontology.org/wiki/index.php/Resource\\_Index\\_REST\\_Web\\_Service\\_User\\_Guide](http://www.bioontology.org/wiki/index.php/Resource_Index_REST_Web_Service_User_Guide)

### **Hetop :**

<http://cispro.chu-rouen.fr/CISMeFhetopservice/>

## Objectives for the future of HeTOP

The objective of HeTOP is to become in the near future the terminology/ontology server in Europe, mainly for European academicians, terminologists, translators, but also health professionals and health students eager to obtain information about health dictionaries, lexicons, classifications, terminologies, thesauri, and ontologies.

### With regard to multilingual management of Reference terminologies

HeTOP could integrate the main health reference terminologies and ontologies in the various European languages (e.g. ICD10 & 11 for epidemiology, ATC for drugs, ICPC2 for primary care, Human Phenoty Ontology, OMIM, and Orphanet for rare diseases).

A huge work will be necessary to translate these main terminologies in the various European languages. Some further work is necessary to map these terminologies (some which are not included in UMLS).

According to the SKOS model, these mappings will need to explicit the relations between two concepts:

Exact match, which means these two concepts are true synonyms

Broader term-narrower term (BTNT); two concepts are in the same hierarchy, the first one is broader than the second

Narrower Term-Broader Term (BTNT): this relation means that the first concept is narrower in a hierarchy than the second concept

Close relation: this relation means that the first concept is related by a see also relation with the second concept, and does not belong to the same hierarchy: e.g. prison has a “close relation” with prisoners

In the next five years, some specific work could be handled to enhance drug semantic knowledge.

### With regard to multilingual management of lexical resources

Lexical resources for medical language, such as lists of words with inflectional and derivational information, are publicly available for the English language with the UMLS Specialist Lexicon since more than 20 years [14]. Since 2002, the goal of the Unified Medical Lexicon for French (UMLF)

project has been to pool and unify existing resources and to add extensively to them by exploiting medical terminologies and corpora [15]. This project is still active thanks to a collaboration between Pierre Zweigenbaum's team (LIMSI, CNRS UPR3251, University of Paris Sud), Fiammetta Namer's team (ATILF - Analyse et Traitement Informatique de la Langue Française, CNRS UMR 7118, Nancy) and the HeTOP team in Rouen, these three labs were UMLF members.

A cooperation has started between the HeTOP team (URL: <http://www.chu-rouen.fr/sibm/>) and the MERITERM consortium (URL: <http://www.meriterm.org/>) [16] to include these lexicons based on Lexical Markup Framework (LMF; URL: <http://www.lexicalmarkupframework.org/>) as a new layer into HeTOP. The HeTOP metamodel is already compliant with the Terminological Markup Framework (TMF; URL: <http://www.loria.fr/projets/TMF/>). The main perspective is to lexically and semantically link T/O and lexical resources.

## References

- [1] Bodenreider O & Stevens R. Bio-ontologies: Current Trends and Future Directions. *Briefings in Bioinformatics*. 2006;7:256-274
- [2] Rubin DL, Shah NH & Noy NF. Biomedical ontologies: a functional perspective. *Briefings in Bioinformatics*. 2008;9(2):75-90.
- [3] Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004.
- [4] Côté R, Reisinger F, Martens L, Barsnes H, Vizcaino JA, Hermjakob H. The Ontology Lookup Service: bigger and better. *Nucleic Acids Res*. 2010.
- [5] Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, Sim I, Hute CG, Solbrig H, Storey MA, Smith B, Day-Richter J, Noy NF & Musen MA. National Center for Biomedical Ontology: Advancing Biomedicine through Structured Organization of Scientific Knowledge. *OMICS A Journal of Integrative Biology*. 2006;10(2):185-198.
- [6] Grosjean J, Merabti T, Dahamna B, Kergourlay I, Thirion B, Soualmia LF & Darmoni SJ. Health Multi-Terminology Portal: a semantics added-value for patient safety. *Patient Safety Informatics - Adverse Drug Events, Human Factors and IT Tools for Patient Medication Safety*. 2011;166 of *Studies in Health Technology and Informatics*:129-138.
- [7] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith NB, Jonquet C, Rubin DL, Storey MA, Chute CG & Musen MA. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*. 2009;37:170-173.
- [8] Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T & Musen MA. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Research*. 2011;39:541545.
- [9] D'Aquin M & Noy NF. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. *Journal of Web Semantics*. 2012;11:96-111.
- [10] Grosjean J, Merabti T, Griffon N, Dahamna B, Soualmia LF & Darmoni SJ. Multi-terminology crosslingual model to create the Health Terminology/Ontology Portal. *American Medical Informatics Association Annual Symposium, Chicago, USA*. 2012:1753.
- [11] Grosjean, J; Soualmia, LF; Bouarech, K; Jonquet, C & Darmoni, SJ. Comparing BioPortal and HeTOP: towards a unique biomedical ontology portal ?. *IWBIO*, 2014.
- [12] Tao C, Pathak P, Solbrig H. R., Wei WQ, Chute C. G. Terminology representation guidelines for biomedical ontologies in the semantic web notations, *Journal of Biomedical Informatics*, Volume 46, Issue 1, February 2013, Pages 128-138
- [13] Grosjean J. Modélisation, réalisation et évaluation d'un portail Multi-terminologique, Multi-discipline, Multi-lingue (3M) dans le cadre de la Plateforme d'Indexation Régionale (PlaIR), PhD Thesis, 2014.
- [14] McCray AT, Srinivasan S, and Browne AC. Lexical methods for managing variation in biomedical terminologies. In: *Proc Eighteenth Annu Symp Comput Appl Med Care*, Washington. Mc Graw Hill, 1994:235-9.
- [15] Zweigenbaum, P; Baud, R; Burgun, A; Namer, F; Jarrouse, E; Grabar, N; Ruch, P; Le Duff, F; Forget, JF; Douyère, M & Darmoni, SJ. UMLF: a unified medical lexicon for French.. *Int J Med Inform* , Mar, Volume 74, Number 2-4, Pages 119-124, 2005.
- [16] Cardillo E, Warnier M, Roumier J, Jamouille M, Vander Stichele RH. *Using ISO and Semantic Web standards for creating a Multilingual Medical Interface Terminology: A use case for Heart Failure*. To be published in *Proceedings of the 10th International Conference on Terminology and Artificial Intelligence (TIA2013)*. October 28, 2013.
- [17] Darmoni, SJ; Sakji, S; Pereira, S; Merabti T; Prieur E; Joubert M & Thirion B. Multiple terminologies in an health portal: automatic indexing and information retrieval. *Artificial Intelligence in Medicine, Lecture Notes in Computer Science*, Pages 255-259, Springer, 2009.

Table 1 HeTOP T/O (included in the UMLS) main figures and translated CUIs in French

Source vocabulary (Translator) [number of languages]	Number of Terms (PT + syn. + acronyms)	Number of CUIs (% Translation in French)
MeSH fr*	105,758	41,229
MedDRA fr [9]	65,071	48,005
WHO-ART fr [5]	3,631	3,091
MTHMST fr	1,833	1,636
ICPC2 [19]	702	722
<b>*MeSH fr HeTOP</b>		
Descriptors (INSERM) [16]	105,274	27,014 (100%)
Supplementary Concepts (CISMeF) [2]	58,514	42,955 (19.29%)
Concepts (INSERM & CISMeF) [2]	99,184	92,437 (26.81%)
<b>Not yet in UMLS</b>		
ICD10 fr (WHO) [12]	26,337	12,143 (100%)
ICD10 PCS (CISMeF) [2]	7,297	7,297 (5%)
ICD9 fr (WHO) [1]	10,716	7,356 (100%)
ICDO fr (WHO+CISMeF) [3]	1,016	1,014 (69.84%)
ICF (WHO) [2]	1,496	1,495
FMA fr (U. of Washington)	4,564	4,452
FMA fr in HeTOP (CISMeF) [7]	15,923	14,781 (18.24%)
ICNP [2]	2,811	1,158 (92%)
SNOMED Int. [2]	139,792	96,756 (94.51%)
ATC (WHO) [3]	5,834	5,758 (100%)
LOINC (APHP) [2]	57,942	57,844 (60.93%)
MEDLINEplus fr (CISMeF & LIMSI) [2]	849	846 (100%)
NCIT fr (CISMeF) [2]	37,545	35,541 (37.94%)
OMIM fr (CISMeF) [2]	7,770	6,890 (88.66%)

**Table 2 HeTOP T/O (not included in the UMLS) main figures and translated CUIs in French**

<b>Source vocabulary (Translator) [N lang.]</b>	<b>Number of Terms fr / en % of translations</b>	<b>Number of CUIs (via CISMef mappings)</b>
WHO-ICPS [2]	424	103
RADLEX [2]	5,948/42,313/14.06	240
HRDO (Orphanet) [2]	13,535/id/100	4,942
HPO (CISMef) [2]	11,100/11,908/93.21	1,541
CCAM (procedure) [1]	10,121	0
BNPC (toxicology) [2]	91,751	11,539