Indexing grey literature in General Practice: Family Medicine in the Era of Semantic Web

Marc Jamoulle[1], Elena Cardillo[2], Ashwin Ittoo[3], Robert Vander Stichele[4], Melissa P. Resnick[5], Julien Grosjean[6], Stefan Darmoni[6], and Marc Vanmeerbeek[1]

[1]**Department of General Practice, University of Liège, Belgium**
[2]**Institute of Informatics and Telematics (IIT), National Research Council, Italy**
[3]**HEC Management School, University of Liège, Belgium**
[4]**Heymans Institute of Pharmacology, University of Ghent, Belgium**
[5]**School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA**
[6]**Department of Information and Medical Informatics (D2IM), University of Rouen, France**

Corresponding author : marc.jamoulle@gmail.com

ABSTRACT

Problem/Goal: Sharing the results of research with General Practitioners (GPs) is crucial for the survival of the discipline of General Practice / Family Medicine (GP/FM). The production of abstracts in GP/FM probably exceeds 15,000 per year worldwide. Each abstract often represents two years of work for its authors and is expressed in local languages. Only 45% of them are published in indexed medical journals. Usual indexing systems like MeSH are not multilingual nor adapted to the particular field of GP/FM. Consequently, these abstracts are lacking bibliographic control and more than half of the research presented by GPs at congresses is lost. Considering the absence of appropriate domain-specific terminologies or classification systems, we propose a new multilingual indexing system. The existing International Classification of Primary Care (ICPC) is currently used for clinical purposes and has now been expanded with a taxonomy related to contextual aspects (called Q-Codes) such as education, research, practice organization, ethics or policy in GP/FM, currently not captured. The set is proposed under the name Core Content Classification in General Practice (3CGP). The aim is to facilitate indexing of GP/FM specific scientific work and to improve performance in information storage and retrieval for research purposes in this field.

Research Method/Procedure: Using qualitative analysis, a corpus of 1,702 abstracts from six GP/FM- related European congresses was analyzed to identify main themes discussed by GPs (e.g., continuity, accessibility or medical ethics), handled in a domain-specific taxonomy called Q-Codes and translated in 8 languages. In addition, a methodology for building a lightweight ontology (in OWL-2) was applied to Q-Codes, adding object and datatype properties to the hierarchical relations, including mapping to the MeSH thesaurus, Babelnet (www.babelnet.org) and Dbpedia. Finally, the Q-Codes in 8 languages have been integrated healthcare terminology service (www.hetop.eu/q) with a companion website (http://3cgp.docpatient.net).

Anticipated Results of the Research: The creation and the on-line publication of this multilingual terminological resource, for indexing abstracts and for facilitating Medline searches, could reduce loss of knowledge in the domain. In addition, through better indexing of the grey literature (congress abstracts, master's and doctoral thesis), we hope to enhance the accessibility of research results of GP/FM domain and promote the emergence of networks of researchers. First result of experimental implementations of the new indexing system will be presented.

Indication of costs related to the project: This project has not been funded.

keywords General practice, Terminology, Electronic publishing, Repository, Grey Literature.

CONTENTS

1 BACKGROUND

1.1 Need for information in family medicine

In the cycle of patient centered information (Jamoulle et al., 2015), the General Practitioner (GP) is simultaneously a heavy user and producer of published/unpublished data. Data could be clinical, (i.e. dealing with symptoms, processes and diseases) or contextual. Contextual data can address particular issues concerning the patient, which may influence the process of care (Schrans et al., 2016). However, contextual data can also deal with issues concerning the doctor, the managerial aspects of care. In particular, it could address the position of GPs within the health care system, the general concepts used in Primary Health Care (PHC), or the delivery services. In this work, the focus will remain on these last contextual medical features of General Practice / Family Medicine (GP/FM), as its tools for training, research, ethics, inquiry, environmental issues, infrastructure and principle of care. These features are central to this field, and family doctors are used to exchange information over them when they meet in training sessions or during congresses.

The realm of GP/FM differs from mainstream health care, as Family Physicians (FPs) address biological, technological, behavioral, sociological and anthropological domains. All of these have a deep impact on the terminologies needed (Helman, 2008; Thompson et al., 2014). As the creation of already available terminologies was focused on specialized domains, the biological and technological fields of medical terminologies are now almost complete (Jonquet et al., 2016; Lelong et al., 2016). However, they sometimes fall short when applied to the field of GP/FM, which relies intensely on complexity and timeline issues (Liang et al., 2014; Madkour, Benhaddou, and Tao, 2016). Albeit well documented clinical

issues, professional contextual issues, like management, teaching, research, and ethics are documented in a fragmented way for the first level of care (Jamoulle et al., 2017a).

1.1.1 GP/FM, a profession without clear limits

Despite elaborate definitions of GP/FM Allen et al. (2011) and Primary Care Physicians (PCPs) (AAFP, 2011), the manner in which the profession of GP/FM or PCP is defined and structured varies greatly across family medicine textbooks (Casado Vicente, 2012; Gusso and Lopes, 2012; Kochen, 2012; Murtagh, 2011; Druais et al., 2009; David et al., 2013; Lakhani, 2003; McWhinney, 1997). This is especially true in regard to managerial and contextual features. These textbooks have offered a top-down expert view of the profession, as the authors of those textbooks themselves chose the subjects addressed. In this research, we rely on what practicing doctors are interested in. In this sense, one can speak of a bottom-up approach.

If one examines the table of contents of these cited works, as far as the general management and the contextual background are concerned, the quoted books are absolutely different. One focuses on communication, the other on the systemic approach, and the third one on the ethics of relationships. None give a similar view of the scope and contextual scope of family medicine. Also, technology is often absent. Only one author or another approaches current technical processes in family medicine.

An extensive review of the vocational training programs in the specialty of General Practice was not done. The author, however, knows from experience and the many contacts he has in family medicine in Europe/the world that these programs have no homogeneity, despite the recommendations of EURACT, the WONCA Europe working group on education (Heyrman, 2005). Also, when considering Continuous Medical Education, the drug industry's influence on the choice of subjects is decisive, which creates multiple conflicts of interests (Davis, 2004). Therefore, it seemed wise to develop an index of concepts dealing with family medicine by listening to practicing GPs, and to develop a bottom-up approach rid of conflicts of interest.

1.1.2 Published and unpublished in GP/FM, what's the meaning?

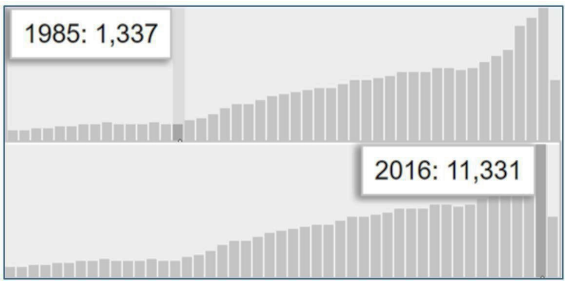
Medical Subject Headings (MeSH) are normalized keywords directed to enter queries in Medline, the bibliographic data base of the National Library of Medicine (NLM) through the interface PubMed (Lowe and Barnett, 1994). Currently the PubMed interface gives access to 27,3 millions of citations. A search with 7 GP/FM relevant MeSH in June 2017 brings a little less than 200.000 citations. The same interrogation with 8 specific MeSH descriptors of Primary Health Care (PHC) gives 480.000 citations.(PHC and GP/FM related MeSH are listed in Fig.1). Together the 15 specific descriptors of the first line of care gather less than 2% of Medline content.

Figure 1. GP/FM & PHC related MeSH in PubMed

Types	MeSH	Year of introduction
Primary Health care (PHC) related MeSH	"Community Health Service" [MeSH:Noexp]	1967
	"Community Mental Health Services" [MeSH:NoExp]	1967
	"Home Care Services" [MeSH:NoExp]	1967
	"Primary Health Care" [MeSH:NoExp]	1974
	"Community Health Centers"[MeSH:NoExp]	1979
	"Community Mental Health Centers" [MeSH:NoExp]	1979
	"Home Care Agencies" [MeSH]	1995
	"Rural Health Services" [MeSH:NoExp]	1996
General Practice / Family Medicine (GP/FM) MeSH	"Physician, Family" [MeSH]	1974
	"Community Medicine" [MeSH]	1977
	"Family Practice" [MeSH]	1978-2010
	"Gatekeeping" [MeSH]	2000
	"General Practice" [MeSH]	2011
	"General Practitioners" [MeSH]	2011
	"Physician, Primary care" [MeSH]	2011

The rise in publications in GP/FM has been inevitable since 1985 (see Fig. 2). The quoted publications considered relevant to GP/FM are obtained through the use of a search strategy composed of seven MeSH concepts related to GP/FM (see Fig. 1). The citations obtained could be certainly appraised as *published* data and claimed white literature. This raises question about the Pisa definition of grey literature, considered as *not controlled by commercial publishing* (GreyNet, 2014), as one can hardly pretend that papers published by the NLM are always commercially controlled.

Figure 2. Evolution of publications in GP/FM (7 MeSH) from 1985 to 2016 on PubMed



Unpublished data in the GP/FM field are numerous. Yet, GP/FM organizations are heavy producers of continuous medical education (VanNieuwenborg et al., 2016). They contribute greatly to training sessions and organize local, regional and national level medical conferences, as well asto research meetings (Buono et al., 2013), virtual conferences (Cavadas, Villanueva, and Gervas, 2010), websites, and blogs. They are also active on social networking (Veuillotte et al., 2015). With so many heavy producers of various nationalities, it is important to note that at local and national events, the local language is the rule.

It is not a secret that knowledge translation between doctor and patient is highly controlled by pharmaceutical companies (Moynihan, 2003) (Moynihan and Bero, 2017). Despite the activities of GP organizations, in some countries, domestic papers as medical newspapers remain the main sources of information for practicing doctors (Tabatabaei-Malazy, Nedjat, and Majdzadeh, 2012). Usually, they are edited within an atmosphere of heavy, silent corruption (Angell, 2017). The translation of information by drug representatives is also a determining factor (Greenway and Ross, 2017) as well as predatory open access publications (Shen and Björk, 2015) or pay for publishing process (Quan, Chen, and Shu, 2017) which dismisses whole sectors of publication. This implies that knowledge management tools in GP/FM must be controlled by dedicated, unbiased GPs. The movement of free lunch doctors (`http://www.nofreelunch.org/`(USA)), the no-gracias one (`http://www.nogracias.eu/`(Spain), the Medico Sin Marca (Chile), (`http://www.medicossinmarca.cl/`), the Therapeutic initiative (Canada) (`http://www.ti.ubc.ca/`) or, more generally, the members of the International Society of Drug Bulletins (ISDD) (`http://www.isdbweb.org`) refuse to adhere the pharmaceutical industry influence. Their work merits distributing its grey literature in a professional multilingual indexing system
that is accessible by all. Finding a way to publish and share information outside the default language of English is attractive to many GPs. However, sharing the results of research with General Practitioners is crucial for the survival of the discipline of GP/FM, which means a universal system must be created (McIntyre et al., 2016).

As an example of hidden grey literature, the website of the World Family Doctor Association in Europe (`www.woncaeurope.org`) edits more than 30,000 non-indexed abstracts of European or world conferences in English. Each abstract often represents two years of work (Master's theses are included in this). If not published, they become lost work. This also represents missed opportunities to develop networks between authors that share similar interests. Only half of this production has the chance to be published in indexed medical journals (Van Royen et al., 2010; Hummers-pradier, 2007).

1.2 Producing Information at the point of care

It could be believed that a simple terminological subset may be sufficient to meet the needs of GP/FM computer systems. Lack of visibility of the complexity of the work of family doctors has allowed for
such biased vision. Moreover, special tools specifically dedicated to primary health care, such as the International Classification of Diseases, tenth revision, for primary care (ICD-10 PC) (Ustün et al., 1995) or the Statistical Manual of Mental Disorders, 4th ed., primary care version (DSM IV PC) (Pingitore and Sansone, 1998) have been developed and nicknamed quickly ICD-10 or DSM-IV for dummies.

Of course, the family physician often sees simple problems. But he is accompanying a set of patients throughout a lifetime. The family physician knows a patient more extensively than most specialists do. He also becomes a specialist in patients bearing rare diseases and of various cultural background. He will, therefore, have extensive terminological needs, even more extensive than many specialists.

1.2.1 Clinical information

The adjective clinical deals here with patient related data, such as: reasons for encounter, symptoms, acts performed or requested and diagnosis. Terminology for clinical information is a highly specialized and difficult

field of current medicinal research (Jamoulle et al., 2014). Clinical information is accumulated in Electronic Medical Records (EMRs) and, if well organized, transferred to study centers where huge database may be used to teach medicine, analyze epidemiological data or be used for secondary searches (Charlton et al., 2010; Carey et al., 2004; Britt et al., 2003). Among others, studies worth citing are, mostly with good validity (Khan, Harrison, and Rose, 2010), produced by the Dutch Transition Project (Soler et al., 2012) (`http://www.transhis.nl`), data produced by the Belgian Intego project (Bartholomeeusen, Buntinx, and Heyrman, 2002) (`https://intego.be/en`) , the Beach project in Australia (Britt et al., 2016) or at a larger scale the UK General Practice Research Database (`http://gprd.com`).

1.2.2 Professional contextual information

The term *contextual* applies as a generic term for the name of the taxonomic product presented here. A concept like *uncertainty*, the usual companion of the doctor or the concept of *quality assurance* or *environmental health*, are all essential elements of professional practice. These are not clinical terms as they do not always deal with current patient problems. The term *contextual* appeared the most relevant, as it was defined in the Meriam-Webster Dictionary as: *the interrelated conditions in which something exists or occurs*.

Answering the question; *What are they discussing?* in a meeting of two, twelve or several hundreds or thousands of GPs may give insight into the details of this well-defined (Jamoulle et al., 2017b) but not limited profession. As stated by Cimino (1998) : *Part of the difficulty with using a standard controlled vocabulary is that the vocabulary was created independent of the specific contexts in which it is to be used*. Adding a contextual supplement to ICPC gives birth to an extended *Controlled vocabulary*, able to take in account the extension and the complexity of the domain covered by GPs. Controlled vocabulary *is a general term for a list of standardized terms used for indexing and information retrieval usually in a defined information domain* (Library and Archives Canada, 2017). In this case the Controlled vocabulary is also a *Vocabulary coding scheme* as defined in Dublin core (see further).

This approach to the doctor's professional context should not be confused with the contextual approach of the patient's universe, as developed by Schrans et al. (2016) who studied the elements of the patient's life context that influence his or her state of health and the health problems he shares with the doctor.

1.3 Consuming information at the point of care

As stated by James (2016), the Internet has triggered a transformational change in the dissemination of science in the form of a global transition to open access (OA) publishing. GPs, the rank and file (Chinitz and Rodwin, 2014) workforce in medicine are using those resources extensively despite sometimes huge material difficulties to access the sources when working in rural or remote areas (Salman Bin Naeem, Shamshad, and Amjid, 2013).

Finding information is sometimes difficult for researchers, even though they may not work with patients and remain mostly in academic laboratories with access to expensive medical journals. So, what about *rank and file* physicians who have only a few seconds to check information and source relevance? (Hubbard, 2008). Availability in the real world, at the point of care, is often clashing with the economic model (paid access) or with copyright issues (Myška and Šavelka, 2013).

Open access (OA) to researchable and usable information (Heilman, 2015) at the point of care is of utmost importance in GP/FM. It is necessary to maintain open access, point of care resources at the high level of quality that patient care demands (PLoS Medicine Editors, 2015). A potential issue with this is that GPs should consider that daily medical journals as well as major papers can also be manipulated by the pharmaceutical industry (Schwitzer, 2017; Dowden, 2015).

1.3.1 Sources of information at the point of care in GP/FM

On-line, directly accessible major documentary databases in open access and local language are not numerous nor specific to the GP/FM field (Hubbard, 2008). The US NLM PubMeD database (`https://www.ncbi.nlm.nih.gov/pubmed/`) gives access to millions of paid and open access publications, mostly in English. The Pan American Health Organization sponsors Scientific Electronic Library on-line (`http://www.scielo.org`). SciELO is the South American champion of free access journals, mostly in Spanish (PAHO Bireme Sao Paulo, 2016). The World Health Organization (WHO) supports the Global Index Medicus (`http://www.globalhealthlibrary.net`). In France, LiSSa (`http://www.lissa.fr`) gives access to more than 10^6 citations of French literature (Cabot et al., 2017a).

To say nothing of Google Scholar, the web is a considerable resource of information in medicine, especially in gray literature. Janamian et al. (2016) have identified and searched for 260 web sites as GP/FM sources. As stated by González-González et al. (2007) "*In primary care, each practitioner encounters more than 500 different clinical topics in any year*". Ten years ago Internet base accounted for only 5% (ibidem) for

search of information by Spanish GPs. This percentage has risen to 59% for German GPs in 2016 (Eberbach et al., 2016). Anyway the use of Internet is now so generalized that studies are performed on influence of Internet the patient-physician relationship (Tan and Goonawardene, 2017).

1.3.2 Use of Medical Subject Headings (MeSH) descriptors in GP/FM

For retrieval of scientific bibliographic information in GP/FM, MeSH descriptors are used by all doctors and researchers. The MeSH is a huge thesaurus of 27,000 hierarchically managed descriptors- i.e., normalized terms, intended to index medical documents and growing yearly (Jamoulle 2016).. Currently, the use of indexing civil data in the health care domain is being studied (Marc et al., 2015).

General Practice is a profession generally regarded as part of the first line of care, PHC, a form of care organization. General Practice and Primary Health Care concepts share the same extension, but not the same intension. The first describes the duty of a profession, the second the management of a service (Jamoulle et al., 2017b).

The confusion could be found in the MeSH thesaurus. Gill et al. (2014) state that: *Constructing a highly efficient search filter to identify primary care relevant articles is challenging, particularly due to the inadequate and ambiguous description of the clinical research setting in title, abstract and MeSH keywords*. Huang, Névéol, and Lu (2011) observes that: *manually assigning MeSH terms to biomedical articles is a complex, subjective, and time-consuming task*. Shultz (2007) argues that: *terminology was observed to be a major factor affecting retrieval and the ability of both systems to obtain unique items*. However, not all aspects of the broad field of GP/FM are covered in a specific area (Sladek et al., 2006). Despite this, interesting advance in MeSH indexation for GP/FM have been proposed (Mendis and Solangaarachchi, 2005) (Jelercic et al., 2010). Theme of interest searched for by GPs have also been studied (Hong et al., 2016). The future looks prepared as MeSH are now available in RDF, ready for Semantic web (Bushman, Anderson, and Fu, 2015).

1.3.3 Use of Health descriptors (DeCS)

The controlled vocabulary of Health Sciences Descriptors (DeCS), initially a translation of MeSH into Spanish and Portuguese, has been expanded with new categories. It was also adopted into the indexing and multilingual search of the scientific and technical literature in South America (`http://decs.bvs.br/I/homepagei.htm`). It's a by-product of the Latin American and Caribbean System on Health Sciences Information (BIREME), the Pan American Health organization (PAHO)
network of libraries and documentation centers (Neghme, 1975). Doctors and students in South America are using DeCS as a standard controlled vocabulary for indexing scientific and technical health-related documents in knowledge databases like the Virtual Health Library (`http://bvsalud.org/`) or SciELO Both sources generally give open access to documents.

1.3.4 ICPC for Classifying Clinical Issues in GP/FM

The International Classification of Primary Care (ICPC) (Soler, Jamoulle, and Schattner, 2015) is routinely used by physicians around the world to categorize the problems encountered in their practice with patients, i.e. their clinical activity. We will see that ICPC has proved effective in many other uses and show that it is adapted for collecting clinical problems that doctors discuss at congresses.

Primary care isn't specialized care. It encompasses specialized care and biotechnological vocabularies along with anthropological, family-based/personal information. A medical record in primary care must encompass all facets of health care, alongside personal and family environment knowledge. This is why, WONCA, the world organization of family doctors, through the ongoing work of its International Classification Committee (WICC), has developed ICPC (Wonca, 1987) (WONCA, 2005) (Lusignan, 2005).

1.4 Grey Literature as source of information

The price of access to international high-level journals, mostly exclusively in English, is prohibitive. The economic aspect is therefore a major obstacle to the spread of knowledge outside academic circles. However, the change of economic model is under way in the world of publishing, the cost of which is largely reflected on the author and not the reader. In the meantime, appeals for public access of research data continue to proliferate (Lin and Strasser, 2014). McKenzie (2017) considers that the explosion of the use of Sci-Hub facilities (`http://sci-hub.cc`) is the beginning of the end of the scholarly publishing.

Whatever the case, the open access grey literature in medicine is in full development (Swan, 2012). According to Schöpfel (2015), *The term gray literature remains ill-defined, imprecise, with fuzzy outlines. Its two handicaps are part of its definition: identification, access and acquisition are often difficult, and quality and reliability are not always assured.* Ferreras Fernández (2016, p.211-216) has done an exhaustive review of the

definitions of grey literature. Those definitions share the negation of commercial involvement like Pisa's (GreyNet, 2014). Grey literature reviews are not always free of access like the Grey Journal itself (`www.greynet.org`).

Practically, when addressing the case of grey literature, authors exchange more pragmatic definitions than the Pisa's one as; *difficult to locate or retrieve* (Moher et al., 2000), or; *has not been formally published* (Hopewell et al., 2007) or; *there is no such peer review or passage through quality filters* (Silva, Garcia, and Cássia, 2009).

Interestingly, Hoffmann et al. (2011) points out that the *grey literature yields more substantial infor- mation* (than white literature) *on the content of interest*. This could be understandable, partly as white medical literature is not free from the influence of the industry (Gotzsche, 2013; Schwitzer, 2017).

We propose to consider grey literature in GP/FM publications that share the following characteristics:

- For the background:
  - Sharing knowledge specific to the field of GP/FM, no matter the format (papers, articles, memo, master thesis, PhD thesis, leaflet, abstracts of presentation, web pages, video, images, YouTube, Facebook, Twitter, Google+, LinkedIn, dataset).
  - Being unreferenced in well-known local or international medical databases (PubMed, LiSSa, Scielo, Lilacs, ORBI, etc.).
  - Being submitted to a scientific quality assurance process (in anthropology or bio-sciences).
- For the format:
  - Being freely accessible in an Open access model.
  - Using a systematic multilingual vocabulary encoding scheme (indexing system).
  - Relying to Dublin Core Metadata Initiative or equivalent standardization process.
  - Being ready for machine use in the semantic web.

1.5 Metadata and Vocabulary Coding Scheme

Metadata consists of statements we make about resources to help us find, identify, use, manage, evaluate, and preserve them (Sutton, 2007). Metadata may be interpreted by machines and people. Dublin Core Metadata Initiative (DCMI) (`http://dublincore.org/`) provides simple standards to facilitate the finding, sharing and management of information. Metadata are basic description mechanism for digital information that, can be used in all domains, for any type of resource, simple, yet powerful, can be extended and can work with specific solutions, making it easier to find information on the Web as it develops. DCMI participates in the development of the "new Web", the Semantic Web and Linked Data (Dekkers, 2009). Allen (2016) states that *The emergence of machine intelligence and machine reading in the second machine age will make it even easier to automate the production of metadata to help people find, filter and organize information.* .

Quality of search results is dependent on the quality of the metadata in the original repositories of which high quality structured metadata are more accessible (Farace and Schöpfel, 2010).

We are thus dealing with knowledge identification process by humans and by machine through well formalized denominations. So we are addressing here the concept of *Controlled Vocabulary*. The reader has to be conscious that the same concept could bear different names. For instance the Australian Metadata Online Registry (MeteOR) uses the term *Classification scheme* for pointing the same issue. *A classification scheme is an official terminological system, recognized and endorsed by a national or international body, that is used to classify data.* (`http://meteor.aihw.gov.au`).

1.6 Grey literature and Semantic web opportunities

Metadata allow the retrieval from data from dedicated repositories. Nevertheless, as stated by Goggi et al. (2015), *documents may contain important information that has not been encoded in the metadata.* Extracting key concepts from unstructured texts is the following step, done by semantic annotators, by-product of research in Natural Language Processing (Cabot et al., 2017b). Key concepts could be added to indexing facilities or tagged as identifiable information for use in Linked Open Data (LOD). This open the *possibility of enhancing the visibility and accessibility of grey literature via its connection to the data it describes and to an advanced full text indexing* (Goggi et al., 2015).

As stated by (Cardillo, 2015) : *During the last ten years ontologies and the use of Semantic Web technologies has been seen as a better solution to semantic interoperability because this allows describing the semantics of information sources and makes its contents explicit by providing a shared comprehension of a given domain of knowledge [. . . . .] Unfortunately, ontologies and their structure are not really familiar and natural to most healthcare providers and their use raises heterogeneity problems to a higher level.*

2 AIM OF OUR RESEARCH: PROPOSAL FOR A NEW CODING SCHEME IN GP/FM

Our work aims to identify the themes in knowledge production by GPs in a new Vocabulary Coding Scheme called Core Classification of General Practice Family Medicine (3CGP). This program encompasses clinical and contextual situations in the GP/FM practice. Simultaneously, we hope to develop our system in such a way that machines (i.e., computer), could deal with that data and reason about it using the Semantic web technologies.

As mentioned above, classifications and terminologies for patient data retrieval are numerous. However, one must know if they could be used to index and retrieve documents in a specific manner. Indeed, units of knowledge managed in historically different terminologies and classifications are interlinked and address the same reality seen by various eyes and interests (Bowker and Star, 1999).

The absence of adapted concepts and descriptors for contextual aspects of GP/FM is one of the reasons why the scientific work of family physicians is hard to retrieve from mainstream bibliographic systems. In addition, more than 50% of the scientific output of GPs at conferences is never published (Van Royen et al., 2010). There are no dedicated indexes of grey literature (Mahood, Eerd, and Irvin, 2014), and abstracts or collections of dissertation titles are often not properly indexed in this field (Lawrence et al., 2014).

This work presents a new taxonomy of contextual aspects of GM/FM, in hopes of helping to improve the situation surrounding GM/FM grey literature. Taxonomies provide schemes to help classify entities and define the relationships between them (Dixon, Zafar, and McGowan, 2007). The purpose of this development is also to provide tools to exploit modern technology - in terms of terminology for information storage and retrieval systems (Vanopstal et al., 2011), such as: machine learning, semantic web techniques, natural language processing (NLP) and linking data. This kind of system is already in use in clinical settings for patient data (Colliers et al., 2016) and one hopes to apply such techniques to an indexing system in a near future for the communication of family doctors in congresses and related grey literature.

In brief, our aims are triple:

- To improve annotation of grey literature in primary care.
- To facilitate indexing of congress abstracts and theses.
- To improve the searchability of repositories for these information artefacts.

3 METHODS

3.1 Referring to METHONTOLOGY steps for the development of the project

The phases of development of the project are shown on Fig. 3 along the time line. Qualitative analysis of communications of GPs during congresses has induced the creation of a controlled vocabulary organized in a taxonomy. To develop a domain-oriented taxonomy (the simplest form of an ontology- i.e., a light- weight ontology), methodology for ontology construction was included (Gómez-Pérez, Fernández-López, and Corcho, 2003). The four main phases of the METHONTOLOGY process are shown Vertically : 1.Knowledge Acquisition and formalization;
2. Integration process;
3. Implementation;
4. Publication and Dissemination

Knowledge acquisition, formalization and integration were added in 2005. The implementation phase in the online Hetop server began in 2014. We have added a dissemination phase through Internet and publications.

3.2 Knowledge acquisition & formalization; Qualitative analysis of GPs' communications by a Computer-Assisted Qualitative Data Analysis Software (CAQDAS)

Using qualitative analysis, a corpus of 1,702 abstracts from six GP/FM-related European congresses was analyzed to identify 182 themes discussed by GPs (e.g., continuity, accessibility or medical ethics), handled in a domain-specific taxonomy called Q-Codes and translated into 8 languages. To identify key concepts in a domain-specific taxonomy, data is analyzed in a grounded theory approach (Glaser and Strauss, 1999). This approach is often used in disciplines

such as: economics, law, and medicine (Wells, 1995; Denzin and Lincoln, 2000). It involves the construction of a hypothesis or discovery of concepts through data analysis (Faggiolani, 2011; Martin and Turner, 2016). After a careful study of existing products, the qualitative analysis software (ATLAS.ti®http://atlasti.com/) was used, as it enabled therequired analyses to be executed at a relatively low cost. ATLAS.ti enabled the ability to map specific words to already-defined ICPC-2 and to find new concepts to feed the new Q-Codes taxonomy. The same theme could not reappear in the same abstract more than once, and (generally) no more than six themes were identified in each abstract. The analysis performed by EGPRN in 2010 on 614 abstracts, using a similar approach, has been used to control the QR (Research) domain and check the consistency of the Q-Codes proposal.

Two additional steps are required to complete the lightweight ontology (taxonomy) construction process according to METHONTOLOGY, namely: Integration and Implementation.

3.3 Integration phase, birth of 3CGP

The Core Content Classification in General Practice/Family medicine (3CGP) is formed by the addition of ICPC-2 for clinical issues and Q-Codes for professional contextual issues, both discussed during meetings between GPs. The Q-Codes taxonomy was elaborated on the model of ICPC, using the letter Q to categorize the contextual elements, for the letter Q was unemployed in ICPC-2.

$$\text{ICPC-2} + \text{Q-Codes} = \text{3CGP}$$

Figure 3. The phases of development of the project on a time-line. The four main phases of the METHONTOLOGY process

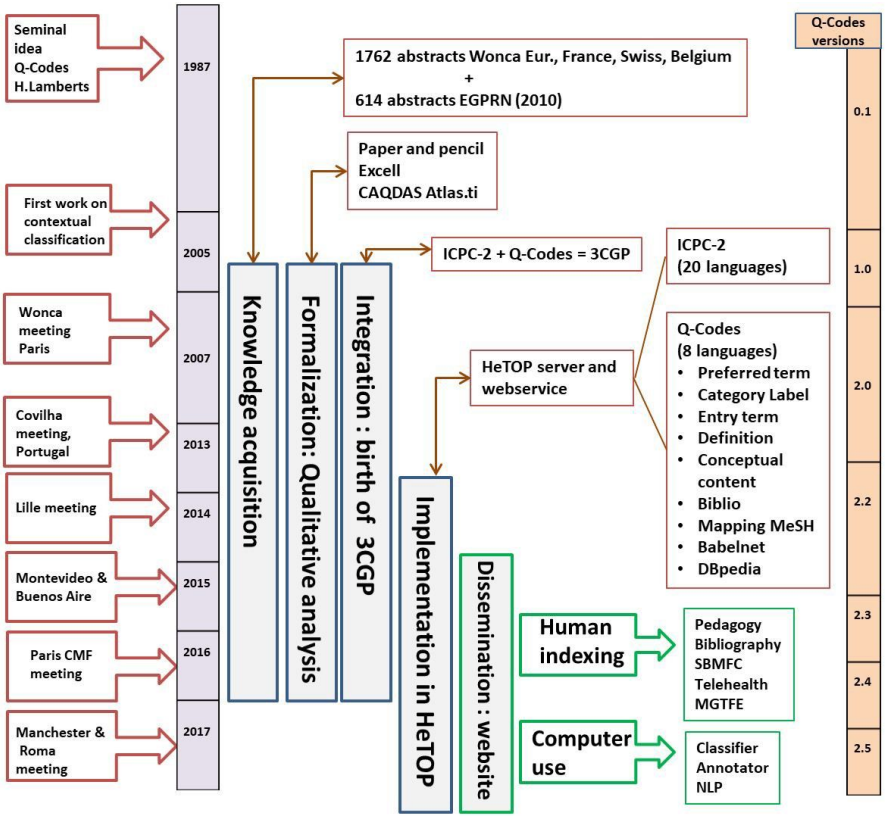
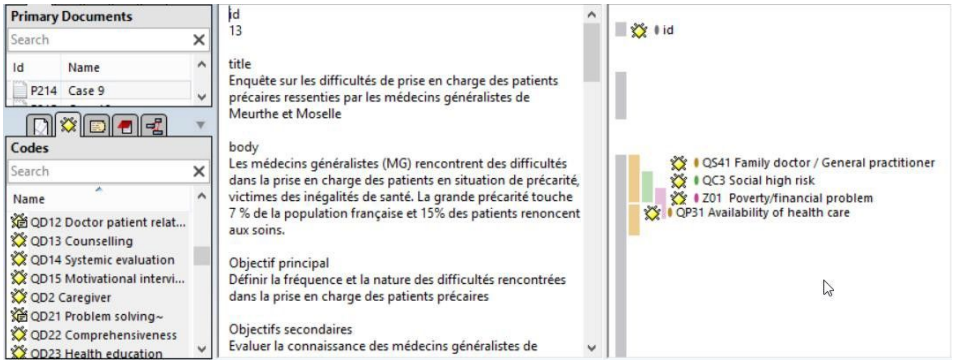
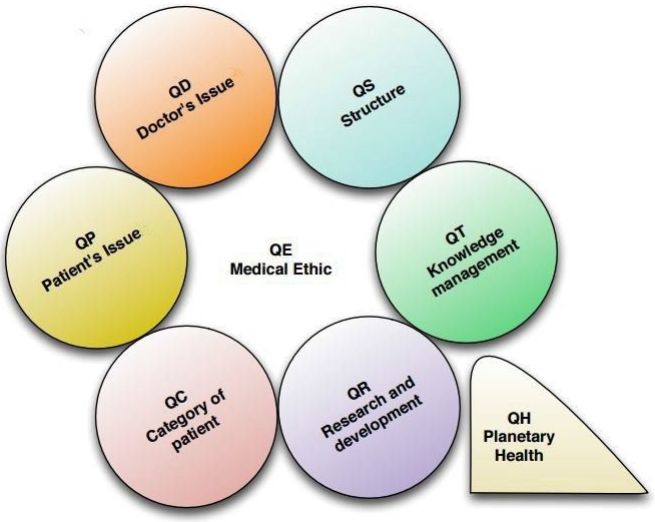


Figure 4. Congress CNGE 2013: After being scanned, coded themes appear on the right column: Q-Codes (QS41, QC3, QP31) and one ICPC code (Z01). Here, the theme identified is the offer of family medicinal services in vulnerable populations. At the bottom left, the software proposes the list of pre-registered Q-Codes. (ATLAS.ti Software)



Q-Codes are born from the analysis of 1,702 abstracts of conferences. Naturally, we hope that future conferences will allow for a surge in new concepts and new entries in the Q-Codes classification. Q- Codes are divided into 8 domains. The taxonomy starts with the QC domain, which represents *Patient's category*, and covers topics such as age, gender issues, and victim-hood. The second one is the QD domain, representing *Family doctor's issue*, which covers issues such as disease management, communication, clinical prevention, and medico legal issues. QE represents *Medical Ethics*. This domain covers bioethics, professional ethics, and info-ethics. The fourth domain is QH, representing "Planetary Health", which deals with such areas as environmental health, biological hazards, and nuclear hazards. The fifth domain is QP, *Patient Issue*, which includes patient safety, patient centeredness, and quality of care. The QR domain is *Research & Development*, covering research methods, research tools, and epidemiology of primary care. QS is the Structure of Practice domain. It covers topics such as primary care settings, primary care providers, and practice relationships. Finally, the QT domain is Knowledge management. This domain deals with teaching, training, and knowledge dissemination. Each domain of Q-Codes is divided in Categories, Sub-Categories, and Sub-Sub-Categories. A ninth domain, QO for Other will be used in the abstract coding process for not precise descriptions or for a concept worth to be considered as a potential candidate for a new theme.

Figure 5. The Q-Codes matrix in the shape of a Q-Letter.



The presentation of the Q-codes under a matrix format is shown on Fig. 5. The matrix takes the shape of the letter Q, representing the 8 domains of the Q-Codes. On the left, the people related domains - Doctor's issue, Patient's issue and Category of patients; On the right, the managerial related domains - Structure, Knowledge management including Teaching and Training, and Research and development; Hazards are the underlying Planetary health conditions represented by the downward oblique tail stylized as a triangle but which are in reality the back-ground of the GPs work; in the center, joining all, Medical Ethics. Note that the Q's tail, which is the Planetary Health, prevents the wheel from turning endlessly. This is a nice demonstration of the importance of the environment on health issues (Graphic design Patrick Ouvrard).

3.4 Implementation ; Organizing the concepts of the taxonomy following a Data Struc- ture Diagram on the HeTOP server

Integration and implementation came to fruition in the meantime. The ICPC-2 classification was edited on the HeTOP web site in 22 languages and the Q-Codes in 9 languages; French, English, Dutch, Spanish, Portuguese, Vietnamese, Turkish, Georgian and Korean. More are coming (Greek, German, Italian, Ukrainian). The Data Structure Diagram, a graphic technique, based on a type of notation dealing with classes of entities and the classes of their relationships (Bachman, 1969) has been used to organize the mappings. In Fig.7, the central concept (here, Overmedicalisation), is linked by its relations (is a - consider - has a definition, conceptually related to) to other formally defined fields of knowledge. This kind of structure is machine readable and forms the basic structure of our taxonomy. It is presented in Excel format in Fig.6.

3.5 Dissemination; The HeTOP server as a GP/FM knowledge resource

The HeTOP server, produced by the Department of Medical Information and Informatics (D2IM) of Rouen University (France) is edited in the Web Ontology Language (OWL), which allows for the linking of data with other data (McGuinness and Harmelen, 2004). HeTOP is based on a multi-terminology meta-model that integrates all terminologies and ontologies into its data core. It is cross-lingual since terminologies and ontologies are often available in several languages. The web site can be used by both humans and machines via a dedicated web service (`http://www.hetop.eu`). HeTOP currently contains 71 health terminologies and ontologies (only 17 are included in UMLS as most of them are French terminologies),

Figure 6. The 14 fields of each Q-Code in the HeTOP interface of which conceptual links are described in Fig.7.

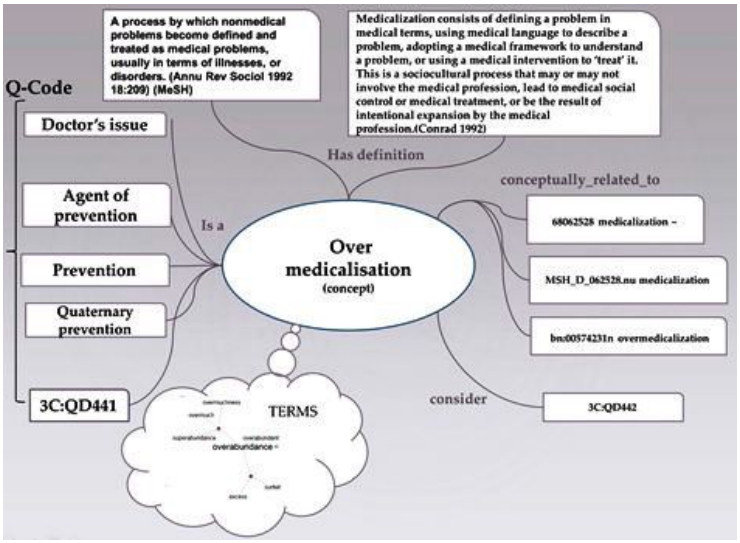
HeTop field	Signification	Remark
Category ID	Alphanumeric identifier of the Q-code	Letter Q followed by one letter (C,D,P,S,T,R,H,O) followed by maximum four numeric digits
Category label	Full title of the category	First letter capital, plural possible, can be compound words
Preferred term (PT)	Normalized title of the category	English, masculine, singular, lower case
Preferred term, other language	Translated (es, pt, nl, vn, tr, fr) version of the PT (more allowed)	Will be used as text word [TW] search term in PubMed equation (in English)
Supplementary entry term(s)	Additional search term(s)	Will be used as text word [TW] search term in PubMed equation(in English)
Category definition	Definition of the title of the category	Reflecting the corporate culture of GP/FM
Category conceptual content	Set of definitions or descriptions from online nomenclature, dictionaries or thesauri	Shows the extension of the concepts, includes the curated MeSH definitions
Automatic HeTop interterminologic relations	Automatic mapping of the PT by the HeTop embedded terminologies	Each proposal could be checked accepted or refused
Terminological features	Broader Than Narrower Term (BTNT) or Narrower Than Broader Term (NTBT)	Establish the hierarchical position of the PT mapping and links it to the HeTop semantic network
Curated MeSH	Accepted MeSH(s) which meaning correspond to the defined content of the PT	Will be used as [MH] in the PubMed equation automatically generated by the HeTop system
Refused links	Manually terminological links judged non-convenient	Will not be used in the search equation
Bibliographic free full text links	URL of citations of free full texts highlighting the content of the Q-Code	Generally chosen in PubMed but also through Google scholar
Babelnet.org link	Link to the URI of the corresponding babelnet.org entry(ies)	Map the Q-Code to an extensive semantic network of multilinguistic knowledge
DBpedia.org link	Link to the URL of the corresponding DBpedia entry(ies).	Map the Q-Code to a major knowledge semantic database. Could be Wikipedia if DBpedia missing

2,538,595 concepts, 9,982,113 terms, 10,120,417 relations and 32 managed languages. This cross-lingual terminology server is dedicated to various usages by different types of users: translators, students, teachers, researchers, librarians, physicians, etc. HeTOP allows users to search and browse Health terminologies and ontologies in a second (Grosjean et al., 2012). Cross-linguality allows matrix navigation: among terminologies, but also among languages. HeTOP is a multilingual terminology server that not only allows one to search for concepts in several terminologies (and several languages) at the same time, but represents their interoperability.

(Friedman et al., 1999). The content of the HeTOP database could be downloaded in CSV (EXCEL), RDF, SKOS or OWL format.

Thanks to contacts all around the world, the collaboration with D2IM team allowed for the completion of the online publication of ICPC-2 in 20 languages (Schuers et al., 2015), Ukrainian and Greek being the last ones. Some colleagues went a step further and translated the HeTOP interface, allowing health professionals to use it in their native language to access ICPC-2. Short after, Q-Codes took place as the last born of terminologies on the HeTOP server. Adding Q-Codes to ICPC-2, it was then possible to develop a complete terminology adapted to GP/FM and PHC needs. Although WONCA retains copyright on the use of ICPC, every effort should be made to disseminate it. The HeTOP database is freely accessible by means of a simple registration process. The Q-Codes belong to the author, but they are licensed under a Creative Commons Attribution Non Commercial (CC-BY-NC) licence.

Figure 7. Data structure diagram (DSD) of a Q-Code, showing the map of concepts and their relationships (conceptual data model)



3.6 ICPC-2 & Q-Codes available under URI format

Each HeTOP rubric could be also expressed under an Unique Ressource Identifyer (URI) format (see Fig. 8. A Uniform Resource Identifier (URI) is *a compact sequence of characters that identifies an abstract or physical resource* (Berners-Lee, Fielding, and Masinter, 1998). It is a string of characters used to identify a resource (Miller, 1998) A URI identifies a resource by either location, name, or both. In addition to identifying a web resource, a URI specifies the means of acting upon or obtaining the representation of it. Each entries of ICPC-2 and Q-Codes individual rubrics are available under URI format on the HeTOP server. The chain of character is stable. Languages are expressed under the ISO 639-3 Codes for the representation of names of languages and ICPC-2 or Q-Codes rubrics by their respective codes (see fig. 8). The following URIs are giving access to the hierarchies and rubrics of the corresponding classification. Note that each entry give access to a detailed terminological description, mappings to other terminologies and to automatic queries on resources like PubMed.

- URIs to reach the hierarchy of ICPC and Q-Codes
  - ICPC-2 `http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBO&tab=1`
  - ICPC-2 Process `http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBOPROC& tab=1`
  - Q-Codes `http://www.hetop.eu/hetop/Q?la=en&rr=CGP_CO_Q&tab=1`
- URIs to reach each rubrics of ICPC and Q-Codes
  - ICPC RFE and diagnosis: `http://www.hetop.org/hetop/?la=en&rr=CIP_D_ A01`

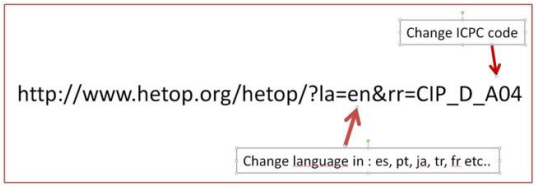
- ICPC Process: http://www.hetop.org/hetop/?la=en&rr=CIP_P_30
- Q-Codes: http://www.hetop.eu/hetop/Q?la=en&rr=CGP_QC_QC1

- To change the language; change the ISO 639 for the language; Ex.: =en for =pt for Portuguese (en,fr,es,pt,tr,vi,ko,nl, ge allowed - more in progress)

- To change the rubric; change the code at the end. Examples :

- ICPC process code #33 in English:
http://www.hetop.org/hetop/?la=en&rr=CIP_P_33
- ICPC-2 S chapter in Japanese:
http://www.hetop.org/hetop/?la=ja&rr=CIP_C_S&tab=1
- Q-Code QC Patient category in English:
http://www.hetop.eu/hetop/Q?la=en&rr=CGP_QC_QC
- Q-Code QD323 Shared decision making in Spanish:
http://www.hetop.eu/hetop/Q?la=es&rr=CGP_QC_QD323.

Figure 8. The URI for the code ICPC-2 A04 (Tiredness) in English



RESULTS

The tools developed to carry out this research are presented in the methods portion of this paper but can also be considered results. The provision of ICPC-2 and the Q-Codes in the form of Unique Resource Identifiers (URIs) was completed by a support web site (http://3cgp.docpatient.net/) and a Q-Code working group (https://tinyurl.com/Q-codesWG). All are technical, communicational or human realizations aimed at the achievement of this endeavor. 3CGP has been designed to be used by both humans and by machine.

3.7 3CGP use by humans

3.7.1 Pedagogical use

The ICPC is used worldwide as the main data producing system in Primary Care. It is incorporated into Health Information Systems and used in Electronic Medical Records in numerous countries. Availability of ICPC-2 in multilingual URIs is a must for teaching ICPC worldwide.

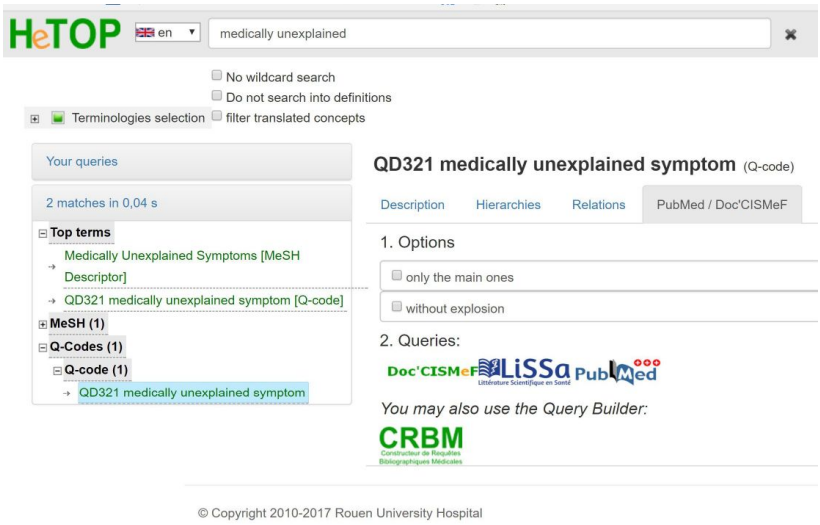
The eight domains addressed by the Q-codes are the embryo of what could become the table of contents of GP/FM. Teaching GP/FM is a must when referring to a rarely taught although so frequent as *Medically unexplained symptoms* or *Indoor pollution*.

The terms and definitions of the 182 Q-Codes are available in multiple languages, stressing the international interest surrounding this database. The terms and definitions have been edited in book format in 6 languages (es, pt, fr, en, nl, vi). All versions minus Vietnamese are available at the printing office (https://www.publier-un-livre.com/en/). All terminologies are available online on http://3cgp.docpatient.net/.

3.7.2 Bibliographic use

GP/FM has no specific indexing system. Twenty per cent of the ICPC-2 codes and all the Q-Codes are mapped automatically to MeSH and each mapping curated manually to MeSH of the National Library of Medicine. Q-Codes are a wonderful tool for teaching specific fields of GP/FM. They are also a useful resource of knowledge for students, researchers and working practitioners at the point of care. Automatic specific citations retrieval system allows access to dedicated bibliography on PubMed but also to LiSSa, the French resources base in medicine (see Fig. 9).

Figure 9. The HeTOP query interface for the Q-Codes *Medically Unexplained Symptom* proposes automatic queries to PubMed and LiSSa bibliographic bases.



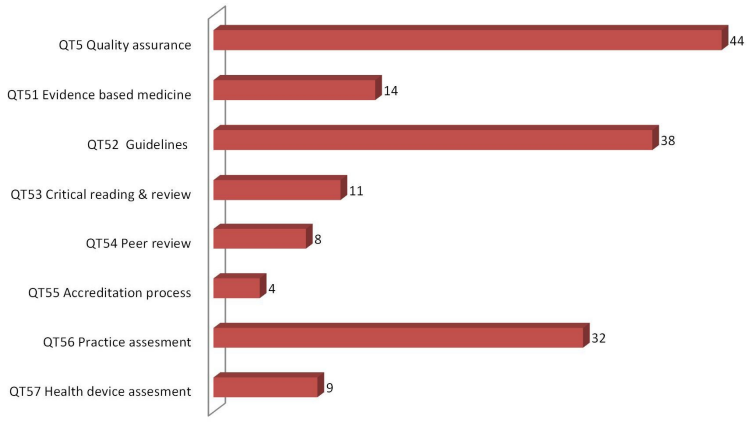
3.7.3 *Indexing master theses*

The figure obtained by manual indexing of conferences to develop the Q-Codes taxonomy are still a source of information on the interest of participating GPs. In Fig.10 we show the distribution of the rubrics of the QT domain in a French congress of GPs.

In this domain, several experiments are ongoing. The work of doctors in general and family medicine training is often of high quality, requires considerable investment from the authors and sometimes represents little-explored research areas. The leaders of the association of the three Belgian francophone universities (CCFFMG) (see `http://www.mgtfe.be/le-guide-dindexation-dun-tfe/`) decided to give visibility to this work by publishing the best of them online. The 3CGP indexing system was chosen to index work by authors At the time of registration. This guide, also available online

in English (`https://tinyurl.com/Q-Codes-guide`), has been incorporated into the online instructions for authors.

Figure 10. Distribution of QT (Training & Teaching) codes in a congress of French GPs showing the main domains of Interest of participating doctors (number of codes used on 212 abstracts)(Q-Codes version 2.3).



3.7.4 *Indexing of congresses*

The system developed to index master's theses is reused by the Brazilian Society of Family and Community Medicine (SBMFC). The 3CGP coding system is in use at the deposit page of their 14th Congress. The participants have to choose at least two codes and a maximum of 4 codes of ICPC-2 and Q-Codes. This experiment is currently underway

(see `http://3cgp.docpatient.net/codificacao-do-congresso-sbmfc`).

So far 1,746 reviewed and coded abstracts with ICPC and Q-Code have been retrieved and will be analyzed.

3.7.5 Indexing question-answer pairs

This approach is now used by researchers on data from Pernambuco, Brazil, for initially manually indexing a sample of 550 questions; with an ultimate goal of semi-automated indexing of larger data sets, measured in the tens of thousands of question-answer pairs. These question-answer pairs originate from the Brazilian Telehealth system, representing communication between rural health care providers and nurses and doctors in the urban Telehealth centers. (Resnick et al. 2013)

3.8 3CGP use by machines

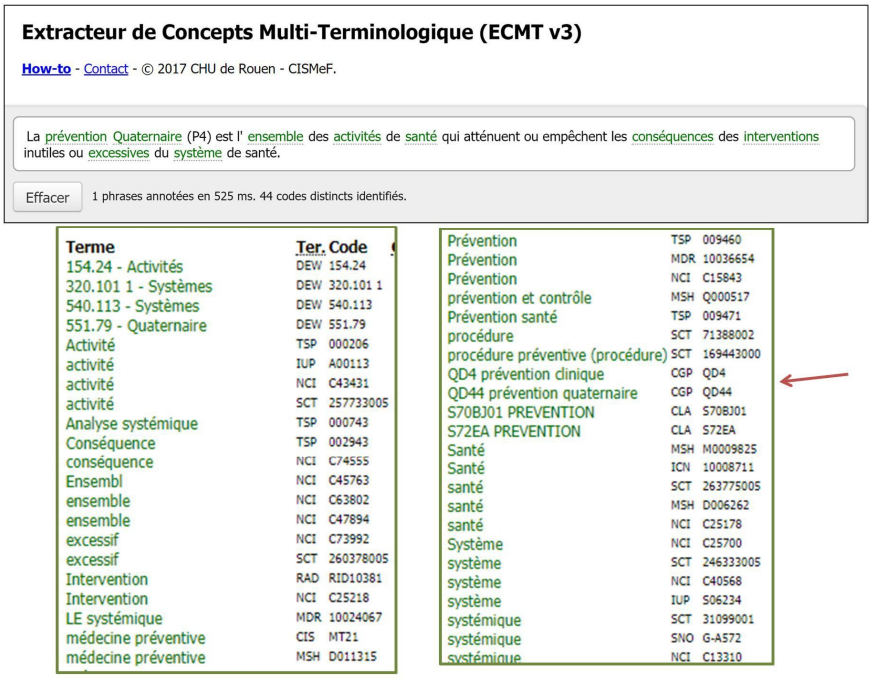
3.8.1 Automated classifier

To find an automated method capable of analyzing the content of non-clinical General Practice articles and predicting the corresponding Q-Codes categories is not an easy task. A classifier was already developed at the department of Information Systems at the University of Liège (HEC, Professor Ashwin Ittoo). The main difficulties arose from the small amount of sample data available, the large number of categories to be identified, and the high specificity of the scope of Q-Codes making categories difficult to discern (Rigaux, 2015). The classifiers also use filtered lemmatization, and they obtain a modest F1-score of 0.452 and 0.344 respectively. The full work is available in French on; `https://tinyurl.com/yc5ej2bw`.

3.8.2 Automated annotator

The tool *Extracting Concepts with Multiple Terminologies* (ECMT) is a web service developed at D2IM, Rouen. It aims to fully, automatically identify clinically relevant entities in medical texts in French with several types of documents: abstracts titles, documents about marketed drugs and death certificates (Cabot, 2016). The extraction is performed at the phrase level of the text. ECMT has also a user-friendly interface accessible after authentication (`http://ecmt.chu-rouen.fr/`).

Figure 11. Automatic annotation of concepts by ECMT v3 by MeSH (MSH), National Cancer Institute (NCI), MedDRA (MDR), SNOMED (SNO) etc. The red arrow shows the automated identification of concepts in Q-Codes (CGP); QD4 Prevention and QD44 Quaternary prevention (in French).



The pertinent terms are retained based on the HeTOP resources. In Fig. 11 an example is given for the processing of the phrase: *La prévention Quaternaire (P4) est l'ensemble des activités de santé qui atténuent ou empêchent les conséquences des interventions inutiles ou excessives du système de santé. [Quaternary Prevention (P4) is the set of health activities that mitigate or prevent the consequences of unnecessary or excessive health system interventions.]* ECMT extracts the terms in French from terminologies in HeTOP like MeSH or the National Cancer Institute terminology (NCI) as well as from Q-Codes terminology (CGP). As observable in Fig.11, QD4

Clinical Prevention and QD44 Quaternary Prevention have been identified.

3.8.3 Using Q-Codes in an e-learning program, Vietnam

Dr. Thành Liêm Võ, from the family medicine unit of the Pham Ngoc Thach Medical University, Ho Chi Minh City, Vietnam (`http://www.pnt.edu.vn/vi/`) has incorporated the Q-codes in Vietnamese in an e-learning system for medical students. The glossary of Family Medicine terminology helps one to understand and standardize the complex concepts of the discipline. It reduces the variety of these interpretations. Vietnamese versions of Q codes are used as a source of reference. For now, Q-Codes has been integrated into the format of glossary in 3 months FM training at Pham Ngoc Thach Medical University.

4 DISCUSSION

4.1 Main findings

Three areas of knowledge are at stake in this study: (i) Family Medicine as a pillar of primary care, (ii) Computational linguistics, and (iii) Information systems. The association of ICPC, in its three components Symptoms, Procedures and Diagnostics, with the Q-Codes forms an indexing system. This system therefore covers clinical and contextual elements specific to General Practice and Family Medicine. This system allows us to identify patients' symptoms and complaints, diagnosis or disease hypotheses, processes used by physicians, either by themselves or by third parties, and, finally, the context of application given by Q-Codes.

The Q-Codes represent a form of controlled medical, multipurpose vocabulary that is subject to further additions. As stated by Cimino *the unit of symbolic processing is the concept - an embodiment of a particular meaning*. Q-Codes can be seen as a medical subject authority list, including medical subject headings, a comprehensive series of mutually exclusive terms. According to guidelines set by Cimino, we have tried to gather a set of non-redundant, shareable, multipurpose, high-quality permanent concepts, in a mono-hierarchic organization, identified by a set of definitions and linked to existing terminologies. This study proposes a system of Knowledge Management (KM) in GP/FM which could potentially fill a major gap in KM of GP/FM. Conceived as a lightweight, multilingual ontology that is fit for new Internet technologies, NLP, and Semantic Web, 3CGP gives the opportunity to unravel GP/FM productivity and establish GP/FM as a professional discipline aiming at an extended range of specific knowledge.

4.1.1 Filling in a major gap in GP/FM and PHC

To the best of our knowledge, there is nothing similar available in GP/FM that has been developed for both human and machine use. There is also not anything of this measure that demonstrates the complexity of GP/FM. Due to the overlap GP/FM with the first line of health service, this tool could also be useful in Primary Care. All doctors and health managers, for whom proximity and health management are of utmost importance, could potentially reuse Q-Codes for their clinical needs, for teaching and for indexing..

4.1.2 Paving the way for an ontology in GP/FM and PHC

Though this project took years of work, it acts only as a base from which future researchers may expand upon. As it was designed according to termino/ontological concepts, is available in OWL and is ready for use with Linked Data. The set of ICPC-2 and Q-Codes is a lightweight ontology; however, because it adapts NLP and automatic and semiautomatic coding (Cabot et al., 2017b), it could serve as the basis for the development of a real ontology in GP/FM. The path to a real ontology is still a long time in the making.

4.1.3 Opening the gates for multilingualism in GP/FM and PHC

English has always been the fall-back language of GP/FM. However, family doctors speak to their patients and with one another in their own language, which leads to confusion in translation and varied context of vocabularies. This study has given a potential solution to this issue, by allowing for ICPC-2 to be published in 20 languages and Q-Codes in 8 languages. Having tools that facilitate various languages, while simultaneously communicating the same concept without variation in context or understanding, is incredibly important and useful/necessary for GPs. Having a tool that accommodates so many different mother-tongues may explain the enthusiasm of so many international colleagues that wished to participate in this multilingual edition.

4.2 Study limitations

4.2.1 A Single-Researcher Study

An important issue to address is that there was a seven-year hiatus in this research, shown by the dates of the conference abstracts analyzed. This was due to an extended illness by the main author. Despite this hiatus,

research was eventually able to move forward. Any negative effects resulting from this hiatus may be offset by the fact that only one researcher analyzed the abstracts. Bradley, Curry, and Devers (2007), qualitative data analysis experts, argue that *a single researcher conducting all the coding is both sufficient and preferred.[...]. In such cases, the researcher is the instrument; data collection and analysis are so intertwined that they should be integrated in a single person who is the choreographer of his/her own dance[...]* However, bias of said researcher could have influence over the collection of data and its analysis. Therefore, disclosure of the researcher's biases and philosophical approaches is essential. In this case, the main researcher is a male of Occidental origin who has been practicing as family doctor for more than 40 years with an expertise in Public health and taxonomy. An evaluation for the appropriateness of the selected terms could be in as future work the identification of the terms in a big sample of documents using semi-automatic term extraction or key phrases tools, to see the coverage with respect to the one selected by the one researcher. Of course a term extraction process needs in any case a further clinical review by one or more (better) domain experts.

4.2.2 An empirical move

The Q-Codes form the initial building blocks of classification in the GP/FM field. However, this approach has been filled with the personal experience of the main researcher, which may lead to unintentional biases. One can argue that the qualitative approach to the coding process is both inductive and deductive, an approach sometimes called abductive (Silver and Lewins, 2014). As an empirical document, one has tried to change, fill in the gaps and modify content of classifications using GP/FM publications, pair experience, critiques, and application to real work. MeSH's corresponding descriptors, searching, and indexation exercises on published documents have been also a good way to verify the applicability of the classifications. For safety, we've chosen to distribute the concept over all the classifications when adequate, rather than creating a special category. Each conference affirmed this thought and allowed for the addition of new elements. The fact that new concepts have emerged within Q-Codes has two reasons. First of all, the issue was addressed several times in conference abstracts. Secondly, the expertise in the field confirmed that the discussed issue was important.

4.2.3 Potentially Eurocentric

Another limitation of this study is that the data is Eurocentric. This is due to the fact that the conference abstracts analyzed present work done mainly by European GPs. Thus, the Q-Codes concepts might not be fully representative of other geographical areas- i.e., North America, South America, and Asia. This, in turn, may limit worldwide usability. Nevertheless, the fact that the Q-Codes have been translated into three non-European languages (Turkish, Vietnamese and Korean) implies that the translators have found points of connection to their own culture in the proposed concepts. However, this illustrates that the globalization of GP/FM concepts are strongly influenced by its Occidental, Anglo-Saxon origins (Simon, 2009; Gutierrez and Scheid, 2002).

4.2.4 Validity and reliability

Another potential issue is the validity of identification and concepts generated. Validity is concerned with *whether a variable measures what it is supposed to measure* (Bollen 1984 cited by Adcock and Collier, 2001). Here, we deal with the identification of concepts in texts. Yet, how can we measure that the same text will generate the same concepts accurately? Adcock and Collier (2001) also state: *Because background concepts routinely include a variety of meanings, the formation of systematized concepts often involves choosing among them*. They distinguish between a *consensual concept* and a *contested concept*. It is supposed that a text about *gender violence* will be identified with the corresponding concept *gender violence* by a reader. But, for more ambiguous terms, like *continuity* which is often confused with *permanence*, or more contestable concepts like *disease mongering* and *deprescription* which some colleagues may have no knowledge of, how does one proceed? This ambiguity may pose issue in the execution of this project.

There are as many definitions of *validity* in qualitative research as there are authors. *Face validity*, in quantitative research, is defined as *the extent to which a test is subjectively viewed as covering the concept it purports to measure.*(Holden, 2010). Noble and Smith (2015) propose a new terminology and criterion to evaluate the credibility of research findings. Usual terms used in quantitative research such as *validity*, *reliability* or *generalisability* are replaced with *Truth value, Consistency* and *Applicability.* Evaluating *Truth Value - Face Validity - Descriptive Validity* is recognizing that the interpretation bias, the particular way in which researcher view reality, corresponds to the reality in his/her colleague's world of reference. Many participants offered to translate and contribute to the development of the tool. This made a good argument in favor of the *Truth Value* of these findings. On the other hand, we have seen that the tool could be applied to very different situations in different countries in different languages. These two last points can bear witness to good *Truth Value* but also to good *Applicability*.

Evaluating *Consistency - Reliability - Interpretive Validity* is referring to whether these Q-Codes could be tested. It was imperative that the Q-Codes could be evaluated through extensive use GP/FM grey literature

indexation before being considered a valid construct. One measure used for testing was inter-indexer reliability. But, according to Funk and Reid (1983), who have studied the PubMed data-base for consistency in indexing, the quality of indexing cannot be directly measured, as there is no right or wrong way to index an article or abstract. In turn, the issue of holding abstracts to ambiguous standards of correctness is a potential downfall.

4.2.5 A searcher bias in need of discussion

We are not proposing a standard; however, we are proposing a searcher bias in need of discussion. The main aim of this research is to facilitate the management of information produced by family doctors and to prepare it for further computerized development/reuse. The current version of this program is named Q-Codes, in honor of its creator Professor Lamberts, but it is still only a preliminary version (ver. 2.5). It will obviously evolve, and names will most likely change. But the need to manage GP/FM information in a structured and standardized way must remain a substantial facet of research. The future of the profession is at stake.

It is important to recognize that Q-Codes have been created from a limited number of abstracts. If a concept was not present in the read abstracts, it will have no place in the Q-codes. This emphasizes that the current program is limited to a small number of abstracts within the GP/FM field. Q-Codes would need to integrate much more information to be considered a fully applicable program to the field of GP/FM. Further conferences will contribute new concepts to this, while simultaneously helping GP/FM to evolve. We hope that the structure of this proposed taxonomy will remain enough strong to support the introduction of new items, but it must be taken into account that as more information is added, the basis could potentially not be strong enough to accommodate all.

One issue with the Q-codes ontology involves the unique identifiers. Cimino (1998) notes that when building an ontology, there is an *irresistible temptation to make the unique identifier a hierarchical code which reflects the concept's position in the hierarchy.* However, there are inherent disadvantages to using unique identifiers. The first issue, which we have encountered here, is that the coding system runs out of room to grow (Cimino, 1996; Cimino, 1998). This can be due to limited depth, limited breadth, or both of the unique identifier. For instance, when the code has a limited number of positions (digits), the depth of the hierarchy is limited.

Further ontological research is needed to determine whether the two main rules of taxonomic thinking have been respected: completeness (all identified) and exclusivity (a place for each concept) (Ittoo and Bouma, 2013).

4.2.6 Advantages and limits of The Semantic Web

The Q-Codes bases, like all the terminologies edited on HeTOP server, are fit for The Semantic Web. Semantic Web technologies promote common data formats and exchange protocols on the Web, like the Resource Description Framework (RDF), the cited OWL (now available OWL-2) and the query language SPARQL. We have seen that Linked Open Vocabulary (LOV) (`http://lov.okfn.org/ dataset/lov/`), a lightweight ontology, differentiates from other ontologies through its characteristics that enable reuse and integration of other vocabularies-i.e., (i) small size, (ii) low formal constraints,
(ii) few instances except for examples, (iii) rich user documentation. Labels, comments, definition, description, etc. are all characteristics of Q-Codes and ICPC-2 on the HeTOP server. We hope that in the near future, ICPC-2 and Q-Codes could find their place in Linked Open Vocabularies, that are published and used by actors in diverse media corporations like BBC, national administrations like INSEE, the European Community, universities and research projects. Or perhaps, we hope to see them reused and published by individuals and put on the community table. (Vandenbussche and Vatant, 2011).

Nevertheless semantic difficulties may arise from the supposed simplicity of the language. The relation *is a* is not an simple relation. Aristotelian logic, which decomposes the proposition into subject and predicate (Younes, 2016), is not sufficient in rendering reality. Following Wittgenstein, the relation *is a* has at least three semantic interpretations. As stated by Wittgenstein (1922) in Tractatus Logico Philospophicus (TLP 3.323): *In the language of everyday life it very often happens that the same word signifies in two different ways – and therefore belongs to two different symbols – or that two words, which signify in different ways, are apparently applied in the same way in the proposition. Thus the word "is" appears as the copula, as the sign of equality, and as the expression of existence [. . . . .] In the proposition "Green is green" – where the first word is a proper name as the last an adjective – these words have not merely different meanings but they are different symbols.* Language could be more complex than its use in Health Information systems. As stated by Elish and Boyd (2017); *Because computational systems require precise definitions and mathematically sound logics, sociocultural phenomena that are typically nuanced and fuzzy are rendered in coarse ways when implemented into code*. Again, the last word will be given to Wittgenstein (TLP 4.002): *Language disguises the thought. So that from the external form of the clothes one cannot infer the form of the thought.*

5 CONCLUSION

Constructed on the basis of Semantic web technologies, Q-Codes could be considered as a lightweight ontology ready to be used in the semantic web domain, to be extracted in OWL. The multilingual classes of the classification could be individually reached through Unique Resource Identifiers (URIs). Note that each entry gives access to a detailed terminological description, mappings to other terminologies like Babelnet and DBpedia and to automatic queries on resources like PubMed.

We have created a terminology that highlights the vastness of GP/FM contributions to medical knowledge. We hope, by doing this, to contribute to the recognition of GP/FM as a professional entity within the scientific community that contributes heavily to all fields of medicine. Several question remain unsolved. Does the current extent of the knowledge base efficiently cover the GP/FM domain? How will this resist the hierarchical structure proposed to the introduction of new themes? Will this system retain enough inter-observers reliability? Nevertheless, given the number of contributions by volunteer translators, such an indexing system seems largely expected by the profession. We hope to transform it into a validated tool for its development.

REFERENCES

AAFP (2011). *Primary Care definitions – AAFP Policies*. URL: http://www.aafp.org/about/ policies/all/primary-care.html(visited on 02/07/2016).

Adcock, Robert and David Collier (2001). "A Shared Standard for Qualitative and Quantitative Re- search". In: *American Political Science Review* 95.3, pp. 529–546. ISSN: 00030554. DOI: 10.1017/ S0003055401003100. arXiv: arXiv:1011.1669v3.

Allen, Bradley P (2016). "The role of metadata in the second machine age". In: *Second International Conference Establishment Surveys.* ISBN: 9788578110796. URL: http://www.amstat.org/ meetings/ices/2000/proceedings/S57.pdf.

Allen, Justin et al. (2011). "The European Definition Of General Practice / Family Medicine". In: URL: http://www.woncaeurope.org.

Angell, Marcia (2017). "Drug Companies & Doctors: A Story of Corruption". In:

Bachman, CW (1969). "Data structure diagrams". In: *SIGMIS Newsletter* 1.2, pp. 4–10. DOI: 10.1145/ 1017466.1017467. URL: http://dl.acm.org/citation.cfm?id=1017467.

Bartholomeeusen, Stefaan, Frank Buntinx, and Jan Heyrman (2002). "Ziekten in de huisartspraktijk: methode en eerste resultaten van het Intego-netwerk". In: *Tijdschrift voor Geneeskunde* 58.863-871.

Bentzen N.(ed) (2003). *WONCA Dictionary of general/family practice*. Ed. by Niels Bentzen. Maanedsskr. Copenhagen. URL: http://www.ph3c.org/PH3C/docs/27/000092/0000052.pdf.

Berners-Lee, T, R Fielding, and Larry Masinter (1998). "Uniform Resource Identifiers (URI): Generic Syntax". In: *Request For Comments* 2396.

Bowker, GC and SL Star (1999). *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Bradley, Elizabeth H, Leslie A Curry, and Kelly J Devers (2007). "Qualitative data analysis for health services research: developing taxonomy, themes, and theory." In: *Health services research* 4, pp. 1758– 72. ISSN: 0017-9124. DOI: 10.1111/j.1475-6773.2006.00684.x.

Britt, H et al. (2003). "Bettering the Evaluation And Care of Health 2001-2002 (summary of results)". In: *Australian Family Physician* 32.1/2, pp. 59–63. URL: http://www.racgp.org.au/ document.asp?id=8998.

Britt, Helena et al. (2016). *A decade of Australian general practice activity 2006–07 to 2015–16. Bettering the Evaluation and Care of Health (BEACH)*. Report. Family Medicine Research Centre. Sydney School of Public Health. University of Sydney. URL: https://ses.library.usyd.edu.au/ bitstream/2123/15482/5/9781743325162{_}ONLINE.pdf.

Buono, Nicola et al. (2013). "40 years of biannual family medicine research meetings – The European General Practice Research Network (EGPRN)". EN. In: *Scandinavian Journal of Primary Health Care*. URL: http://www.tandfonline.com/doi/full/10.3109/02813432.2013.847594.

Bushman, B., D. Anderson, and G. Fu (2015). "Transforming the Medical Subject Headings into Linked Data: Creating the Authorized Version of MeSH in RDF". In: *J Libr Metadata* 15.3-4, pp. 157–176. ISSN: 1938-6389 (Print)1937-5034. DOI: 10.1080/19386389.2015.1099967.

Cabot, Chloé et al. (2017a). "Evaluation of the Terminology Coverage in the French Corpus LiSSa." In: *Studies in health technology and informatics* 235, pp. 126–130.

Cardillo, Elena (2015). "Mapping between international medical terminologies to SHN Work Package 3". In: *SemanticHealthNet*. Chap. Deliverable 3.3, 18p.

Carey, Iain M et al. (2004). "Developing a large electronic primary care database (Doctors' Independent Network) for research". In: *International Journal of Medical Informatics* 5, pp. 443–453. ISSN: 1386-5056. DOI: 10.1016/j.ijmedinf.2004.02.002.

Casado Vicente, Verónica (2012). *Tratado de medicina de familia y comunitaria*. Ed. by Verónica Casado Vicente. Médica Panamericana, p. 2563. ISBN: 8498355850.

Cavadas, L F, T Villanueva, and J Gervas (2010). "General practice innovation: a Portuguese virtual conference". In: *Med Educ* 44.5, pp. 514–515. ISSN: 0308-0110. DOI: 10.1111/j.1365- 2923.2010.03649.x.

Charlton, Rachel A et al. (2010). "Identifying major congenital malformations in the UK General Practice Research Database (GPRD)". In: *Drug Safety: An International Journal of Medical Toxicology and Drug Experience* 9, pp. 741–750. ISSN: 0114-5916. DOI: 10.2165/11536820-000000000- 00000.

Chinitz, David P and Victor G Rodwin (2014). "Perspective On Health Policy and Management (HPAM): mind the theory-policy-practice gap". In: *International Journal of Health Policy Management* 3.x, pp. 1–3. DOI: 10.15171/ijhpm.2014.122.

Cimino, J J (1996). "Review paper: coding systems in health care." In: *Methods of information in medicine* 35.4-5, pp. 273–84. ISSN: 0026-1270. URL: http://www.ncbi.nlm.nih.gov/pubmed/ 9019091.

— (1998). "Desiderata for controlled medical vocabularies in the twenty-first century". In: *Methods of Information in Medicine* 37.4-5, pp. 394–

403.
Colliers, Annelies et al. (2016). "Improving Care And Research Electronic Data Trust Antwerp (iCARE- data): a research database of linked data on out-of-hours primary care". In: *BMC Research Notes* 9.1, p. 259. ISSN: 1756-0500. DOI: 10.1186/s13104-016-2055-x.
David, A.K. et al. (2013). *Family Medicine: Principles and Practice*. Springer Science & Business,p. 1240. ISBN: 0387217444.
Davis, D. A. (2004). "CME and the pharmaceutical industry: two worlds, three views, four steps". In: *CMAJ*. Vol. 171, pp. 149–50. ISBN: 0820-3946 (Print)1488-2329 (Electronic). DOI: 10.1503/cmaj.1040361.
Dekkers, Makx (2009). "History, objectives and approaches of the Dublin Core Metadata Initiative". In: December, pp. 1–24.
Denzin, Norman and Yvonna S Lincoln (2000). "The Sage handbook of qualitative research (2nd ed.)" In: *Sage Publications*. Thousands Oaks: Sage., p. 784. ISBN: 9781412974172. DOI: Doi10.1177/ 1354067x07080505.
Dixon, Brian E, Atif Zafar, and Julie J McGowan (2007). "Development of a taxonomy for health information technology." In: *Studies in health technology and informatics* 129.Pt 1, pp. 616–620. ISSN: 0926-9630.
Dowden, John (2015). "Conflict of interest in medical journals". In: *Australian Prescriber* 38.1, pp. 2–3. DOI: 10.18773/austprescr.2015.001.
Druais, PL et al. (2009). *Médecine générale*. Ed. by D Pouchain. Masson SA, p. 460.
Eberbach, Andreas et al. (2016). "A simple heuristic for Internet-based evidence search in primary care: a randomized controlled trial". In: *Advances in Medical Education and Practice*, pp. 433–441. DOI: 10.2147/AMEP.S78385.
Elish, M. C. and Danah Boyd (2017). "Situating methods in the magic of Big Data and AI". In: *Communication Monographs*, pp. 1–24. ISSN: 0363-7751. DOI: 10.1080/03637751.2017.1375130.
Faggiolani, Chiara (2011). *Perceived Identity: applying Grounded Theory in Libraries*. it. DOI: 10.4403/jlis.it-4592.
Farace, Dominic John and Joachim Schöpfel. (2010). "Collection building with special regards to Report Literature" In: *Grey Literature in Library and Information Studies*. Ed. by Walter de Gruyter,
Ferreras Fernández, Tránsito (2016). "Visibilidad e impacto de la literatura gris científica en repositorios institucionales de acceso abierto. Estudio de caso bibliométrico del repositorio Gredos de la Universi- dad de Salamanca". PhD thesis. URL: https://gredos.usal.es/jspui/handle/10366/132444.
Friedman, Carol et al. (1999). "Representing Information in Patient Reports Using Natural Language Processing and the Extensible Markup Language". In: *Journal of the American Medical Informatics Association* 6.1, pp. 76–87. URL: http://www.jamia.org/cgi/content/abstract/6/1/76.
Funk, M E and C A Reid (1983). "Indexing consistency in MEDLINE." In: *Bulletin of the Medical Library Association* 71.2, pp. 176–83. ISSN: 0025-7338. URL: https://www.ncbi.nlm.nih. gov/pmc/articles/PMC227138/.
Gill, P. J. et al. (2014). "Development of a search filter for identifying studies completed in primary care". In: *Fam Pract* 31.6, pp. 739–45. ISSN: 0263-2136. DOI: 10.1093/fampra/cmu066. URL: http://dx.doi.org/10.1093/fampra/cmu066.
Glaser, Barney G. and Anselm L. Strauss (1999). *Discovery of Grounded Theory: Strategies for Qualitative Research - - Livres*. AldineTransaction, p. 284. ISBN: 0202302601.
Goggi, S. et al. (2015). "A semantic engine for grey literature retrieval in the oceanography domain". In: *Seventeenth International Conference on Grey Literature*. Amsterdam, pp. 76–77. URL: http:// greyguide.isti.cnr.it/wp-content/uploads/2017/04/GL17{_}Program{_}Book-min.pdf.
González-González, A I et al. (2007). "Information Needs and Information-Seeking Behavior of Primary Care Physicians". In: *Ann Fam Med* 4, pp. 345–352. ISSN: 1544-1709 (Print)1544-1717 (Electronic). DOI: 10.1370/afm.681.
Gotzsche, Peter G (2013). *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. Radcliffe Publishing Ltd.
Greenway, Tyler and Joseph S Ross (2017). "US drug marketing: how does promotion correspond with health value?" In: DOI: https://doi.org/10.1136/bmj.j1855.
GreyNet (2014). *Pisa Declaration on Policy Development for Grey Literature Resources*. URL: http: //greyguiderep.isti.cnr.it/Pisadeclapdf/Pisa- Declaration- May- 2014. pdf(visited on 06/16/2017).
Grosjean, Julien et al. (2012). "Teaching medicine with a terminology/ontology portal." eng. In: *Stud- ies in health technology and informatics* 180, pp. 949–53. URL: http://europepmc.org/ abstract/MED/22874333.
Gusso, Gustavo and José Mauro Ceratti Lopes (2012). *Tratado de Medicina de Família e Comunidade: 2 Volumes: Princípios, Formação e Prática*, p. 2180. ISBN: 8536327979.
Gutierrez, Cecilia and Peter Scheid (2002). "The History of Family Medicine and Its Impact in US Health Care Delivery". In: *AAFP Foundation*, pp. 1–31.
Gómez-Pérez, A., M. Fernández-López, and O. Corcho (2003). "Ontological Engineering and the Semantic Web". URL: http://www.exa.unicen.edu.ar/escuelapav/cursos/corcho/01_ introduction.pdf.
Heilman, J (2015). "Point of care Information in Open Access: A Time to Sow?" In: *PLOS Medicine* 12.8, e1001870. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001870.
Helman, Cecil G (2008). *Medical Anthropology*. Ashgate, p. 580. ISBN: 978-0-7546-2655-8.
Heyrman, J (ed.) (2005). *EURACT Educational Agenda, European Academy of Teachers in General Practice EURACT,* Leuven.
Hoffmann, Kathryn et al. (2011). "Antibiotic resistance in primary care in Austria - a systematic review of scientific and grey literature". In: *BMC Infectious Diseases* 11.1, p. 330. ISSN: 1471-2334. DOI: 10.1186/1471-2334-11-330.
Holden, Ronald B. (2010). "Face validity". In: *The Corsini encyclopedia of psychology*. Ed. by Irving B. Weiner and W. Edward. Craighead. Wiley, pp. 637–638. ISBN: 9780470170267.
Hong, Y. et al. (2016). "Knowledge structure and theme trends analysis on general practitioner research: A Co-word perspective". In: *BMC Fam Pract* 17, p. 10. ISSN: 1471-2296. DOI: 10.1186/s12875- 016-0403-5.
Hopewell, S et al. (2007). "Grey literature in meta-analyses of randomized trials of health care interven- tions." In: *Cochrane database of systematic reviews (Online)* 2, MR000010. ISSN: 1469-493X. DOI: 10.1002/14651858.MR000010.pub3.
Huang, Minlie, Aurélie Névéol, and Zhiyong Lu (2011). "Recommending MeSH terms for annotating biomedical articles." In: *Journal of the American Medical Informatics Association : JAMIA* 18.5, pp. 660–7. ISSN: 1527-974X. DOI: 10.1136/amiajnl-2010-000055.
Hubbard, Derek (2008). *How to Find Clinical Information Quickly at the Point of Care*. Vol. 15. 6. American Academy of Family Physicians, p. 23.
Hummers-pradier, Eva (2007). "Which Abstracts Do Get Published ? – Output Of German Gp Research 1999-2003". In: *Wonca Europe Paris 2007*.

Ittoo, Ashwin and Gosse Bouma (2013). “Term extraction from sparse, ungrammatical domain-speci c documents”. In: *Expert Systems with Applications* 40, pp. 2530–2540.

James, Jack E (2016). “Free-to-publish, free-to-read, or both? Cost, equality of access, and integrity in science publishing”. In: *Journal of the Association for Information Science and Technology* 68.6, pp. 1584–1589. ISSN: 2330-1643. DOI: 10.1002/asi.23757.

Jamoulle, M et al. (2014). “Mapping French terms in a Belgian guideline on heart failure to international classifications and nomenclatures: the devil is in the detail”. eng. In: *Inform Prim Care* 21.4, pp. 189– 198. DOI: 10.14236/jhi.v21i4.66.

Jamoulle, Marc (2015). “Quaternary prevention, an answer of family doctors to overmedicalization”. In: *International Journal of Health Policy and Management* 4.2, pp. 61–64. ISSN: 2322-5939. DOI: 10. 15171/ijhpm.2015.24. URL: http://ijhpm.com/article{_}2950{_}0.html.

Jamoulle, Marc, Julien Grosjean, and Stefan Darmoni (2017). “Access to multilingual individual rubrics in URI format for ICPC-2 and the Q-Codes”. In: URL: http://orbi.ulg.ac.be/handle/ 2268/211268.

Jamoulle, Marc et al. (2015). “Semantic Web and the Future of Health Care Data in Family Practice”. In: *Merit Research Journal of Medicine and Medical Sciences* 3.12, pp. 586–594. URL: http: //orbi.ulg.ac.be/handle/2268/189292.

Jamoulle, Marc et al. (2017a). “A terminology in General Practice / Family Medicine to represent non-clinical aspects for various usages : the Q-Codes”. In: *Medical Informatics Europe (MIE2017) Informatics for Health 2017 /* April, pp. 1–5. URL: http:// orbi. ulg. ac. be/ handle/ 2268/206527.

Jamoulle, Marc et al. (2017b). “Analysis of definitions of General Practice/Family Medicine and Primary Health Care”. In: *British Journal of General Practice - Open*, 050 ISSN: 0960-1643. URL: http://orbi.ulg.ac.be/handle/2268/210049.

Janamian, T et al. (2016). “Quality tools and resources to support organisational improvement integral to high-quality primary care: a systematic review of published and grey literature”. In: *Med J Aust* 204.7 Suppl, S22–8. ISSN: 0025-729x.

Jelercic, S. et al. (2010). “Assessment of publication output in the field of general practice and family medicine and by general practitioners and general practice institutions”. In: *Fam Pract* 27.5, pp. 582–9. DOI: 10.1093/fampra/cmq032.

Jonquet, Clement et al. (2016). “SIFR BioPortal : Un portail ouvert et générique d ’ ontologies et de terminologies biomédicales françaises au service de l ’ annotation sémantique”. In: *16e Journées Francophones d’Informatique Médicale(JFIM)* June.

Khan, Nada F, Sian E Harrison, and Peter W Rose (2010). “Validity of diagnostic coding within the General Practice Research Database: a systematic review”. In: 60.572, e128–e136. ISSN: 0960-1643. DOI:10.3399/bjgp10X483562.

Kochen, Michael M. (2012). *Allgemeinmedizin und Familienmedizin*. Thieme; Auflage: 4., vollständig überarbeitete und erweiterte Auflage, p. 652. ISBN: 3131413840.

Lakhani, Mayur K. (2003). *A Celebration of General Practice*. Radcliffe Publishing, p. 203. ISBN: 1857759230.

Lawrence, Amanda et al. (2014). “Where is the evidence: realising the value of grey literature for public policy and practice”. In: *Australian Policy Online*. DOI: 10.4225/50/5580B1E02DAF9. URL: http://apo.org.au/node/42299.

Lelong, R et al. (2016). “Semantic Search Engine to Query into Electronic Health Records with a Multiple- Layer Query Language”. In: *MEDIR workshop*. URL: http://medir2016.imag.fr/data/ MEDIR_2016_paper_8.pdf.

Liang, S. F. et al. (2014). “Semi Automated Transformation to OWL Formatted Files as an Approach to Data Integration”. In: *Methods of Information in Medicine* 54.1, pp. 32–40. ISSN: 0026-1270. DOI: 10.3414/ME13-02-0029.

Library and Archives Canada (2017). *Canadian Subject Headings*. URL: http://www.bac-lac.gc/.

Lin, Jennifer and Carly Strasser (2014). “Recommendations for the role of publishers in access to data.” In: *PLoS biology* 12.10, e1001975. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1001975.

Lowe, H J and G O Barnett (1994). “Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches.” In: *JAMA* 271.14, pp. 1103–8. URL: http://www.ncbi.nlm.nih.gov/pubmed/8151853.

Lusignan, Simon de (2005). “Codes, classifications, terminologies and nomenclatures: definition, develop- ment and application in practice”. In: *Informatics in primary care* 13.1, pp. 65–70. URL: http://www.ncbi.nlm.nih.gov/pubmed/15949178.

Madkour, Mohcine, Driss Benhaddou, and Cui Tao (2016). “Temporal data representation, normalization, extraction, and reasoning: A review from clinical domain”. In: *Computer Methods and Programs in Biomedicine*, pp. 52–68. DOI: 10.1016/j.cmpb.2016.02.007.

Mahood, Quenby, Dwayne VanEerd, and Emma Irvin (2014). “Searching for grey literature for systematic reviews: challenges and benefits”. In: *Research synthesis methods* 5.3, pp. 221–34. DOI: 10.1002/ jrsm.1106.

Marc, D. T. et al. (2015). “Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Term Coverage”. In: *Stud Health Technol Inform* 216, pp. 529–33. URL: https://www.ncbi.nlm.nih.gov/pubmed/26262107.

Martin, Patricia and Barry A. Turner (2016). “Grounded Theory and Organizational Research”. In: *http://dx.doi.org/10.1177/002188638602200207*. DOI: 10.1177_002188638602200207.

McGuinness, DL and Frank van Harmelen (2004). *Web Ontology Language*. URL: http://www.w3. org/TR/owl-features/.

McIntyre, Ellen et al. (2016). “The contribution of a knowledge exchange organisation in primary healthcare.” In: *Australian family physician* 45.9, pp. 684–7. ISSN: 0300-8495. URL: http://www. ncbi.nlm.nih.gov/pubmed/27606374.

McKenzie, Lindsay (2017). “Sci-Hub’s cache of pirated papers is so big, subscription journals are doomed, data analyst suggests”. In: *Science*. ISSN: 0036-8075. DOI: 10.1126/science.aan7164.

McWhinney, Ian R. (1997). *A Textbook of Family Medicine*. Oxford University Press, p. 448. ISBN:019511518X.

Mendis, Kumara and Indragit Solangaarachchi (2005). “PubMed perspective of family medicine research: Where does it stand?” In: *Family Practice* 22.5, pp. 570–575. ISSN: 02632136. DOI: 10.1093/ fampra/cmi085.

Miller, Eric (1998). “An Introduction to the Resource Description Framework”. In: *Bulletin of the Association for Information Science and Technology* 25.1, pp. 15–19. ISSN: 1550-8366. DOI: 10. 1002/bult.105. URL: http://onlinelibrary.wiley.com/doi/10.1002/bult.105/abstract.

Moher, David et al. (2000). “Mejora de la calidad de los informes de los metaanálisis de los ensayos clínicos controlados: el acuerdo QUOROM”. In: *Rev. Esp. Salud Publica* 74.2. URL: http://scielo.isciii.es/pdf/resp/v74n2/mejora.pdf.

Moynihan, R (2003). “Who pays for the pizza? Redefining the relationships between doctors and drug companies. 1: entanglement”. In: *Bmj* 7400, pp. 1189–1192. ISSN: 0959-535x. DOI: 10.1136/bmj. 326.7400.1189.

Moynihan, R and L Bero (2017). “Toward a Healthier Patient Voice: More Independence, Less In- dustry Funding”. In: *JAMA Intern Med* 177.3, pp. 350–351. ISSN: 2168-6106. DOI: 10.1001/ jamainternmed.2016.9179.

Murtagh, John (2011). *John Murtagh’s General Practice*. McGraw-Hill Medical Publishing Division, p. 1508. ISBN: 0070285381.

Myška, Matěj and Jaromír Šavelka (2013). “A Model Framework for Publishing Grey Literature in Open Access Defining Grey Literature”. In: *Open Access* 4.JIPITEC 2, para 104, pp. 1–12.

Neghme, A. (1975). “Operations of the Biblioteca Regional de Medicina (BIREME)”. In: *Bull Med Libr Assoc* 63.2, pp. 173–9. ISSN: 0025-7338 (Print).

Noble, Helen and Joanna Smith (2015). “Issues of validity and reliability in qualitative research”. In: *Evidence-Based Nursing* 18.2, pp. 34–35. ISSN: 1367-6539. DOI: 10.1136/eb-2015-102054. URL: http://ebn.bmj.com/cgi/doi/10.1136/eb-2015-102054.

PAHO Bireme Sao Paulo (2016). *Virtual Health Library Search Portal of the Regional library of Medicine*. URL: http://regional.bvsalud.org/http://www.paho.org/bireme/.

Pingitore, D and R A Sansone (1998). “Using DSM-IV primary care version: a guide to psychiatric diagnosis in primary care.” In: *American family physician* 58.6, pp. 1347–52. URL: http://www.ncbi.nlm.nih.gov/pubmed/9803199.

PLoS Medicine Editors (2015). “Point of care Information in Open Access: A Time to Sow?” In: *PLoS medicine* 12.8, e1001870. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001870.

Quan, Wei, Bikun Chen, and Fei Shu (2017). “Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016)”. In: *ArXiv e-prints*. arXiv: 1707.01162. URL: http://arxiv.org/abs/1707.01162.

Resnick, M. P., Santana, F., de Araujo Novaes, M., Shamenek, F. S., Frieden, L., & Iyengar, M. S. (2013). Representing second opinion requests from primary care within the Brazilian tele-health program: international classification of primary care, second edition. Studies in Health Technology and Informatics, 192, 1190. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23920964

Rigaux, Sébastien (2015). “Classification automatisée de résumés médicaux – Belgique 2015 [Multi-Label Text Classification of Medical Abstracts]”. Master thesis.

Salman Bin Naeem, Salman, Ahmed Shamshad, and Khan Amjid (2013). “Information seeking in primary care: a survey of doctors working in remote government health facilities in Pakistan”. In: *Library Philosophy and Practice (e-journal)* Paper 1009. URL: http://digitalcommons.unl.edu/ libphilprac/1009.

Schöpfel, Joachim (2015). “Littérature ≪ grise ≫ : de l’ombre à la lumière”. In: *I2D – Information, données & documents*. Vol. Volume 52. 1. Chap. Introduction, pp. 28–29. URL: https://www. cairn.info/revue-i2d-information-donnees-et-documents-2015-1-page- 28.htm.

Schrans, D. et al. (2016). “The search for person-related information in general practice: a qualitative study”. In: *Fam Pract* 33.1, pp. 95–9. ISSN: 0263-2136. DOI: 10.1093/fampra/cmv099.

Schuers, Matthieu et al. (2015). “Mise en ligne de la CISP en près de 20 langues au sein d’un portail terminologique de santé”. In: *Congrès de la Médecine Générale, Paris, 26 et 27 mars 2015*, poster. URL: http://orbi.ulg.ac.be/handle/2268/207421.

Schwitzer, Gary (2017). *Conflicts of interest in health care journalism. Who’s watching the watchdogs? We are. Part 1 of 3 - HealthNewsReview.org*. URL: https://www.healthnewsreview.org/ 2017/06/conflicts- of- interest- in- health- care- journalism- 1- of- 3/ (visited on 06/16/2017).

Shen, Cenyu and Bo-Christer Björk (2015). “‘Predatory’ open access: a longitudinal study of article volumes and market characteristics”. In: *BMC Medicine* 13.1, p. 230. ISSN: 1741-7015. DOI: 10. 1186/s12916-015-0469-2.

Shultz, M. (2007). “Comparing test searches in PubMed and Google Scholar”. In: *J Med Libr Assoc* 95.4, pp. 442–5. ISSN: 1536-5050 (Print)1558-9439 (Electronic). DOI: 10.3163/1536-5050.95.4. 442.

Silva, Caio da, Regina Garcia, and Rita Bonadio Inacio de Cássia (2009). “Literatura Cinzenta : teses , eventos e relatórios”. Thesis. URL: http://rabci.org.

Silver, Christina and Ann Lewins (2014). *Using Software in Qualitative Research*. SAGE Companion. URL: https://study.sagepub.com/using-software-in-qualitative-research.

Simon, C. (2009). “From generalism to specialty–a short history of General Practice”. In: *InnovAiT* 2.1, pp. 2–9. ISSN: 1755-7380. DOI: 10.1093/innovait/inn171.

Sladek, Ruth et al. (2006). “Development of a subject search filter to find information relevant to palliative care in the general medical literature.” In: *Journal of the Medical Library Association : JMLA* 94.4, pp. 394–401. URL: http://www.ncbi.nlm.nih.gov/pubmed/17082830.

Soler, Jean karl, Marc Jamoulle, and Peter Schattner (2015). “The International Classification of Primary Care”. en. In: *The World Book of Family Medicine – European Edition 2015*. Ljubljana. ISBN: 978-961-281-983-5. URL: http://orbi.ulg.ac.be/handle/2268/187050.

Soler, Jean K. et al. (2012). “The interpretation of the reasons for encounter ’cough’ and ’sadness’ in four international family medicine populations”. In: *Informatics in Primary Care* 20, pp. 25–39. ISSN: 14760320.

Sutton, Stuart A (2007). “Tutorial 1: Basic Semantics”. In: August. URL: http://dublincore. org/resources/training/dc-2007/T1-BasicSemantics.pdf.

Swan, Alma. (2012). *Policy guidelines for the development and promotion of open access*. United Nations Educational, Scientific, and Cultural Organization, p. 76. ISBN: 9789230010522.

Tabatabaei-Malazy, O, S Nedjat, and R Majdzadeh (2012). “Which information resources are used by general practitioners for updating knowledge regarding diabetes?” In: *Arch Iran Med* 4, pp. 223–227. ISSN: 1029-2977. DOI: 012154/aim.0010.

Tan, Sharon Swee-Lin and Nadee Goonawardene (2017). “Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review”. In: *Journal of Medical Internet Research* 19.1, e9. ISSN: 1438-8871. DOI:

10.2196/jmir.5729.

Thompson, C. A. et al. (2014). "Patient and provider characteristics associated with colorectal, breast, and cervical cancer screening among Asian Americans". In: *Cancer Epidemiol Biomarkers Prev*. Vol. 23. United States: (c)2014 American Association for Cancer Research., pp. 2208–17. ISBN: 1538-7755 (Electronic)1055-9965 (Linking). DOI: 10.1158/1055-9965.epi-14-0487.

Ustün, T B et al. (1995). "New classification for mental disorders with management guidelines for use in primary care: ICD-10 PHC chapter five." In: *The British journal of general practice : the journal of the Royal College of General Practitioners* 45.393, pp. 211–5. ISSN: 0960-1643. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239204/.

Van Royen, Paul et al. (2010). "Are presentations of abstracts at EGPRN meetings followed by publication?" EN. In: *The European journal of general practice* 16.2, pp. 100–5. ISSN: 1751-1402. DOI:10.3109/13814788.2010.482582.

Vandenbussche, Py and Bernard Vatant (2011). "Metadata recommendations for linked open data vocabu- laries". In: *Version*. URL: http://lov.okfn.org/dataset/lov/Recommendations{_}Vocabulary{_}Design.pdf.

VanNieuwenborg, L et al. (2016). "Continuing medical education for general practitioners: a prac- tice format". In: *Postgrad Med J* 92.1086, pp. 217–222. ISSN: 0032-5473. DOI: 10.1136/ postgradmedj-2015-133662.

Vanopstal, Klaar et al. (2011). "Vocabularies and retrieval tools in biomedicine: disentangling the termino- logical knot." In: *Journal of medical systems* 35.4, pp. 527–43. ISSN: 0148-5598. DOI: 10.1007/ s10916-009-9389-z.

Veuillotte, I et al. (2015). "General practice and the Internet revolution. Use of an Internet social network to communicate information on prevention in France". In: *Health Informatics J* 1, pp. 3–9. ISSN: 1460-4582. DOI: 10.1177/1460458213494905.

Wells, K (1995). "The strategy of grounded theory: possibilities and problems." In: *Social work research* 19.1, pp. 33–7. ISSN: 1070-5309. URL: http://www.ncbi.nlm.nih.gov/pubmed/ 10140997.

Wittgenstein, Ludwig (1922). *Tractatus logico-philosophicus*. London: Kegan Paul, p. 108. ISBN: 1602064512. URL: http://people.umass.edu/klement/tlp/.

WONCA (1987). *ICPC: International Classification of Primary Care*. Ed. by H Lamberts and M Wood. Oxford: Oxford University Press. ISBN: 0 19 261633 1.

WONCA (2005). *ICPC-2-R: International Classification of Primary Care (Oxford Medical Publications)*. Oxford University Press, USA, p. 204. ISBN: 019262802X.

Younes, Nora (2016). *Introduction à Wittgenstein*. Paris: Editions La Découverte, p.11