

# The Creation of a Standardized Citizen Health Vocabulary

## A Linked Open Data and Semantic Web example

Martijn Hartog  
eSociety Institute  
The Hague University of Applied Sciences  
The Hague, The Netherlands  
m.w.hartog@hhs.nl

Bert Mulder  
eSociety Institute  
The Hague University of Applied Sciences  
The Hague, The Netherlands  
a.w.mulder@hhs.nl

**Abstract**—The urgency behind this research is twofold. Firstly, the on-going development of an aging society could potentially lead to a 40% workforce reduction in the Dutch public sector, and with insufficient financing in the healthcare sector requires a participatory society in which citizens will need to start taking more responsibility regarding their healthcare. Secondly, we also live in an information society, which keeps developing itself. This brings new and innovative digital opportunities and information services, which potentially makes a participatory society viable. Yet, the development of such opportunities also brings new design requirements since systems world and life-world are converging. Based on an explorative study one of the requirements appeared the creation of clear conditions regarding the comprehensibility, usability and integration of such information technologies. More specifically, this would mean creating digitally accessible and usable terminologies in order for citizens and governmental bodies and professionals to work together for the quality of life. For healthcare this would mean creating so-called a citizen/consumer health vocabulary, a machine-readable standardized list of terms which links ‘lay-language’ and the standard terms of professional terminologies.

**Keywords**—participatory society; health; vocabularies; Linked Open Data; semantic web.

### I. INTRODUCTION

Citizen health vocabularies are created to try to overcome the communication gap between patients and professionals, as the daily life-world of citizens is vastly different from the systems-world of professionals. Professionals use technical terms whilst citizens use lay-language (their own “technical terms”). During face-to-face communication, it is often possible to ask for an explanation of the technical terms used by professionals. However, online this is usually not an option. Online information is formatted in a “one-size-fits-all” package, based on the knowledge and opinions presented. If the patient lacks the means of understanding of technical terms this could lead to faulty associations or misinterpretations [1].

Smith interprets the obstacle described above from two viewpoints, on the one hand there are patients who have difficulty understanding medical terminology and on the other hand, information systems designed specially to

recognize, understand and process medical technical terms, have difficulty interpreting lay-language, since patients are generally not involved in the creation of (medical) terminology [5], whilst Dutch citizens themselves indicate that they’d like to be more informed about and more in control of their own medical records [6].

The advancement of technology in recent years, like Open Data, Linked Data and the Semantic Web has allowed us to have access to any information, anywhere, at any given time and make connections with other information / data for more adequate information and analysis. Large (amounts of) databases, websites, digital books, et cetera are all searchable with only a few actions needed.

Most of our searches are performed by known search engines, yet we don’t always get the best results/results we are looking for. Health is an important issue and if necessary citizens want and need qualitative results. A search result may fall short partly because the query is too vague or because the wrong search terms were used. Almost everyone wants to find information that is complete, relevant and reliable in order to rightly participate in realizing the quality of life for themselves and their environment [7].

The digital communication between citizens and professionals has not been optimal up to now. Not professionally trained citizens may find difficulty in reading and understanding medical documents on the Internet and other medical information coming from healthcare professionals. At the same time, there are care-related computer programs, applications, designed to handle only medical terminologies, which have difficulty interpreting expressions of citizens and accompanied care intentions.

The aim of this research is to explore the possibilities of citizen vocabularies and a possible design of a Dutch citizen health vocabulary in combination with semantic web technologies. On the base of the approach were two main aspects. One was an explorative and inventory research of current International initiatives of consumer health vocabularies, among others: the American open access, collaborative (OAC) consumer/citizen health vocabulary (CHV) [1][2] and the Italian Consumer Medical Vocabulary (ICMV) [3][4]. The other was a

conceptual study enriching patient friendly terms linked to SNOMED terms with the National competence Centre for standardization and eHealth (Nictiz) and the Dutch Patients and Consumer Federation (NPCF).

In Section 3, we provide more context for the Dutch initiative and situation which created an urgency for this research. In Section 4, we reflect on the possibilities for a semantic citizen health vocabulary. In the last section we conclude with some discussion points we intend to continue upon in the research project.

## II. CONTEXT

Before we explicate the possibilities for a citizen health vocabulary it is of importance to elaborate on the specific Dutch context in which the research was developed. In the following paragraphs we will emphasize on the development and different design requirements of digital citizens and the main challenges if the ‘systems world’ and the ‘life world’ converge in order to realize and manage functionality and participation.

### A. *The Dutch context*

Close to 100% of the population is connected to the Internet and more than 50% of them have access to broadband with speeds that are the highest in European benchmark studies. All people up to 75 years of age use the Internet regularly, mostly daily. Adoption of technology is pervasive, with many intensive users of Twitter and LinkedIn, the most web shops and the most online therapies per capita worldwide.

At the same time the Dutch society is confronted with the transformation of the healthcare and welfare sectors, and moving towards a participatory society where citizens need to take care of themselves more than they have been used to. That transformation is partly due to financial reasons, but partly the consequence of demographics with the effects of an aging society (larger percentage of older people) with a smaller percentage of younger people [8]. The Dutch Ministry of Internal Affairs and Kingdom Relations predict that those compounded developments will cause a 40% reduction of employees in government by 2022 [9][10].

### B. *Digital citizens*

Citizens use digital means to take responsibility for the safety and quality of life in their environment. They may make reports digitally, develop ideas together digitally how to deal best with situations, decide using digital systems and organize the activities involved in maintaining quality and safety in their neighborhood or village. Their digital community systems are connected to all other stakeholders that might be involved in safety and quality of life – they may inform or alarm others to take action.

Citizens in neighborhoods, towns, cities and regions have access to actual and recent data about safety and quality of life of their environment, integrated, analyzed and presented in an understandable way. The use of ICT by citizens in their daily life is fundamentally different from that of professionals: that creates very specific requirements for a digital society. The structural use of digital solutions by citizens means the shift from the ‘systems world’ of professionals to the ‘life world’ of people in their day-to-day lives and environment.

That citizens will become a dominant factor in the next phase of the information society is the expected next step in its development. Technology has become personal, prices have come down, communication is effortless which results in the development of integrated information infrastructures *now centered around the citizen*. This may sound self-evident but it is a new development without prior history. We actually have very little notion of the specific requirements that citizens have in their personal lives, even though we have spent years optimizing the usability of individual applications. But that is no longer sufficient and we need to direct our attention to the broader infrastructure supporting a digital nation, where people digitally coordinate their lives, work and activities together in smaller and larger groups. This development is characterized by three new challenges that set it apart from professional information infrastructures [8][9].

### C. *Main challenges*

Quality – people speak their own language. In our own lives we speak our own language. We may suffer from a stroke but not call it CVA (‘Cerebro Vascular Accident’) or TIA (‘Transient Ischaemic Attack’). We talk about money and not ‘liquidity’. People use their own language, where professionals need more precise terminology to do their work. For digital solutions to work for citizens, they need to speak the language of people. For health providers in a country that will mean that they - in order to communicate consistently and interoperable with consumers - should use a standardized consumer/citizen health vocabulary in order to realize understandable and usable information [11].

Complexity – diversity requires integration. The growing diversity of digital solutions in the household creates technical challenges, data inconsistencies and information overload. Products and services may be designed to work effectively for a single user, but often not to interoperate with the products and services of other providers. Healthcare apps generally don’t integrate their data with other apps. Government services provide individual services, often not integrated or providing a comprehensive overview of all interactions with government.

Scale – networks require effective solutions. People live and grow in groups and networks. That is why

citizens need to be supported and empowered at a different scale than the individual. Digital solutions need to be effective at the level of groups, neighborhoods, towns, cities, regions and society. Their data need to be interchangeable and consistent and their services scalable and understandable at every level.

By misinformation, patients can for example: misjudge their symptoms, take their medication wrong and could be inadequate to individually maintain a certain level of quality of life for themselves and others. Citizens and health care providers (and other sectors) may experience inadequate digital communication as a problem. But the problem at the moment is mainly present in western countries where there is an aging population, an increase in the use of ICT and less capacities for face-to-face communication, with healthcare providers/professionals.

### III. HEALTH VOCABULARY AND SEMANTIC WEB: LESSONS LEARNED

A semantic citizen health vocabulary (CHV) could lead to better understanding of health related issues and more efficient communication of these health related issues to, with and between patients. Moreover, such a vocabulary could also help to enhance search engines and could form the basis for automatic knowledge and advice [8][9]. As Smith and Wicks mention “consumers have gained increased access online to the literature of healthcare professionals, they have also formed their own powerful communities of expertise, and so the very notion of “expertise” has undergone expansion” [12]. There are numerous examples of online platforms in which citizens are able to share questions and answers related to shared interests and illnesses, like ‘PatientsLikeMe’. A CHV extends further:

- Recording and standardizing citizen terminology.
- Connecting terminology with systems-world.
- Referring to extra information or explanations.
- Adding qualities.
- Better disclosure and optimizing search results.
- Realizing automatic processing by machines.

Much like creating thesauri or ontologies, CHVs include phases as connecting terminology, structuring accordingly using SKOS (Simple Knowledge Organization System) and RDF (Resource Description Framework) enabling standardized exchange of information and data. Afterwards, a CHV ontology structurally classifies categories, classes, subclasses, instances, individuals, properties creating ‘concepts’ that are machine-readable and interchangeable. In this way, knowledge domains can be mapped and qualities are defined, which in turn creates the possibilities to defining relations, characteristics and meaning of concepts/terminology [13] from a citizens’ point of view.

SNOWMED for instance contains a systemically organized computer processable collection of medical terms (codes, terms, synonyms and definitions) merely used in clinical documentation and reporting, and doesn’t match communicating or integrating in services towards citizens.

A possible function of a CHV is increasing the comprehensibility of medical information. An integrated CHV would increase the comprehensibility of for example electronic health records and personal health records, by linking patient-friendly terms with professional terms, as done with the International initiatives. Much like translation applications which are currently build upon CHV’s [2].

Another possibility/function of a CHV is supporting the patient by providing the right words for describing their medical history, editing and reading their personal health records and during online consultations. The implementation of a CHV in personal health records ensures that such records are easily understood by the patient. Which in turn gives the patient a better understanding of, and control over his/her own health. A personal health record is only useful for the patient, provided that the health information and associated tools are understandable and can form a base for action [4].

There are multiple ways to give shape to the concept of CHV. In essence, it’s about patient-friendly terms being “linked” to their professional counterparts. This means nothing more than an organized list of data in a machine-readable open file format. Such CHV’s are freely available and can be used by computer applications (e.g., search engines and translation applications). Once the patient-friendly terms are collected, this option would be relatively easy to develop. CHV’s open a world of possibilities, connecting them using semantic web technologies/linked open data. The added value of linked open data is the ability of machine-readability and usage combining other linked open datasets. It becomes possible to query extensively more data and adding uniform resource identifiers (URI’s), unique names, to entities will optimize search results and eventually enriches knowledge and analysis.

### IV. CONCLUSION AND DISCUSSION

The development of a CHV is based on research which contains the information needs of the citizens and connecting them digitally with professional information. The challenge is to study/research how people communicate with each other in their own circles/environment, to capture the most common used terms, and then connect these terms with professional standards. Additionally, patient-friendly terms have to be sufficiently consistent: “The feasibility of developing useful CHVs relies on the fundamental question about the stability of lay health language. That is, are the health-related forms, concepts, and relationships familiar to and

used by citizens sufficiently consistent to allow for the development of a useful vocabulary?" [1].

A Dutch CHV would ideally make more extensive use of semantic web technologies. Moreover, it should include more focus on qualitative applicability in order to generate patient-friendly terms, allowing to generate relations between these terms and with the professional terms. Based on the explorative study of International initiatives it is very important that citizens and patients are actively involved in the creation and attribution of terminology for a CHV [14].

As the resources for developing or translating professional terms are quite restricted an interactive process with a strong co-creation component could help intensify a truly patient-friendly consumer health vocabulary and will ensure the general applicability and sustainability of the vocabulary (Figure 1).

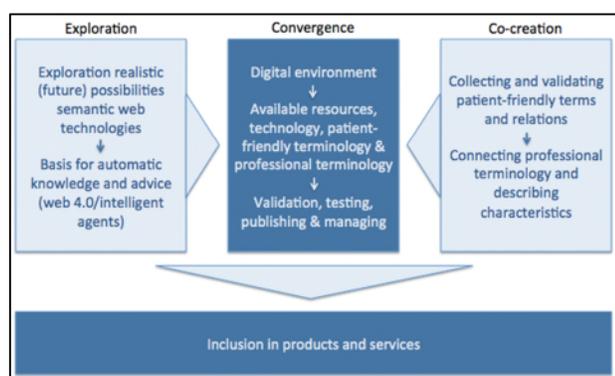


Figure 1. Process patient-friendly citizen health vocabulary

A healthcare vocabulary would give patients more control and possibilities for active participation with ensuring they are equipped and empowered for their quality of life. The design and development process of a digital environment for the collection, validation, linking of terms and relations can start directly. At the same time a further exploration of the realistic future possibilities of semantic web technologies is needed. More specifically, how a short term semantic CHV could be developed and enable automatic knowledge and advice (intelligent agents) in the near future [15]. After this is completed validation is required by domain professionals on the used technology, patient-friendly collection and standards. After a testing phase the CHV could be published and managed. Lastly, the integration in products and services will follow.

#### ACKNOWLEDGMENTS

This study on which the paper is build is part of an ongoing larger and broader exploration within the multiannual research and development programs 'Government of the Future' and specifically the National program 'Citizen Information Management', of the eSociety Institute of The Hague University with the

Municipality of The Hague, the Province of South-Holland and VNG/KING (Quality Institute Dutch Municipalities), focusing on the possibilities to structurally engage and equip a participatory society in Public Administration and healthcare/welfare.

#### REFERENCES

- [1] Q. T. Zeng and T. Tse, "Exploring and developing consumer health vocabularies" *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp. 24-29, 2006
- [2] Q. T. Zeng, S. Goryachev, H. Kim, A. Keselman and D. Rosendale, "Making texts in electronic health records comprehensible to consumers: a prototype translator" *Annual Symposium Proceedings, American Medical Informatics Association*, pp. 846, 2007
- [3] E. Cardillo, "A Lexi-ontological Resource for Consumer Healthcare: The Italian Consumer Medical Vocabulary" *Doctoral dissertation, University of Trento*, 2011
- [4] E. Cardillo, "Medical terminologies for patients", 2015
- [5] C. A. Smith, "Consumer language, patient language, and thesauri: a review of the literature" *Journal of the Medical Library Association*, vol. 99, no. 2, pp. 135, 2011
- [6] Nictiz and Nivel, "eHealth-monitor", 2015
- [7] P. Becker, "Organize your information: Working with thesauri, taxonomies, tags and topics (in Dutch: Organiseer je informatie: Aan de slag met thesauri, taxonomieën, tags en topics)", Leidschendam: Biblion, 2010
- [8] B. Mulder and M. Hartog, "Digital Citizens | Digital Nations: the next agenda", The Hague, 2015
- [9] Dutch Ministry of Internal Affairs and Kingdom Relations, "The Great Departure (in Dutch: De Grote Uittocht)", 2010
- [10] Dutch Ministry of Internal Affairs and Kingdom Relations, "The Great Exit revised (in Dutch: De Grote Uittocht herzien)", 2013
- [11] Nictiz and Nivel, "eHealth-monitor", 2016
- [12] C. A. Smith and P. J. Wicks, "PatientsLikeMe: Consumer health vocabulary as a folksonomy" *Annual Symposium Proceedings, American Medical Informatics Association*, pp. 682, 2008
- [13] N. F. Noy and D. L. McGuinness, "Ontology Development 101: A guide to creating your first ontology", 2008
- [14] N. Oudshoorn and A. Somers, "Constructing the digital patient: Patient organizations and the development of health websites" *Information, Community & Society*, vol. 9, no. 5, pp. 657-675, 2006
- [15] J. Fowler and E. Rodd, "The Ultra-Intelligent Electronic Agent is Coming", 2012