Evaluation of LOINC semantic coverage for coding clinical observations related to environmental influenced diseases

Maria Teresa Chiaravalloti
Instituto di Informatica e Telematica,
Consiglio Nazionale delle Ricerche,
Cosenza, Italy

Abstract

Terminology standards ensure semantic interoperability by allowing efficient exchange and pool of data coming from many idiosyncratic sources. This is especially true when operating among multiple domains that are connected to each other through different kinds of semantic relationships. Environment and health are strictly related domains, as the first one produces a variety of effects on the second one. While terminologies exist for coding and classifying almost all the aspects of the healthcare process (actors, products, observations, etc.), a specific one focused on environmental related or caused health problems has not yet been developed. Nonetheless, concepts related to the link between the two domains can be found in different standards. Among them, Logical Observation Identifiers Names and Codes (LOINC®) is the most widely used for identifying parameters connected to clinical conditions. Considering this, the work presented in this paper aims to: 1) identify a subset of existing LOINC codes of clinical observations influenced or related to environmental factors; 2) search for environmental concepts not covered by LOINC in the UMLS Metathesaurus to see if they are covered and how they are expressed in other medical terminologies; and consequently 3) detect LOINC concept gaps related to the link between environment and health to be filled by proposing new LOINC terms for standardizing identified clinical parameters.

1.0 Introduction

Widespread adoption of terminology standards offers the promise of enabling efficient processing and storage of data that comes from many independent sources (McDonald 1997). Only if data carrying the same semantic meaning are identified by the same code, even if they are labeled with idiosyncratic names, can be pooled and merged. Existing standard terminologies cover multiple domains and they are often interrelated through meta-thesauri, mappings, transcodings, etc., so they can interoperate according to different types of semantic relationships. Multiple relations exist be-

tween *health* and *environment* domains, covering more than one aspect of each. However, the influence of environmental factors on human health is proved by clinical studies of different medical specialties (Mekonnen 2021, Nakayama 2021 and Stringer 2021, just to cite the most recent ones), and it is has been recently confirmed by the WHO Global Air Quality Guidelines (AQGs) (World Health Organization 2021). Compared to their last edition of 2005, AQGs reduce levels of key air pollutants that need to be controlled to keep air quality levels within limits which are not dangerous for human health. The guidelines, in fact, provide clear evidence of the damage inflicted by air pollution on the health of the population, at even lower concentrations than previously recommended.

Both *health* and *environment* domains are almost completely semantically covered by terminologies for coding and classifying different aspects of those fields. Nonetheless, to the best of our knowledge, a specific resource focused on environmental related or caused health problems has not yet been developed. However, concepts expressing the relation between the two domains can be found in different terminological standards. Among them, this work is focused on LOINC, which is the most widely used standard for identifying clinical observations connected to health conditions (not necessarily already resulted in problems or diseases). Investigating clinical parameters that can be influenced by environmental determinants is fundamental to monitor the trend and evolution of possible risk situations and consequently plan corrective actions. For this reason, this work aims to identify a subset of existing LOINC codes of clinical observations influenced or related to environmental factors, finding possible concept gaps and consequently proposing new LOINC terms to standardize them. The choice of LOINC is also due to the fact that it is an open and nimble standard, compared to many others of the clinical domain, and this makes it easy to quickly update.

2.0 Background

LOINC (McDonald 2003) is a universal code system for laboratory and clinical observations used in more than 164 countries to enable semantic interoperability. It was created in 1994 by the Regenstrief Institute, Inc, which continues as the standards development organization (SDO) responsible for its development. The current release, version 2.71 (August 2021), contains more than 96,000 concepts covering the full scope of laboratory testing (hematology, microbiology, etc.) and a broad range of clinical measurements (e.g. vital signs, EKG, patient reported outcomes, etc.). Based on formal naming conventions, LOINC also carries names for document titles (Frazier 2001) radiology reports (Vreeman 2005) and section headings (social history, objective, etc.).

At present, LOINC is used by many kinds of organizations, including large reference laboratories, healthcare organizations, insurance companies, regional health information networks and national standard bodies. Global LOINC adoption has been accelerated by an innovative approach to facilitating translation (Vreeman 2012). The official LOINC Italian version has been translated and it is currently mantained by the Institute of Informatics and Telematics of the National Council of Research. It has been adopted as national standard for identifying document type in the national Electronic Health Record, a plethora of clinical parameters in Health Level 7 (HL7) Implementation Guides and, specifically, for coding performed tests in laboratory reports.

LOINC codes are uniquely assigned numbers and LOINC names are defined 'fully specified' because they contain all the information needed to unambiguously identify a test, distinguishing it from others who might apparently seem identical. These names are given by the concatenation of six fundamental axes:

1. Component: the substance that is measured (e.g., sodium, glucose, etc.);
2. Property: the measurement type;
3. Timing: it distinguishes measurements made at a given time by those covering a time interval;
4. System: the type of sample on which the observation is performed;
5. Scale: the scale of measurement;
6. Method: the method used in test performing.

Studies about the application of specific subsets of LOINC codes have been carried out in multiple domains, such as workers' health in the workplace (Park 2021), social determinants of health (Watkins 2020) and, last but not least, the Covid-19 (Dong 2020), but there are not a lot of works about the use of LOINC for coding clinical observations in some way related to environmental factors, despite the fact that it is the most widely used standard for identifying measurements and parameters that allow the definition or monitoring of health problems and conditions. It is worth mentioning the Pan-Canadian REspiratory STandards INitiative for Electronic Health Records (PRESTINE), whose

> "goal is to recommend respiratory data elements and standards for use in electronic medical records across Canada that meet the needs of providers, administrators, researchers and policy makers to facilitate evidence-based clinical care, monitoring, surveillance, benchmarking and policy development" (Lougheed 2012, 117).

They identified clinical nomenclatures such as the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT®) and LOINC as standards currently available for clinical variables that are likely to be included in elec-

tronic medical records in primary care for diagnosis, management and patient education related to asthma and Chronic Obstructive Pulmonary Disease (COPD), and other respiratory conditions. Among the identified LOINC codes, there are many regarding clinical parameters influenced or determined by environmental aspects, thus making evident direct or indirect relationships between the two observed diseases and environmental determinants.

3.0 Materials and methods

The first step of this work was identifying existing LOINC terms that could be related, both broadly or narrowly, to environment conditioned factors, so as to evaluate the degree of semantic coverage of the standard in the target domain. They had been retrieved using entries of specific controlled terminologies as search keywords. Among the most used Knowledge Organization Systems (KOS) of the domain, we considered GEMET (General Multilingual Environmental Thesaurus),[1] AGROVOC (https://agrovoc.fao.org/browse/agrovoc/en/), SWEET (Semantic Web for Earth and Environment Technology) Ontology,[2] Earth (which includes GEMET)[3] and INSPIRE Feature Concept Dictionary,[4] as references. Firstly, we used the search keyword *health* in all of them, as the aim was looking for how healthcare related concepts are represented in terminological resources primarily dedicated to the environmental domain. We retrieved 29 concepts from GEMET, 19 from AGROVOC, 4 from SWEET Ontology, 46 from Earth and 5 from INSPIRE. Then they were matched among each other so to create a list of unique concepts and avoid to double use them. This operation revealed that GEMET has the major part of its health-related concepts covered by the other considered terminological resources, while INSPIRE has no intersections with the other four vocabularies of the domain. The unique concepts of the list were 70 and they were used as searching keywords in LOINC to recall codes of those clinical observations or parameters that could be in some way related

1 European Environment Information and Observation Network (EIONET), "GEMET Thematic Listings," EIONET Portal, last accessed October 11, 2021, https://www.eionet.europa.eu/gemet/en/themes/.

2 National Center for Biomedical Ontology, "Semantic Web for Earth and Environment Technology Ontology," BioPortal, last accessed October 11, 2021, https://bioportal.bioontology.org/ontologies/SWEET/?p=classes&conceptid=root.

3 National Research Council – Institute of Polar Sciences (CNR-ISP), "EARTh – Environmental Applications Reference Thesaurus," last accessed October 11, 2021, https://www.isp.cnr.it/index.php/en/earth.

4 European Commission, "INSPIRE feature concept dictionary," INSPIRE Knowledge Base, last accessed October 11, 2021, https://inspire.ec.europa.eu/featureconcept.

LOINC code	Component	Property	Scale	Time	System	Method	Class
89551-6	Consultation note	Find	Doc	Pt	{Setting}	Environmental health	DOC.ONTOLOGY
	A consultation note is generated by a provider upon request for an opinion or advice from another provider. Consultations may involve face-to-face time with the patient, telemedicine visits, or a second opinion on a diagnosis that does not involve interaction with a patient. A consultation note is typically sent to the referring provider when the consultation is completed.						
42561-1	Event description.environmental hazard &or safety	Find	Nom	Pt	^Patient	Observed.MERSTH	PATIENT SAFETY
	Normative Answer List LL415-1 AnswerCodeScoreAnswer ID Body fluid exposureLA7393-7 Chemical exposureLA7394-5 Chemotherapy spillLA7395-2 Hazardous material spillLA7396-0 OtherLA46-8						
89552-4	Note	Find	Doc	Pt	{Setting}	Environmental health	DOC.ONTOLOGY
96336-3	Note	Find	Doc	Pt	Outpatient	Environmental health	DOC.ONTOLOGY

Figure 1. Pertinent LOINC codes recalled by the searching keyword environmental health

to environmental aspects. Some concepts were excluded because too generic for a health-related terminology such as LOINC and they would have recalled a lot of noise (i.e *health*, *healthcare*, *health service*, *public health*). Obviously narrower keywords recalled the same LOINC codes of their correspondent broader terms, i.e. typing *environmental health hazard* we found one LOINC code, which is comprised in the ten codes recalled by *environmental health*. So narrower keywords were not considered, if they are covered by a related broader term. Then, we analyzed the LOINC codes recalled by each searching keywords in order to eliminate those not relevant for the purpose of this work. For example, *environmental health* recalled ten LOINC codes, but only four of them are pertinent for the considered domain (Figure 1).

As 25 out of 70 searching keywords of our unique concepts list did not recall any LOINC code and also many of the retrieved ones are not relevant for the aims of the study, we decided to search in LOINC using the upper categories of the considered KOSs as keywords. Also in this case we excluded extremely generic words such as *administration* in GEMET. This choice created more noise, but it also gave the chance to retrieve more LOINC codes, which are logically related to the "health and environment" domain even if they do not contain those two words in the code description string. Consequently, more analysis work has been required on the recalled codes to fully understand the context of use. At the end of this process, we have identified a subset of 476 LOINC codes of clinical observations in some way related to environmental influence.

As second step, in order to verify if there could be possible LOINC concept gaps to fill by submitting requests for the creation of new codes to the LOINC Committee, we searched the UMLS (Unified Medical Language System) Metathesaurus[5] for those concepts not semantically covered by LOINC,

5 «The Metathesaurus is the biggest component of the UMLS. It is a large biomedical thesaurus that is organized by concept, or meaning, and it links similar names for the

but that are representative of the reference domain (i.e *soil*). This activity will allow a cross-checking of the reference concepts in multiple KOSs, especially those coding problems and diseases such as the International Classification of Diseases (ICD). Obtained results will support the need to create new LOINC codes of clinical observations immediately referable to diseases determined by environmental factors encoded in other medical terminologies.

4.0 Results

The LOINC codes retrieved according to the methodology described in the previous paragraph are quite representative enough of the target domain of this work. They are not systematically organized under a common and dedicated Class tag within the standard, but if considered together they actually well represent those clinical parameters in some way related or influenced by environmental factors.

The keywords that matched more pertinent LOINC codes are *air* and *water*, with 195 and 156 retrieved concepts respectively. That is because they are values of the LOINC System axis, expressing the means in which polluting substances can be measured. The majority of these codes belong to the Toxicology LOINC Class and report observations about substances in water or air that could in some way affect human health. Figure 2 and Figure 3 show the descriptions associated to the LOINC codes retrieved, where we underlined in red the parts containing the effects on health of the measured substances, in *air* and in *water* respectively.

Also the LOINC codes retrieved using the keyword *environmental health* belong to the Toxicology Class. They mainly search for the presence of toxic or harmful substances for the human body or metals or volatile organic compounds, both present in nature and by chemical composition. Most of them are panels required by the Center for Environmental Health (CEH),[6] which is an American organization aiming at protecting people from harmful chemicals in air, food, water and in everyday products. A panel in LOINC is a list of clinical observations usually made together. For example, in Figure 4 there is a panel about air contaminants that could be detected in a house.

---

same concept from nearly 200 different vocabularies. The Metathesaurus also identifies useful relationships between concepts and preserves the meanings, concept names, and relationships from each vocabulary». Cfr. National Institute of Health (NIH), "UMLS Metathesaurus," National Library of Medicine (NLM), last reviewed April 12, 2016, https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html.

6 "Center for Environmental Health," Wikipedia, last accessed October 11, 2021, https://en.wikipedia.org/wiki/Center_for_Environmental_Health.

38624-3	**1,1,2-Trichloroethane [Mass/volume] in Air**
	1,1,2-Trichloroethane, or 1,1,2-TCA, is an organochloride solvent with the molecular formula C2H3Cl3. It is a colorless, sweet-smelling liquid that does not dissolve in water, but is soluble in most organic solvents. It is an isomer of 1,1,1-trichloroethane. It is used as a solvent and as an intermediate in the synthesis of 1,1-dichloroethane. 1,1,2-TCA is a central nervous system depressant and inhalation of vapors may cause dizziness, drowsiness, headache, nausea, shortness of breath, unconsciousness, or cancer. Trichloroethane may be harmful by inhalation, ingestion and skin contact. It is a respiratory and eye irritant. Although no definitive studies currently exist, trichlorethane should be treated as a potential carcinogen since laboratory evidence suggests that low molecular weight chlorinated hydrocarbons may be carcinogenic
38625-0	**Trichloroethylene [Mass/volume] in Air**
	A highly volatile inhalation anesthetic used mainly in short surgical procedures where light anesthesia with good analgesia is required. It is also used as an industrial solvent. Prolonged exposure to high concentrations of the vapor can lead to cardiotoxicity and neurological impairment.
38594-8	**Epichlorohydrin [Mass/volume] in Air**
	A chlorinated epoxy compound used as an industrial solvent. It is a strong skin irritant and carcinogen
38693-8	**Ethyl acrylate [Mass/volume] in Air**
	Ethyl acrylate is an organic compound primarily used in the preparation of various polymers. It is a clear liquid with an acrid penetrating odor. The human nose is capable of detecting this odor at a thousand times lower concentration than is considered harmful if continuously exposed for some period of time.

Figure 2. LOINC codes and descriptions of effects on human health of some substances measured in the air

62534-3	PhenX - environmental exposures - air contaminants in the home environment protocol 061101
Indent63790-0	In the past year has there been a major renovation to this house or apartment, such as adding a room, putting up or taking down a wall, replacing wiindows, or refinishing floors?
Indent63791-8	What type of renovation?
Indent63792-6	When was the last renovation?
Indent63793-4	Within the last six months were rugs, drapes or furniture professionally cleaned?
Indent63794-2	Were they cleaned inside the house?
Indent63795-9	When were they cleaned?
Indent63796-7	What items were cleaned?
Indent63797-5	In the past year, was the inside of this house or apartment painted?
Indent63798-3	When was the last time?
Indent63799-1	On how many rooms?
Indent63845-2	In the past year were new carpets or rugs installed?
Indent63846-0	When was the installation?
Indent63847-8	On how many rooms?
Indent63800-7	Did you go to the dry cleaners during the past week?
Indent63801-5	Did you bring home any items from the cleaners that were dry-cleaned during the past week?
Indent63802-3	Specify air contaminant in the home [RIOPA]
Indent63808-0	Which cleaning solutions [RIOPA]
Indent63803-1	Have you used Paints or solvents (paint thinners and removers, typewriter corrective fluids) or has someone used near you in the last 48 hours?
Indent63804-9	Did you handle Paints or solvents (paint thinners and removers, typewriter corrective fluids) yourself? If Yes, for how long?
Indent63805-6	How long did you handle Paints or solvents (paint thinners and removers, typewriter corrective fluids) yourself?

Figure 3. LOINC codes and descriptions of effects on human health of some substances measured in the water

LOINC was born to provide unique identifiers for 'what' is measured, so if an observation is a question and its result is an answer, LOINC codes are created for the question. The answer can be coded using other standardized terminologies, such as SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms) (https://www.snomed.org/). Nonetheless, as

> "Across many domains, the meaning of a particular observation can be best understood in the context of the set of possible answers (result values)" (Regenstrief Institute 2021, 114).

38334-9	**1,1,2-Trichloroethane [Mass/volume] in Water**
	1,1,2-Trichloroethane, or 1,1,2-TCA, is an organochloride solvent with the molecular formula C2H3Cl3. It is a colorless, sweet-smelling liquid that does not dissolve in water, but is soluble in most organic solvents. It is an isomer of 1,1,1-trichloroethane. It is used as a solvent and as an intermediate in the synthesis of 1,1-dichloroethane. 1,1,2-TCA is a central nervous system depressant and inhalation of vapors may cause dizziness, drowsiness, headache, nausea, shortness of breath, unconsciousness, or cancer. Trichloroethane may be harmful by inhalation, ingestion and skin contact. It is a respiratory and eye irritant. Although no definitive studies currently exist, trichlorethane should be treated as a potential carcinogen since laboratory evidence suggests that low molecular weight chlorinated hydrocarbons may be carcinogenic
38292-9	**1,1-Dichloroethylene [Mass/volume] in Water**
	1,1-Dichloroethene, commonly called 1,1-dichloroethylene or 1,1-DCE, is an organochloride with the molecular formula C2H2Cl2. It is a colorless liquid with a sharp odor. Like most chlorocarbons, it is poorly soluble in water, but soluble in organic solvents. 1,1-DCE was the precursor to the original cling-wrap for food, but this application has been phased out. The health effects from exposure to 1,1-DCE are primarily on the central nervous system, including symptoms of sedation, inebriation, convulsions, spasms, and unconsciousness at high concentrations
38332-3	**1,2,4-Trichlorobenzene [Mass/volume] in Water**
	1,2,4-Trichlorobenzene is an organic compound used as a solvent, and is one of the best known solvents used to dissolve fullerenes and pentacene. It is a benzene derivative with three chlorine atoms substitutents, in the 1, 2 and 4 positions of the benzene ring. Trichlorobenzene (TCB) may refer to any of three isomeric chlorinated derivatives of benzene with the molecular formula C6H3Cl3: - 1,2,3-Trichlorobenzene - 1,2,4-Trichlorobenzene - 1,3,5-Trichlorobenzene
38288-7	**1,2-Dibromo-3-Chloropropane [Mass/volume] in Water**
	1,2-Dibromo-3-chloropropane, (dibromochloropropane) better known as DBCP, is the active ingredient in the nematicide Nemagon, also known as Fumazone. It is a soil fumigant formerly used in American agriculture. In mammals it causes male sterility at high levels of exposure.

Figure 4. LOINC panel 62534-3 about air contaminants in the home environment

63802-3	Specify air contaminant in the home [RIOPA]
	Paints or solvents, paint thinners and removers, typewriter corrective fluids
	Glues and adhesives
	Gasoline lawn mower
	Chain saw or other gasoline equipment
	Sander and/or saw
	Pesticides sprayed
	Vacuuming
	Sweeping indoors
	Dusting
	Cleaning solutions, including household cleaners and chemicals
	Gardening
	Woodworking
	Metal working or welding
	Broiling, smoking, grilling or frying inside the house
	Broiling, smoking, grilling or frying outside the house

Figure 5. LOINC Answers List of the code 63802-3 specifying home air contaminants

the LOINC Committee decided to create LOINC answer lists for certain types of observations. They can be *example*, *preferred* or *normative*, which bind the choice of that LOINC code to the use of its values. For example, LOINC Answers (LA) Lists in Figure 5 show which air contaminants of the home can be specified as results of the LOINC observation 63802-3.

Exploring other results, the concept *disaster* retrieved only two pertinent LOINC codes (Figure 6). One (LOINC 66886-3) investigates if the patient has ever been involved in a natural disaster and it is part of Lifetime Trauma and Victimization History (LTVH) survey, which aims at inquiring about

66886-3	Have you ever been involved in a natural disaster, such as a tornado, hurricane, flood, or earthquake [LTVH]	
69463-8	Suspected intentional or unintentional disaster NEMSIS	
	Biologic agent	4507001
	Building failure	4507003
	Chemical agent	4507005
	Explosive device	4507007
	Fire	4507009
	Hostage event	4507011
	Mass gathering	4507013
	Mass illness	4507015
	Nuclear agent	4507017
	Radioactive device	4507019
	Secondary destructive device	4507021
	Shooting/sniper	4507023
	Vehicular	4507025
	Weather	4507027

Figure 6. LOINC codes retrieved by the keyword disaster *and LA list of the second one*

92036-3	**Action taken based on risk assessment meeting**	PUBLICHEALTH
91022-4	Date and time of risk assessment	PUBLICHEALTH
91880-5	Date and time of risk assessment meeting	PUBLICHEALTH
28191-5	**Poisoning risk [CCC]**	
	Exposure to or ingestion of dangerous products	
91578-5	**Probability of an event occurring**	
	The level of probability (e.g. very unlikely, unlikely, likely, highly likely, almost certain) that the event will occur or re-occur. An event that is very likely (almost certain) is one that will occur (or re-occur) in most circumstances when the activity is undertaken. An event that is very unlikely (rare) may happen only in exceptional circumstances when the activity is undertaken.	
74019-1	**Probability of developing disease assessed [Likelihood]**	
	The estimated likelihood of a patient developing a disease. Probability may be based a number of factors, including family history (pedigree) analysis, genetic algorithm results, pre- or post-test results, and general population data for disease prevalence.	
91023-2	**Reason for risk assessment**	
	The reason the risk assessment is conducted. The reason may simply be the incident or event name.	PUBLICHEALTH
89933-6	**Relative risk of developing disease assessed**	
	Relative risk or risk ratio (RR) of developing disease is the ratio of the probability of developing a disease in an exposed group to the probability of the event occurring in a non-exposed group.	
71482-4	risk assessment document	

Figure 7. LOINC codes retrieved by the keyword risk *and their descriptions*

30 traumas and victimization experiences people can undergo during their lifetime (Widom 2005). The other aims to code the type of disaster the patient experienced, giving the chance to choose among some in a normative LA list, and it was created upon request of the National Emergency Medical Services Information System (NEMSIS) (https://nemsis.org/).

The keyword *risk* gave back 12 pertinent LOINC codes out of 492 retrieved records (Figure 7). Among them, there are some to be noticed: *Poisoning risk* (LOINC 28191-5), which registers the patient exposure to or ingestion of dangerous products; *Probability of developing disease assessed* (LOINC 74019-1), which is estimated on a number of factors, including family history (pedigree) analysis, genetic algorithm results, pre- or post-test results, and general population data for disease prevalence, but not on risks related to environmental factors, both natural or determined by the place where the patient lives or works. So, in this case, the code could work for the purpose

75274-1	**Characteristics of residence**
	The patient's home-setting circumstances reflecting the status of the accommodations, living situation, and environment
63839-5	**Do you have wall-to-wall carpeting in your home [FEAS]**
	excluding the bathroom
63833-8	**Floor material in your workspace if other [FEAS]**
63837-9	Has there been renovation in your home during the past 12 months [FEAS]
63836-1	Has there been renovation or repairs in your home due to moisture damage [FEAS]
63831-2	Has there been renovation or repairs in your workspace because of moisture damage [FEAS]
71802-3	**Housing status**
	Describes patients living arrangement
63840-3	**Is the floor material in your home plastic or vinyl [FEAS]**
63842-9	**Is the wall material in your home plastic [FEAS]**
63841-1	**Is the wall material in your home textile, cloth, jute, etc [FEAS]**
63835-3	**Is the wall material of your work space plastic [FEAS]**
63834-6	**Is the wall material of your work space textile, cloth, jute, etc [FEAS]**
96778-6	**Problems with place where you live**
	Pests such as bugs, ants, or mice
	Mold
	Lead paint or pipes
	Lack of heat
	Oven or stove not working

Figure 8. Some of the LOINC codes retrieved by the keyword pollution

of registering a clinical observation related to the influence of environment on health, but its description needs to be updated. Similarly, for the LOINC 71482-4 *Risk assessment document*, though it was originally created for use within the Cardiac Imaging Report CDA Implementation Guide developed by IHE (Integrating the Healthcare Enterprise), the description says it is not limited in use to it, but the chance to refer it to environmental related risks is not made explicit. They are all codes belonging to the LOINC Class *Public health* and that is indicative of the scope and purpose for which they were created.

Last but not least, it is worth mentioning the concept *pollution*. Looking at the retrieved LOINC codes (Figure 8), they are mostly related to the house environment or workspace, as they investigate aspects such as the floor and wall material, the presence of wall-to-wall carpeting and other general problems (i.e. pests, mold, lead paint or pipes, lack of heat, oven or stove not working). There are no codes related to the pollution that can be caused by natural factors or human activities (such as industries) and could cause effects on human health.

Despite the numerous and pertinent LOINC codes retrieved through the described keywords, many concepts of the upper categories of the considered KOSs, which are representative of the reference domain, did not find any correspondence in the standard. Among them, those for which it would be appropriate to have a representation in LOINC, because in some way relatable to environmental factors influencing human health, are: *agriculture*, *biology*, *climate*, *energy*, *forestry*, *natural areas*, *ecosystems*, *natural dynamics*,

vibrations, *soil*, *transport*, *urban environment*, *urban stress*. The second step of this work, as described in the previous paragraph, was searching for these high-level concepts in the UMLS Metathesaurus, in order to see if and how they are represented in other medical terminologies and so determining clinical semantic areas that would need to be covered by LOINC. In particular, searching results belonging to KOSs that encode diseases and problems were considered, as it is consequential that clinical observations or parameters are mostly defined in relation to them.

Using the concept *agriculture* as search keyword in the UMLS Metathesaurus Browser[7] and looking for vocabulary codes,[8] we obtained 49 results. Among them, one is relevant for the purpose of this study: *occupational exposure to toxic agents in agriculture*, which is encoded in ICD10[9] (Z57.4), ICD10CM[10] (Z57.4), ICD10AM[11] (Z57.4) and MEDCIN[12] (129080). This concept could be formalized in LOINC similarly to codes that evaluate occupational risks due to the presence harmful substances, in this case specifically related to the agricultural work.

Searching for *biology* gives back 155 results, but none of them is useful in identifying clinical conditions determined by environmental factors, because retrieved codes are mainly of biology subcategories so the term in this form is surely inclusive of concepts of interest but too high level for our aims.

Typing the word *climate*, 56 resulting codes are recalled. Most of them are for climate types, but some can be considered as in some way expressing the cause effect relationship between climate changes (intended both as a

7 National Institute of Health (NIH), "UMLS Metathesaurus Browser," National Library of Medicine, last accessed October 11, 2021, https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html.

8 It is also possible to search for UMLS concepts. In this case, codes belonging to different vocabularies but carrying the same semantic meaning are grouped together under the same UMLS CUI (Concept Unique Identifier). For example, searching for agriculture will retrieve 19 different CUIs against 49 vocabulary codes, meaning that those 49 carry 19 different concepts, some of which are repeated in more than one code system so they can be conceptually gathered together.

9 World Health Organization (WHO), "International Classification of Diseases 10th revision," last accessed October 11, 2021, https://www.who.int/standards/classifications/classification-of-diseases.

10 The Centers for Disease Control (CDC), "International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)," Centers for Diseases Control and Prevention, last accessed October 11, 2021, https://www.cdc.gov/nchs/icd/icd10cm.htm.

11 Independent Hospital Pricing Authority (IHPA), "International Classification of Diseases 10th revision Australian Modification," last accessed October 11, 2021, https://www.ihpa.gov.au/what-we-do/icd-10-am-achi-acs-current-edition.

12 Medicomp Systems, "The Medicomp MEDCIN Engine," last accessed October 11, 2021, https://medicomp.com/medcin/.

change in climate due to a change in geographical area and as a major climate change in general) and health status. They are:

- *recently traveled to or resided in tropical climate* (MEDCIN 3152);
- *recently traveled to or resided in desert climate* (MEDCIN 3153);
- *recently traveled to or resided in temperate climate* (MEDCIN 3150);
- *recently traveled to or resided in subtropical climate* (MEDCIN 3151);
- *traveled to or resided in climate which is:* (MEDCIN 3149);
- *climate change* (MEDLINEPLUS 5144) (https://medlineplus.gov/), which is a subcategory of *Poisoning, Toxicology, Environmental health;*
- *climatotherapy* (MESH[13] D013790), which consists into the temporary or permanent relocation of a patient in a place with a climate that can help the recovery or improvement of a pathological condition.

The 1,628 codes recalled by the searching keyword *energy* are mostly referred to therapies and treatments that involve the use of some form of electric energy, to energy intended as physical vigor or to the thermal energy potential enclosed in a food. While only the code 242720004 *Accident caused by heat energy release from machine* from SNOMED CT refers to accidental harms caused by heat energy, therefore considering the concept from a negative point of view respect to the other uses.

The concept *forestry* does not have any code from the medical terminologies in the UMLS that could be related to some influence on human health, but it just recalls 24 codes mostly about forestry as a place or as a job sector. Similarly, the queries using the concepts *natural areas* and *natural dynamics* yielded no results, as probably they are too high level for our purposes. While using *ecosystems* as search keyword gives only 2 results, meaning that it has not been semantically explored in any medical terminology of the UMLS yet.

Useful to be encoded in LOINC would be the concept conveyed by the MedDRA French Editon[14] code 10030023 *Exposition professionnelle aux vibrations*, which is among the 6 codes recalled by the search term *vibrations*.

The search keyword *soil* retrieved the highest number of meaningful resulting codes. Many of them refer to different types of *soil bacteria* and *soil fungi* and come from the NCBI (National Center for Biotechnology) Taxonomy.[15] There are two codes for *soil pullutants*, one of which specifying if they

13 National Center for Biotechnology Information (NCBI), "Medical Subject Headings," last accessed October 11, 2021, https://www.ncbi.nlm.nih.gov/mesh/.

14 International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH), "MedDRA Support Documentation, Medical Dictionary for Regulatory Activities," last accessed October 11, 2021, https://www.meddra.org/how-to-use/support-documentation/french.

15 National Center for Biotechnology Information (NCBI), "Taxonomy," last accessed October 11, 2021, https://www.ncbi.nlm.nih.gov/guide/taxonomy/.

are *radioactive*, from MESH (D012989 and D012990, respectively). Similarly, a number of codes from different medical terminologies encodes the concept *soil pollution*:

- A-A3040 of SNMI;[16]
- 2733-6435 of CSP;[17]
- 102417007 of SNOMEDCT_US;[18]
- 0000038184 of CHV (https://consumerhealthvocab.org/).

Other than these codes, which are essentially used to identify organisms, substances or a certain condition of the soil, there are those specifically expressing a cause-effect relation between a polluted or contaminated soil and the health status. Among them, it is also possible to make a distinction between codes expressing an exposure or contact because of occupational or non-occupation reasons:

- *Exposure to polluted soil* (SNOMEDCT_US 102432000);
- *Exposure to polluted soil, occupational* (SNOMEDCT_US 102433005);
- *Exposure to polluted soil, non-occupational* (SNOMEDCT_US 102434004);
- *Contact with and (suspected) exposure to soil pollution* (ICD10CM Z77.112);
- *Exposure to polluted soil, NOS* (SNMI A-A3240);
- *Exposure to polluted soil, occupational* (SNMI A-A3241);
- *Exposure to polluted soil, non-occupational* (SNMI A-A3242);
- *Exposure to polluted soil* (MedDRA 10063603) (https://www.meddra.org/);
- *Non-occupational exposure to polluted soil* (MedDRA 10056402);
- *Exposure to soil pollution* (ICD10AM Z58.3);
- *Exposure to soil pollution* (MEDCIN 129110);
- *Exposure to possibly contaminated soil* (MEDCIN 3293).

The concepts vehicled by these codes are not covered by LOINC at all, so they are meaningful candidates for formulating a new term request submission to the LOINC Committee.

16 SNOMED International 1998, which «has been superseded by the Systematized Nomenclature of Medicine – Reference Terminology (SNOMED RT®). SNOMED RT has been merged with Clinical Terms Version 3 (CTV3), also known as Read Codes, resulting in the creation of SNOMED Clinical Terms (SNOMED CT®)». National Library of Medicine (NLM), "SNMI (SNOMED Intl 1998) – Synopsis," last accessed October 11, 2021, https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/SNMI/index.html.

17 National Institute of Health (NIH), "CSP (CRISP Thesaurus) – Synopsis," National Library of Medicine, last accessed October 11, 2021, https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/CSP/index.html.

18 National Institute of Health (NIH), "SNOMED CT United States Edition," National Library of Medicine, last accessed October 11, 2021, https://www.nlm.nih.gov/healthit/snomedct/us_edition.html.

5.0 Discussion and conclusions

Standardized KOSs need continuous updates if they want to keep up with the ongoing discoveries and evolution of the scientific world and, moreover, to maintain their usefulness as representation and access tools to the knowledge of a specific domain. Cross references among them are vital as they avoid leaving them as single islands and create a complete representation of multiple and interacting aspects of different domains. The links between environment and health are increasing hand in hand with the influence that environmental factors have on the development and progression of certain diseases. Nonetheless, this semantic area is not yet fully represented in standardized medical terminologies especially for what pertains to observable clinical parameters (as trigger factors, consequences, indicators, etc.) related to a disease. This paper aimed to investigate existing LOINC codes for this semantic area, then searching concepts not covered in the UMLS Metathesaurus to see if they are covered and how they are expressed in other medical terminologies and consequently detecting LOINC concept gaps related to the link between environment and health to be filled by proposing new LOINC terms for standardizing identified clinical parameters.

The study gave good results in terms of LOINC evaluation as both gold standard for identifying clinical observations also in the environmental domain and for detecting semantic area not covered by the standard and therefore susceptible to being modeled by requesting the creation of new *ad hoc* LOINC codes. Despite the fact that it is the most widely used standard for uniquely identifying laboratory and clinical observations, it has not fully developed sections or health related domains that need to be addressed more completely. On the other hand, compared to other medical terminologies, LOINC has on its side that it has an open, rapid and pragmatic updating process as it is based on user requests and therefore on the real coding needs connected to new scientific discoveries or evolutions. So, the results of this work could concretely contribute to the development of the standard and they will therefore be formalized according to the LOINC naming conventions to be submitted to the LOINC Committee, that will evaluate the creation of new LOINC codes for the identified clinical observations. Finally, based on the results of this study, the LOINC Committee could evaluate the opportunity to create a LOINC Class entirely dedicated to clinical observations related to environmental factors.

Although this work could have been difficult to approach due to the many facets that the concept of *environment* applied to the world of *health* can assume, the defined working methodology has brought significant results both in terms of numbers and quality. However, it is possible to trace some limits of the study, which can actually be considered as starting points for

future refinements or in-depth studies. Among them, the chance to consider, in order to create the subset of searching keywords: I) more environmental vocabularies; II) not only high-level concepts, but also their subordinate, if significative; III) clinical concepts not only traceable through the term *health*. Moreover, the UMLS Metathesaurus is surely an important source of information as it allows to access multiple terminologies speedly and simultaneously, but it would be further useful to use it just as a starting point and then investigate each concept in its own vocabulary in order to consider its position in a hierarchical tree, possible meaningful branches, etc.

In the future, this work could be expanded through the collaboration with a medical expert for double-checking the identified concept gaps and evaluating if there are clinical conditions or problems that are in some way related to the environment influence but that are not yet encoded in any medical terminology.

In conclusion, the relationships between the environmental and medical domains are proven by various studies and it is of fundamental importance to make them explicit through measurable parameters as they allow them to be monitored and corrective actions taken in both sectors, for the benefit of both the environment and human health. To pursue this aim, standard terminologies ensure semantic interoperability by uniquely identifying concepts and creating semantic relationships among them using tools like the UMLS Metathesaurus. Open, nimble and pragmatic vocabularies like LOINC offer the chance to model the topic of this paper, which crosses multiple domains, according to the needs of the final users and without getting stuck in the rigidity of the branches of a classification system. Representing the knowledge of transversal domains is a non-trivial task, which has to be approached from multiple points of view, but which is necessary to have a shared identification of new arising concepts.

References

Centers for Disease Control (CDC). "International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)." Centers for Diseases Control and Prevention. Last accessed October 11, 2021. https://www.cdc.gov/nchs/icd/icd10cm.htm.

"Center for Environmental Health." Wikipedia. Last accessed October 11, 2021. https://en.wikipedia.org/wiki/Center_for_Environmental_Health.

Dong, Xiao, Jianfu Li, Ekin Soysal, Jiang Bian, Scott L. DuVall, Elizabeth Hanchrow, Hongfang Liu, Kristine E. Lynch, Michael Matheny, Karthik Natarajan, et al. 2020. "COVID-19 TestNorm: A tool to normalize COVID-19 testing names to LOINC codes." *Journal of American Medical Informatics Association* 27, no. 9: 1437-42.

European Commission. "INSPIRE feature concept dictionary." INSPIRE Knowledge Base. Last accessed October 11, 2021. https://inspire.ec.europa.eu/featureconcept.

European Environment Information and Observation Network (EIONET). "GEMET Thematic Listings." EIONET Portal. Last accessed October 11, 2021. https://www.eionet.europa.eu/gemet/en/themes/.

Frazier, Pavla, Angelo Rossi-Mori, Robert H. Dolin, Liora Alschuler, and Stanley M. Huff. 2001. "The creation of an ontology of clinical document names." *Studies in Health Technologies and Informatics* 84: 94-8.

Independent Hospital Pricing Authority (IHPA). "International Classification of Diseases 10th revision Australian Modification." Last accessed October 11, 2021. https://www.ihpa.gov.au/what-we-do/icd-10-am-achi-acs-current-edition.

International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). "MedDRA Support Documentation. Medical Dictionary for Regulatory Activities." Last accessed October 11, 2021. https://www.meddra.org/how-to-use/support-documentation/french.

Lougheed, M. Diane, Janice Minard, Shari Dworkin, Mary-Ann Juurlink, Walley J. Temple, Teresa To, Marc Kohen, Anne Van Dam, and Louis-Philippe Boulet. 2012. "Pan-Canadian REspiratory STandards INitiative for Electronic Health Records (PRESTINE): 2011 National Forum Proceedings." *Canadian Respiratory Journal* 19: 117-26. https://doi.org/10.1155/2012/870357.

McDonald, Clement J. 1997. "The barriers to electronic medical record systems and how to overcome them." *Journal of the American Medical Informatics Association* 4: 213-21. https://doi.org/10.1136/jamia.1997.0040213.

McDonald, Clement J., Stanley M. Huff, Jeffrey G. Suico, Gilbert Hill, Dennis Leavelle, Raymond Aller, Arden Forrey, Kathy Mercer, Georges DeMoor, et al. 2003. "LOINC, a universal standard for identifying laboratory observations: a 5-year update." *Clinical Chemistry* 49: 624-33. https://doi.org/10.1373/49.4.624.

Medicomp Systems. "The Medicomp MEDCIN Engine." Last accessed October 11, 2021. https://medicomp.com/medcin/.

Mekonnen, Tesfaye H., Awrajaw Dessie, and Amensisa H. Tesfaye. 2021. "Respiratory symptoms related to flour dust exposure are significantly high among small and medium scale flour mill workers in Ethiopia: a comparative cross-sectional survey." *Environmental health and preventive medicine* 26, 1: 96. https://doi.org/10.1186/s12199-021-01019-y.

Nakayama, Kazumi, Yoshihide Sugita, Hideaki Tezuka, Koichi Takizawa, and Shoichi Muto. 2021. "Evaluation of the effect of radioactivity to the environment in wooden houses within the evacuation area at Fukushima." *Journal of radiological protection: official journal of the Society for Radiological Protection* 41, no. 4. https://doi.org/10.1088/1361-6498/ac2b8d.

National Center for Biotechnology Information (NCBI). "Medical Subject Headings." Last accessed October 11, 2021. https://www.ncbi.nlm.nih.gov/mesh/.

National Center for Biotechnology Information (NCBI). "Taxonomy." Last accessed October 11, 2021. https://www.ncbi.nlm.nih.gov/guide/taxonomy/.

National Institute of Health (NIH). "CSP (CRISP Thesaurus) – Synopsis." National Library of Medicine. Last accessed October 11, 2021. https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/CSP/index.html.

National Institute of Health (NIH). "UMLS Metathesaurus." National Library of Medicine. Last reviewed April 12, 2016. https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html.

National Institute of Health (NIH). "UMLS Metathesaurus Browser." National Library of Medicine. Last accessed October 11, 2021. https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/index.html.

National Library of Medicine (NLM). "SNMI (SNOMED Intl 1998) – Synopsis." Last accessed October 11, 2021. https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/SNMI/index.html.

National Research Council – Institute of Polar Sciences (CNR-ISP). "EARTh – Environmental Applications Reference Thesaurus." Last accessed October 11, 2021. https://www.isp.cnr.it/index.php/en/earth.

Park, Hyun S., Kwang I. Kim, Ho-Young Chung, Sungmoon Jeong, Jae Y. Soh, Young H. Hyun, and Hwa S. Kim. 2021. "A Worker-Centered Personal Health Record App for Workplace Health Promotion Using National Health Care Data Sets: Design and Development Study." *JMIR medical informatics* 9, 8:e29184. https://doi.org/10.2196/29184.

Regenstrief Institute (RI). 2021. "LOINC Users' guide." Last accessed October 11, 2021. https://loinc.org/downloads/loinc-table/.

Stringer, Leandra, Tina L. Ly, Nicolas V. Moreno, Christopher Hewitt, Michael Haan, and Nicholas Power. 2021. "Assessing geographic and industry-related trends in bladder cancer in Ontario: A population-based study." *Canadian Urological Association journal*. https://doi.org/10.5489/cuaj.7263.

Vreeman, Daniel J., and Clement J.McDonald. 2005. "Automated mapping of local radiology terms to LOINC." In *AMIA Annual Symposium Proceedings, 22-26 October 2005 Washington, DC*, 769-73.

Vreeman, Daniel J., Maria T. Chiaravalloti, John Hook, and Clement J. McDonald. 2012. "Enabling international adoption of LOINC through translation." *Journal of Biomedical Informatics* 45, no. 4: 667-73. https://doi.org/10.1016/j.jbi.2012.01.005.

Watkins, Michael, Benjamin Viernes, Viet Nguyen, Leonardo Rojas Mezarina, Javier Silva Valencia, and Damian Borbolla. 2021. "Translating Social Determinants of Health into Standardized Clinical Entities." *Studies in health technology and informatics* 270: 474-78. https://doi.org/10.3233/SHTI200205.

Widom, Cathy S., Mary A. Dutton, Sally J. Czaja, and Kimberly A. DuMont. 2005. "Development and validation of a new instrument to assess lifetime trauma and victimization history." *Journal of Traumatic Stress* 18, no. 5: 519-31. https://doi.org/10.1002/jts.20060.

World Health Organization (WHO). "International Classification of Diseases 10th revision." Last accessed October 11, 2021. https://www.who.int/standards/classifications/classification-of-diseases.

World Health Organization (WHO). "WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide." Last accessed October 11, 2021. https://apps.who.int/iris/handle/10665/345329. License: CC BY-NC-SA 3.0 IGO.