A Repository of Codes of Ethics and Technical Standards in Health Informatics

Hamman W. Samuel[1], Osmar R. Zaïane[1]

1 Department of Computing Science, University of Alberta, Edmonton, AB, Canada

Abstract

We present a searchable repository of codes of ethics and standards in health informatics. It is built using state-of-the-art search algorithms and technologies. The repository will be potentially beneficial for public health practitioners, researchers, and software developers in finding and comparing ethics topics of interest. Public health clinics, clinicians, and researchers can use the repository platform as a one-stop reference for various ethics codes and standards. In addition, the repository interface is built for easy navigation, fast search, and side-by-side comparative reading of documents. Our selection criteria for codes and standards are two-fold; firstly, to maintain intellectual property rights, we index only codes and standards freely available on the internet. Secondly, major international, regional, and national health informatics bodies across the globe are surveyed with the aim of understanding the landscape in this domain. We also look at prevalent technical standards in health informatics from major bodies such as the International Standards Organization (ISO) and the U. S. Food and Drug Administration (FDA). Our repository contains codes of ethics from the International Medical Informatics Association (IMIA), the iHealth Coalition (iHC), the American Health Information Management Association (AHIMA), the Australasian College of Health Informatics (ACHI), the British Computer Society (BCS), and the UK Council for Health Informatics Professions (UKCHIP), with room for adding more in the future. Our major contribution is enhancing the findability of codes and standards related to health informatics ethics by compilation and unified access through the health informatics ethics repository.

Keywords: health informatics ethics, repository, codes of ethics, technical standards, findability

Introduction

Ethics deals with decisions about right versus wrong or good versus bad. These normative and moral questions involve people and how they affect each other (1). In the field of medicine, ethical issues have been well-known due to the nature of the profession in dealing with life-and-death situations. In the present information age of ubiquitous computing, medicine and technology have become overlapped. The field of health informatics focuses on using computers to enhance the way health information is processed. There is extensive literature on the subject of ethical conduct and principles in medicine. Similarly, there are existing bodies of knowledge on ethics in computing and information technology. It can be argued that these two bodies of literature cover ethical issues in health informatics (2,3). However, we point out that information technology and medicine together lead to new and more complicated issues in ethics. For example, the Therac-26 incident is one of the many examples of software bugs causing loss of lives. Because of a bug in the underlying operating system of the Therac-26 x-ray device, patients died or were seriously injured due to over-exposure (4). Who was to blame: developers, testers, or operators?

Medical ethics also does not adequately cover modern innovations in medicine. For instance, the United Nations has a document covering the principles of medical ethics, but this has no provision for electronic health records (5). Also, the American Medical Association's (AMA) Code of Medical Ethics identifies physician records, yet does not address electronic transmission of patient records (6). On the other hand, computer ethics can deal with issues with data transmission and software development, but does not consider ethical requirements of the medical profession. Ethics issues in health informatics require more than just interdisciplinary cooperation in the fields of ethics, medicine, and computing. These issues are very important due to the vulnerability of people needing care and potential risks of using information technology to provide this care.

Moreover, while there are various codes of ethics and technical standards covering health informatics, the degree of findability of the documents could be enhanced. These documents are available on the respective websites of international, regional, and national bodies and organizations, and given the number of documents, as well as the diversity of content, synthesizing the viewpoints presented by the different documents can prove challenging. The web pages that host the documents also have limited search capabilities that could be improved. The proposed repository aims to address these issues by compilation and unified access to the various documents, along with state-of-the-art search and comparison tools.

The rest of the paper is structured as follows. We firstly present the methods used in carrying out this research, and declare any risks of bias. Next, we look at health informatics ethics by reviewing codes of ethics and technical guidelines from leading organizations across the globe. Later, we discuss the concept of findability and how it relates to codes and standards in health informatics. Finally, we present a repository of codes of ethics informatics and outline the specifications of the state-of-the-art search algorithms and technologies used.

Methodology

We carried out a review of literature, codes of ethics, and technical standards from major policy drivers and stakeholders. Firstly, we used the Health Informatics World Wide (HIWW) online database to identify health informatics organizations. Next, we identified ethical codes within the organizations and common memberships to international and regional bodies using the respective organization's website. Based on common threads, we then analyzed the major ethical principles being discussed. We also accessed pertinent literature via online subscription databases. We selected all international, regional and national bodies that were currently actively listed in the HIWW online database. We measured their state of activeness by visiting their websites and looking at the published time stamps on their web pages. Our selection criteria for the codes of ethics ultimately available in our repository were two-fold; firstly, to maintain intellectual property rights, we indexed only codes and standards freely available on the internet. Secondly, major international, regional and national health informatics organizations were surveyed, and representative bodies from the various continents were selected.

Limitations of the Study

The survey of health informatics bodies and organizations is limited by availability of information. Organizations that do not have some online portfolio were not included in the survey. We also acknowledge that some online resources might not have been accessible due to temporary server downtime. The study is also limited by language selection, and focuses on

English-language ethics codes and standards in health informatics. Furthermore, to respect intellectual property rights, only those codes and standards that are freely available on the World Wide Web are indexed in the health informatics ethics repository. Codes that required paid subscriptions or memberships are excluded from the repository.

Health Informatics Ethics

Ethics in health informatics, or health informatics ethics is the application of the principles of ethics to the domain of health informatics. Essentially, there are three aspects of health informatics: healthcare, information, software (7)□. Information systems are developed to facilitate dispensation of healthcare or the auxiliary activities involved in healthcare. Health informatics also deals with efficiently processing information. Large volumes of information about patients' needs to be stored for future reference, and retrieved when needed. Transfer of information between healthcare organizations needs to be handled with proper security. Ultimately, information is processed, stored and retrieved by software. Indeed, software systems are needed to manage clinics and hospitals, as well as information needed by larger healthcare providers. We can further define health informatics ethics using three hierarchies of ethics: general ethics, informatics ethics, and software engineering ethics, as illustrated in Figure 1 (8).
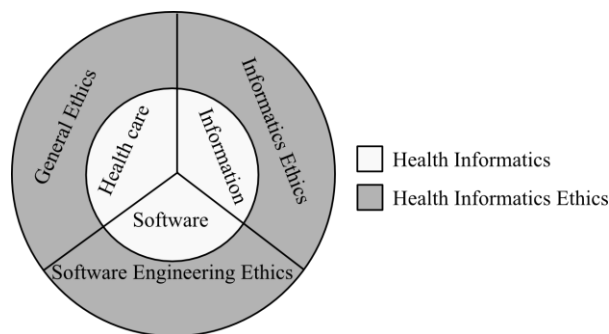

Figure 1. Aspects of health informatics ethics

All our social interactions dealing with norms and values are guided by general ethics, which covers six major principles: non-malfeasance, integrity, equality/justice, beneficence, autonomy, and impossibility (8). We have a duty to prevent harm to others without undue harm to ourselves, and a duty to fulfill our obligations to the best of our abilities. We also have the right to be treated equally, while advancing the good of others. We have the right to self-determination, and all our duties are subject to our ability to do them.

Informatics ethics, on the other hand, is a specific set of ethics principles for informatics. It deals with ethical behavior required of anyone handling data and information. Informatics ethics covers seven principles: privacy, openness, security, access, legitimate infringement, least intrusive alternatives, and accountability (8). Every person has the right to decide how much information they wish to disclose about themselves, and what information they wish to withhold, how it is stored, and used. The recent "Right To Be Forgotten" ruling against Google is an example (9). The person about whom the data is being collected needs to be informed of the intent for collecting data and what the data will be used for. Once data is collected, it ought to be safeguarded against unauthorized access. In addition, data needs to be protected against malicious and unintentional manipulation. Furthermore, everyone has the right to access and correct their own data. An individual's data should not be kept in isolation and allowed to

become outdated. However, access is subject to legitimate data needs of a free society. The individual's right to privacy may be infringed for the larger good of society. Any legitimate infringement should be done with minimum interference to individual rights. Legitimate infringement should be reported to the person affected in due time. This implies that eventually the affected individual should be notified of the infringement.

Health informatics ethics also involves use of software to store and process information. Consequently, activities carried out by software developers have the potential of affecting end-users (10). The software developer has ethical duties and responsibilities to the following six stakeholders: public, client, employer, management, colleagues, and self (11). Activities ought to be done with the best interest of the society in mind. Developers should be aware of social impacts of software systems. This includes disclosing any dangers or known defects in software. Also, activities ought to be done in the best interests of clients and employers, while balancing their duties to the public, including being forthright about personal limitations and qualifications. Software products should meet expected professional standards. Developers should strive to build products that are not sub-standard by thoroughly testing and documenting unsolved problems. Moreover, managers and leaders should subscribe to ethical approaches in software development. Realistic and effective costs, schedules, and procedures should be promoted. And continuing further, colleagues should be supported and treated fairly. Developers should fully credit their colleagues for their work, including intellectual contributions and code re-use. Finally, self-improvement and re-training ought to to be pursued by the software developer. In addition, developers should not let prejudices lead to unfair treatment of others.

Ethical conflicts arise as a result of interactions between various stakeholders involved in the health informatics setting (8). To help resolve ethical conflicts, there are various ethical resources available which can help to determine which course of action to take: codes of ethics, case studies, ethics committees and personnel, and informal discussions (12). Codes of ethics are formal documents that list ethical principles and duties. Members of the profession or organization are required to adhere to the principles of these codes to guide their ethical conduct. In addition, these codes serve an educational purpose by correcting any wrong notions about ethical principles. On the other hand, case studies are often used as a reference for similar ethical conflicts and situations in the past that may have been resolved in a certain manner. These cases can be applied as jurisprudence. Organizations also can have committees and trained staff to discuss and resolve ethics issues. These may include ethics boards or ethics professionals that are contacted for consultation when ethical conflicts occur. Chats with friends or colleagues can equally lead to informal advice about how an ethical conflict can be resolved.

Codes of Ethics

For this study, we focus on codes of ethics as a means of resolving conflicts. There are various national and international bodies related to health informatics. Table 1 gives a summary of health informatics bodies extracted primarily from the Health Informatics World Wide (HIWW) database (13). In addition, membership to regional and international bodies is indicated, such as the International Medical Informatics Association (IMIA) (14), the European Federation for Medical Informatics (EFMI), the Asia-Pacific Association for Medical Informatics (APAMI), and the Health Informatics in Africa (HELINA) group (15–17)□. It can be noted that many bodies do not have their own ethical codes.

Table 1. Health informatics organizations. ○ indicates has stated code of ethics on website ♦ IMIA member ◊ EFMI member ■ APAMI member □ HELINA member. Countries and regions are identified using ISO Alpha-3 codes (18).

International		North America		Europe		South & Central America		Africa		Middle East		Asia Pacific	
001	IMIA ○	USA	AMIA ○♦	PRT	APIM	ARG	AAIM ○	MWI	MIAM □♦	SAU	SAHI ♦	CHN	CMIA ♦
009	ACHI ○♦	USA	AHIMA ○◊	BIH	BHSMI ◊♦	CHL	ACHISA ♦	ZAF	SAHIA □♦	IRN	IrMIA ♦	PAK	eHAP ♦
142	APAMI ♦	CAN	COACH ○	BEL	BMIA ◊	VEN	AVIS ♦	MAR	SMIMS			NZL	HINZ ♦
150	EFMI ♦	CUB	SOCIM ♦	DNK	DSMI ◊	BRA	SBIS ♦	MLI	SOMIBS □♦			LKA	HISSL ■
002	HELINA ♦			NOR	FDH ◊	URY	SUIS ♦	NGA	AHIN			IND	IAMI ♦
826	UKCHIP ○			FIN	FinnSHIA ◊	MEX	AMIM ♦	CMR	CAHIS ♦			JPN	JAMI ■♦
001	iHC ○			DEU	GMDS ◊							KOR	KOSMI ■♦
				IRL	HISI ◊♦							MYS	MHIA ■♦
				ROU	RSMI ◊♦							PHL	PMIS ■♦
				SVN	SDMI ◊							SGP	AMBIS ■♦
				ESP	SEIS ◊							THA	TMI ♦
				SWE	SFMI ◊							AUS	HISA ♦
				CHE	SSMI ◊							HKG	HKSMI ♦
				TUR	TurkMIA ◊							TWN	TAMI ♦
				GBR	BCS ♦○								
				HRV	CSMI ♦								
				FRA	AIM ♦								
				GRC	GHIA ♦								
				ITA	AIIM ♦								

International Codes

Within the international codes, the International Medical Informatics Association (IMIA) code of ethics is comprehensive and covers duties of health informatics professionals from three perspectives: fundamental ethics principles, informatics ethics principles, and rules of ethical conduct in health informatics. We also use the eHealth code of ethics by the iHealth Coalition (iHC), which is specifically for health-related websites (19). The iHealth Coalition code was also referenced by the World Health Organization (WHO).

North American Codes

North American codes of ethics include the "COACH high-level ethical principles", published by Canada's Health Informatics Association and an abridged version containing ten aspirational high-level principles is available to the public (20). The American Health Information Management Association (AHIMA) code of ethics contains eleven ethical obligations expected of AHIMA members (21). Another U.S. code, by the American Medical Informatics Association (AMIA) is a relatively recent publication. The AMIA code references the AHIMA and IMIA codes and focuses more on duties of health informatics professionals expected towards key stakeholders in healthcare (22).

European Codes

In Europe, the British Computer Society (BCS) code of ethics and IMIA codes have shared authorship by Dr. Eike-Henner W. Kluge, and the BCS code contains the same material as the IMIA code (23). The BCS code is acknowledged by the IMIA as an accompanying handbook (22). Similarly, the UK Council for Health Informatics Professions (UKCHIP) code of conduct is adapted from the BCS code (24). The European Federation for Medical Informatics (EFMI) does not explicitly state any code, but is a member of the IMIA (16).

Other Regional Codes

Other regional codes include the Australasian College of Health Informatics (ACHI), which provides a code of conduct based on the underlying principles of the IMIA, but is identical in wording to the UKCHIP version (25). Similarly, the Argentine Association of Medical Informatics, AAIM provides a code of ethics that is a translation of the IMIA code (26). Bodies in the Asia-Pacific, Middle East and Africa region do not provide a separate code of ethics but are regional or international members (15–17). Notably, the Medical Informatics Association of Portugal, APIM does not state any codes and are also not a member of any other body (27,28). Figure 2 shows how some of the codes are related in terms of inter-referencing one another.
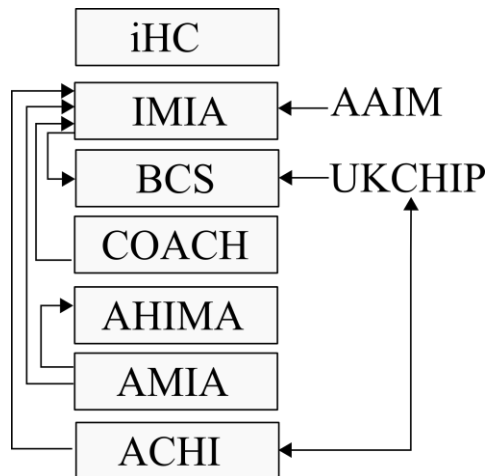
Figure 2. Inter-references between health informatics codes of ethics

Technical Standards

Aside from a set of principles that should be adhered to, health informatics ethics can also be defined as a set of activities. Ethical conflicts may be preventable even before they happen if these activities are followed during development of health informatics systems. The U.S. Food and Drug Administration (FDA) provides a document that aims to minimize software errors and failures through life cycle and risk management. In addition, the International Standards Organization (ISO) has a large collection of standards on health informatics set forth by its Technical Committee ISO/TC 215. We analyze these technical guidelines to identify how these can be applied to ethical issues.

Food and Drug Administration (FDA)

The FDA document titled "General principles of software validation" was created to reduce software defects and recalls resulting from these defects. The document applies to medical software or devices containing software. Its goals include reducing risk to patients and users, and also reduce liability to manufacturers of medical software. The FDA document achieves this through providing explicit guidelines on how to validate software during the stages of software development (29).

Development of health informatics systems involves going through the software development life-cycle (SDLC). In general, the software development life-cycle consists of all the stages during the development of software. There are various SDLC models based on different software engineering paradigms. The FDA document does not conform to any one model, but provides the following generic SDLC pathway: quality planning, requirements, design, coding, testing by software developer, user site testing, and maintenance (29). Quality planning involves tasks and procedures for reporting risks anomalies in software, assigning them to personnel, and resolving them. Requirements and information about the device or software and its intended use are identified, including proper documentation of the software or device functions. These requirements are translated into a logical and physical representation of the software through design activities. In addition, possible usage errors need to have appropriate testing measures. Ultimately, software is constructed by either coding or by assembling coded components. This is the culmination of the design into a medical device or software product. Depending on the coding guidelines being opted for by the developers, there needs to be verification of this

compliance. Once software is complete, it needs to be tested. This involves running the software with pre-defined inputs and known outputs in order to compare its accuracy. Even though testing cannot be exhaustive, all efforts need to be taken for thorough testing of the code. Another level of testing involves testing on-site with the user with realistic data. Tests are needed to determine if the user correctly understands usage of the interface. Finally, maintenance covers both installation and deployment of the software, as well as future changes to software that will require re-validation.

Moreover, the FDA document provides a template of tasks and activities for each stage of the SDLC, depending on the level of risk associated with the device or software. This validation process involves providing documentation of plans and procedures for risk management during each of these stages of the SDLC. Needless to say, items that are more risky go through more activities before approval. Eventually, the plans, procedures, and documented activities need to be evaluated and approved by the FDA. The FDA recommends independent or third-party evaluation, and also self-evaluation before final submission to the FDA.

In the SDLC, the specific tasks recommended by the FDA for validation depend on the level of concern, which is an estimate of the danger a device or software poses. There are three levels of concern detailed in the FDA supplementary document titled "Guidance for the content of premarket submissions for software contained in medical devices": major, moderate, minor (30). The level of concern for a given software is determined by answering a set of key questions. Major concerns involve possibility of death or serious injury to patient or operator. Moderate concerns have a possibility of minor, non-fatal injury to patient or operator, while minor concerns are almost unlikely to cause any injury to patient or operator.

The emphasis on validation in the FDA document facilitates the eventual production of software that reduces software faults. In other words, it leads to the development of software that mitigates harm and upholds the ethical principle of non-malfeasance. However, other principles such as privacy are not addressed.

International Standards Organization (ISO)
Another set of technical standards by the International Standards Organization (ISO), specifically addresses health informatics via the standards and guidelines developed by the ISO/TC 215 committee. The ISO standards in general are globally recognized and the TC 215 committee has international representation in its membership (31). The ISO/TC 215 standards address a wide range of issues in healthcare, and at the time of this publication, we looked at eighty-two active standards on different areas in health informatics. Three main categorizations can be presented based on the artifacts addressed by these standards: health software, health devices, and health information. Majority of these standards focus on communication protocols between devices and computers. The standards also aim to promote common vocabularies, address EHR harmonization, and facilitate health informatics systems. Figure 3 summarizes most of the currently active ISO/TC 215 standards categorized by artifacts and sub-themes. In addition, many of these standards are ethics-aware, and specifically address some ethics issues directly or indirectly. We identify four matches to the IMIA ethics principles that are covered via the ISO/TC 215 standards: non-malfeasance, privacy, security, and accountability.

Information	Vocabulary	ISO 17115	ISO TR 12309	ISO TR 25257			
		ISO 18104	ISO TS 17117	ISO TS 22224			
		ISO 21667	ISO TS 22789				
		ISO 25720	ISO TS 25237				
	EHR	ISO 13606-1	ISO 18308	ISO TS 13606-4	ISO 20301	ISO 21549-3	ISO 21549-6
		ISO 13606-2	ISO HL7 21731	ISO TS 21547	ISO 20302	ISO 21549-4	ISO 21549-8
		ISO 13606-3	ISO TR 20514	ISO TS 27527	ISO 21549-1	ISO 21549-5	
		ISO 13606-5	ISO TR 21548	ISO TS 27790	ISO 21549-2	ISO 21549-7	
Software	HI systems	ISO 27799	ISO 12967-1	ISO TR 11487	ISO TR 17119	ISO TS 29585	
		ISO TS 21091	ISO 12967-2	ISO TS 21298	ISO TS 22600-1	ISO TR 22221	
		ISO TR 11633-1	ISO 12967-3	ISO TS 22220	ISO TS 22600-3	ISO TR 27809	
		ISO TR 11633-2	ISO 18812	ISO TR 22790	ISO TS 22600-2	ISO TS 25238	
	Data Comm	ISO 11073-91064	ISO TR 21089	ISO TR 11636	ISO TR 16056-1		
		ISO 12052	ISO 17090-1	ISO TR 21730	ISO TR 16056-2		
		ISO 17432	ISO 17090-2	ISO TS 16058	ISO TR 18307		
		ISO 18232	ISO 17090-3	ISO 21090			
Device	Personal	ISO 11073-10404	ISO 11073-10471				
		ISO 11073-10407	ISO 11073-20601				
		ISO 11073-10408	ISO 22857				
		ISO 11073-10415					
	Point-of-care	ISO 11073-20101	ISO ISO/IEEE 11073-10101				
		ISO 11073-30200	ISO ISO/IEEE 11073-10201				
		ISO 11073-30300	ISO TS 11073-92001				
		ISO 11073-90101					

Figure 3. Overview of the ISO/TC 215 standards

There are two main standards that address safety: ISO/TR 2 7809 and ISO/TS 25238. While similar to the FDA approach of risk assessment within the SDLC, the ISO/TC 215 standards more comprehensively address risk assessment. While acknowledging and referencing the FDA standards, the ISO/TC 215 propose five classes for risk assessment instead of the FDA's three categories: catastrophic, major, considerable, significant, and minor (32,33). Catastrophic risks could involve multiple deaths, or injuries causing disability, while major risks could include death of an individual patient, or severe injuries to multiple patients. Considerable risks include severe injury for an individual patient or minor injuries for multiple patients, while significant risks could include minor injuries for an individual patient, or minor shock for multiple patients. Finally, minor risks could induce mild shock for an individual patient.

Additionally, the ISO/TC 215 include an assessment of the likelihood of failure of a device or software: very high, high, medium, low, very low. Very high likelihood implies failure most likely will occur, while high likelihood means failure is uncertain but very likely to occur. Medium failures are likely to occur, but not certain, while low failures only have a small chance of occurrence. Very low categories would most likely not occur.

There are various standards in the ISO/TC 215 catalog that deal with information, and specifically patient information in the form of Electronic Health Records (EHRs). ISO 22857 looks at transfer of personal health information generally, and the consequent privacy risks (34). In addition, the standards on EHR communication, archiving, and transmission such as ISO13606-1 and ISO/TS 21547 acknowledge the importance of privacy. Closely related to the need for privacy is its realization via security and access control mechanisms (35,36). Security is addressed for communication, transmission, and storage of data.

There are various standards for personal health devices and point-of-care health devices that look at communication between these devices and computer systems. Additionally, security of EHR communication and storage is also covered. ISO/TS 22600-1, ISO/TS 22600-2, and ISO/TS 22600-3 are stipulations on how security is managed via access control. Access control covers 4 main aspects: authentication, access, authorization, and function (37). With authentication, the identity of the user is established. Next, there is a policy about what information aspects are allowed access. Authorization gives a mapping between what can be accessed and who has access. Functionally, the possible actions that can be taken upon authorization are known. In addition to controlling access to data, security also covers data transfer and secure exchange. ISO 17090-1, ISO 17090-2, ISO 17090-3 look at digital certificates, while ISO TR 11633-1 and ISO TR 11633-2 look at remote access. Also, ISO TR 21089 looks at the role of trust in access control (31).

Various standards with the ISO/TC 215 address the need for accountability via provisioning for audit trails. ISO/TS 13606-4 and 13606-1 address security specifically for EHRs, and include an audit log in its guidelines. In addition, the term "auditability" is introduced in 13606-4, and defined as the ability to examine actions of users (36). ISO 22857 stipulates that audit trails should be tamper-proof (34). In general, depending on the nature of the artifact, whether health software, device, or information, the appropriate ISO/TC 215 standards can be applied. The risks of ethics problems arising are thereby reduced because these standards implicitly or explicitly incorporate ethics aspects.

Findability

Findability is a well-known concept in the area of information architecture, and means the ease of locating information on a website (38). Findability quantifies the ease of accessing and finding specific content by an information consumer. There are two viewpoints of findability in general: the visibility of a website from the internet via search engines, or visibility of specific content on the website itself. For the development of the health informatics ethics repository, we focus on the latter of these viewpoints. The motivation behind this focus is because of the limited search capabilities of websites hosting the health informatics ethics documents. Also, the documents are hosted independent of each other, and findability of content is even further restricted. For instance, a domain researcher may be interested in getting information and references on privacy within the codes, but would be restricted to finding these references per website. In other words, cross-cutting findability of the content is limited.

There is no unified point of access to the vast amount of codes and standards for health informatics ethics, and even existing websites that host these documents have limited search capabilities. All of this makes it cumbersome for researchers in this field to comparatively study the various documents. Our contribution is compilation and unified access to these documents through a repository built using using state-of-the-art search technologies.

Harmonization of Ethics Codes and Standards

Health informatics ethics deals with questions of wrong and right resulting from the use of technology in managing healthcare information. It involves adherence to general, informatics, and software engineering ethics principles. In the preceding sections, we presented health informatics ethics as a set of activities that are incorporated into the SDLC. In our definition,

health informatics ethics as an activity involves software validation and best practices, in accordance with the ISO/TC 215 and FDA guidelines. Whereas the codes provide guidelines on 'what' is ethical, the ISO/TC 215 and FDA documents define 'how' to directly or indirectly make it ethical. Figure 4 shows the different ethical principles and which of these are directly or indirectly addressed by activities through the ISO/TC 215 and FDA technical guidelines.
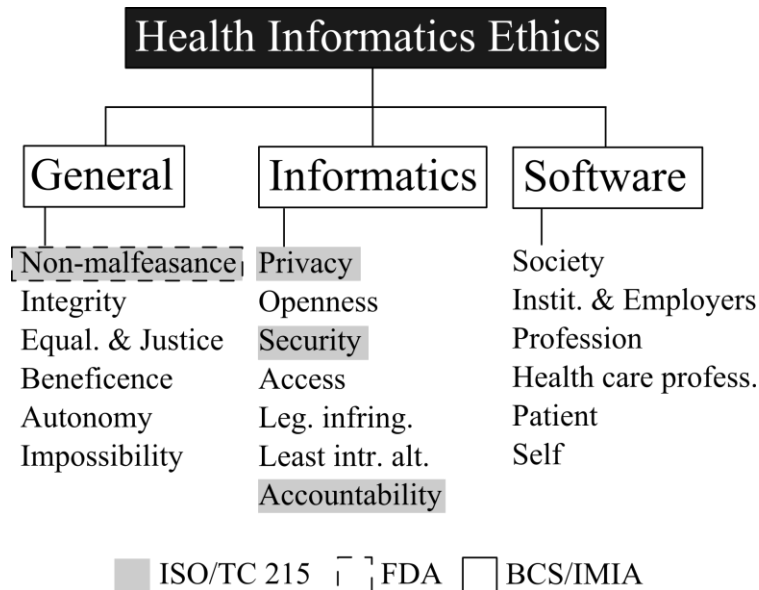

Figure 4. Ethical codes and technical guidelines in health informatics

Health Informatics Ethics Repository

Our survey reiterates a large diversity of literature and sources pertaining to health informatics ethics. These sources may not always be easy to access and comparatively analyze. As our contribution to the research community in this domain, we developed a repository that indexes the major codes of ethics and standards related to health informatics that are freely available on the internet. This repository contains codes of ethics from the IMIA, BCS, iHC, ACHI, UKCHIP, and AHIMA. The system is built using state-of-the-art search technologies to make it easier for researchers to find and compare items of interest across codes. The repository can be freely accessed online on the World Wide Web.

Indexing

One important step in the creation of the health informatics ethics repository is indexing, which involves parsing and extracting the contents the codes and standards, and saving the contents to a database designed for fast retrieval. In the field of information retrieval, indexing converts unstructured text to a structured format (39,40). Contents in the health informatics ethics repository are stored as an inverted index data structure. The inverted index structure allows storing mappings from words in the codes and standards to the source documents. Each word is stored uniquely, but can have multiple mappings to different source documents. For example, the word "privacy" would occur in multiple codes and standards, but would have only one instance in the inverted index.

Content Life-Cycle

In order to properly index the relevant documents, we have guidelines in place for the content life-cycle (41). Each indexed document has three meta-data properties to facilitate awareness of the document's version to the readers, as well as to the maintainers of the repository: publication year, source, and date indexed. Publication year is the reported year of the current document's publication, as reported by the publisher/organization. The source is a link to the original document on the Internet, as being hosted by the publisher/organization of the code or standard. The date indexed property is the date the document was last indexed.

To prevent outdated documents, the date of publication will be checked semi-annually against the organization's latest version. If the dates are the same, then no further re-indexing will be done. In the case a new version or edition of the code or standard has been released, that document will be re-indexed.

Comparative Searching

Our consolidated repository has various enhancements to facilitate comparatively searching across different codes of ethics. Researchers can enter a search queries and see its occurrence visually highlighted across all the indexed codes and standards. Documents can also be hidden if they are not of interest. Researchers can also navigate from one occurrence to the next within a particular document via the search result navigation hyperlinks. Moreover, we incorporate standard query expansion techniques such as stemming and fuzzy searching to expand user search queries. These expanded queries enables discovery and comparative analysis using words and corrections related to the search queries.

Stemming

In everyday usage of the English language, words have different forms. For instance, a word has a singular and plural form, or a referential form, and so on. The word 'privacy' can have related forms such as 'private', while the word 'publish' can have other forms such as 'publishes', 'published', 'publishing'. The job of stemming and lemmatisation algorithms is to determine the other relevant forms of a given word (39). For instance, one form of the Porter stemming algorithm works by stripping all suffix or prefix matches from a word. What remains is an approximation of the etymological root of that word. This can then be iterated to create new forms. As an example, the word 'private' leads to 'priv' as an approximate root, because '-ate' is a common suffix. Consequently, 'priv-acy' can be related to 'private' by adding other suffices to the root. In this fashion, if a researcher searches for the query 'private', the search results would also contain matches for 'privacy'.

Fuzzy Search

This is a technique for finding approximate string matches to a given pattern (42). The most common application for this technique is to detect user spelling errors and suggest corrections. For instance, if the search query is typed in as 'haert', a fuzzy search algorithm would be able to suggest 'heart' as a possible alternative to use. A simple approach to fuzzy search is the edit distance algorithm, also referred to as the Levenshtein distance (42). This distance is calculated based on three main single-character operations that can be performed to convert one string to another, namely, insertion, deletion, and substitution. For example, to change 'haert' to 'heart' two substitutions are needed. More generally, the transformed word needs to be checked for

validity against some language dictionary to avoid coming up with meaningless word suggestions.

Search Interface Demo

Figure 5 shows a demonstration of the health informatics repository's search interface. The end-user can control which documents are being shown, and can view various documents side-by-side. Also, when searching for keywords, the interface highlights them in the documents and provides a navigation panel to go through all occurrences of the search matches. The interface also allows parallel reading of multiple codes and standards.
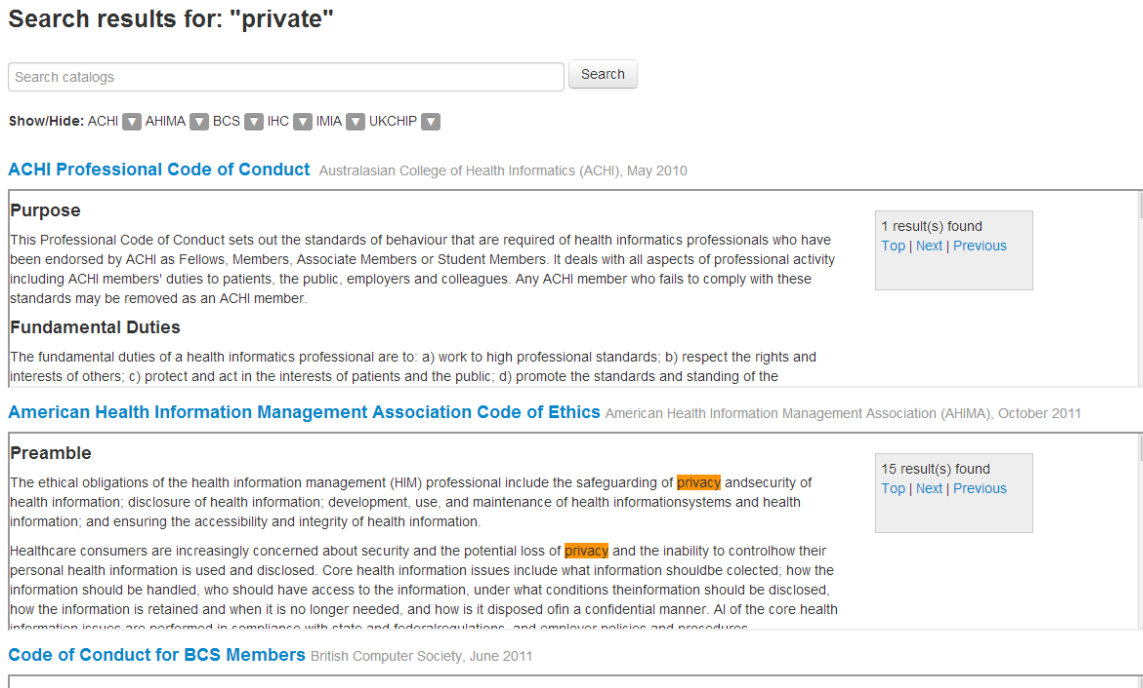

Figure 5. Comparative and enhanced search available through the health informatics ethics repository

Based on our selection criteria, the repository contains codes of ethics from the IMIA, BCS, iHC, ACHI, UKCHIP, and AHIMA. Codes of ethics from other organizations were not indexed because these organized either did not have a code or were referring to international or regional codes. Also, to maintain intellectual property rights, we indexed documents freely available on the internet. This excluded the ISO/TC 215 standards. We focused solely on codes of ethics presently, and will be looking into indexing technical standards in future versions of the repository.

Conclusion and Future Work

In this study, we comprehensively examined codes of ethics from major international, regional, and national policy drivers and stakeholders such as the IMIA, BCS, iHC, ACHI, UKCHIP, and AHIMA. We also investigated how the ISO and FDA guidelines are applied to management of software, devices, and data. Ultimately, we presented a searchable repository of codes of ethics and standards in health informatics built using state-of-the-art search algorithms and technologies, which enhanced the findability of codes and standards related to health informatics

ethics. The repository platform will potentially benefit public health practitioners, researchers, and software developers as a one-stop reference for various ethics codes and standards in health informatics. Our contributions included a systematic review and comparative analysis of various health informatics ethics codes, affiliations, and practices from different bodies around the world. Also, as part of our support for the research community in this domain, we introduced a repository with state-of-the-art comparative searching for freely available codes of ethics and standards related to health informatics. For future work, we will explore new search, comparison and feedback features for the repository and plan to index new health informatics codes, as well as technical standards.

Acknowledgments

Conflicts of Interest

The authors declare that they are not affiliated of any of the health informatics organizations mentioned in this paper. Furthermore, all authors are not affiliated with the Food and Drug Administration (FDA) or the International Standards Organization (ISO). However, the authors are members of the Association for Computing Machinery (ACM). Publications from this professional organization have been used in our research.

References

1. Thomson AJ, Schmoldt DL. Ethics in Computer Software Design and Development. Comput Electron Agric. 2001;30:85–102.

2. Trzęsicki K. Medical Informatics Ethics and its Subject. Ann Acad Medicae Bialostoc. 2005;50(2):3.

3. Trzęsicki K. Medical Informatics Ethics (Subject and Major Issues). Stud Health Technol Inform. 2005;8(21):18.

4. Anderson JG, Goodman KW. Ethics and Information Technology: A Case-Based Approach to a Health Care System in Transition. Springer New York; 2002.

5. Principles of Medical Ethics [Internet]. United Nations (UN). 1982. Available from: https://www.un.org/documents/ga/res/37/a37r194.htm

6. AMA's Code of Medical Ethics [Internet]. American Medical Association (AMA). Available from: http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page

7. Samuel HW, Zaïane OR, Sobsey D. Towards a Definition of Health Informatics Ethics. Proceeding of 1st ACM International Health Informatics Symposium. 2010. p. 257–64.

8. The IMIA Code of Ethics for Health Information Professionals [Internet]. International Medical Informatics Association (IMIA). 2011. Available from: http://www.imia-medinfo.org/new2/node/39

9. Court of Justice of the European Union. Judgment in Case C-131/12 Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González. CURIA;

10. Gotterbarn D. Software Engineering Ethics. Encyclopedia of Software Engineering. 2001.

11. Anderson RE, Johnson DG, Gotterbarn D, Perolle J. Using the New ACM Code of Ethics in Decision Making. Commun ACM. 1993;36(2):98–107.

12. Nelson W, Rosenberg M-C, Weiss J, Goodrich M. New Hampshire Critical Access Hospitals: CEOs' Report on Ethical Challenges. J Healthc Manag. 2009;54(4):273–83.

13. Health Informatics World Wide (HIWW) [Internet]. Available from: http://www.hiww.org

14. International Medical Informatics Association (IMIA). Member Societies [Internet]. Available from: http://www.imia-medinfo.org/new2/member_societies

15. Health Informatics in Africa (HELINA) [Internet]. Available from: http://www.helina-online.org

16. European Federation for Medical Informatics (EFMI) [Internet]. Available from: http://www.efmi.org

17. Asia-Pacific Association for Medical Informatics (APAMI) [Internet]. Available from: http://www.apami.org

18. Standard Country or Area Codes for Statistical Use [Internet]. United Nations (UN). Available from: http://unstats.un.org/unsd/methods/m49/m49.htm

19. Rippen H, Risk A. eHealth Code of Ethics. J Med Internet Res. 2000;(May):1–9.

20. Gabor N, MacGregor J, Murray M. Ethical Principles: A Step Toward "Professionalism" for COACH. COACH: Canada's Health Informatics Association. COACH: Canada's Health Informatics Association; 2006. p. 33–5.

21. American Health Information Management Association (AHIMA). AHIMA Code of Ethics [Internet]. Available from: http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_024277.hcsp?dDocName=bok1_024277

22. Hurdle JF, Adams S, Brokel J, Chang B, Embi PJ, Petersen C, et al. A Code of Professional Ethical Conduct for the American Medical Informatics Association: An AMIA Board of Directors Approved White Paper. J Am Med Informatics Assoc. 2007;14(4):391–3.

23. Committee BCSHI. A Handbook of Ethics for Health Informatics Professionals. 2003.

24. UKCHIP Code of Conduct [Internet]. UKCHIP. Available from: http://www.ukchip.org/?page_id=1607

25. Australasian College of Health Informatics (ACHI) [Internet]. Available from: http://www.achi.org.au

26. Argentine Association of Medical Informatics, AAIM [Internet]. Available from: http://www.imia-medinfo.org/new2/node/57

27. Association for Health Informatics of Nigeria (AHIN) [Internet]. Available from: http://ahin.wordpress.com/

28. Portuguese Association of Medical Informatics, APIM [Internet]. Available from: http://apim.med.up.pt

29. General Principles of Software Validation; Final Guidance for Industry and FDA Staff. U.S. Food and Drug Administration (FDA). U.S. Department Of Health and Human Services, Food and Drug Administration (FDA); 2002.

30. Guidance for the Content of Premarket Submissions for Software Contained in Medical Devices. U.S. Food and Drug Administration (FDA). U.S. Department Of Health and Human Services, Food and Drug Administration (FDA); 2005.

31. International Standards Organization (ISO). TC 215 - Health Informatics.

32. Health Informatics - Classification of Safety Risks from Health Software, ISO/TS 25238. International Standards Organization (ISO). 2007.

33. Health Informatics - Measures for Ensuring Patient Safety of Health Software, ISO/TR 27809. International Standards Organization (ISO). 2007.

34. Health Informatics - Guidelines on Data Protection to Facilitate Trans-border Flows of Personal Health Information, ISO 22857. International Standards Organization (ISO). 2004.

35. Health Informatics - Security Requirements for Archiving of Electronic Health Records - Principles, ISO/TS 21547. International Standards Organization (ISO). 2010.

36. Health Informatics - Electronic Health Record Communication - Part 1: Reference Model, ISO 13606-1. International Standards Organization (ISO). 2008.

37. Health Informatics - Privilege Management and Access Control - Part 1: Overview and Policy Management, ISO/TS 22600-1. International Standards Organization (ISO). 2006.

38. Morville P. Ambient Findability. O'Reilly Media; 2005.

39. Manning CD, Raghavan P, Schütze H. An Introduction to Information Retrieval. Cambridge: Cambridge University Press; 2009.

40. Frakes WB, Baeza-Yates R. Information Retrieval: Data Structures and Algorithms. Prentice-Hall, Inc.; 1992 Jun 1;

41. McKeever S. Understanding Web Content Management Systems: Evolution, Lifecycle and Market. Ind Manag Data Syst. MCB UP Ltd; 2003 Jan 12;103(9):686–92.

42. Gusfield D. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press; 1997 Jul 1