Effective use of personal health records to support emergency services

Alba Catalina Morales Tirado[1[0000-0001-6984-5122]], Enrico Daga[1[0000-0002-3184-5407]] and Enrico Motta[1[0000-0003-0015-1952]]

[1] Knowledge Media Institute, The Open University, Milton Keynes, United Kingdom
{alba.morales-tirado, enrico.daga , enrico.motta}@open.ac.uk

Abstract. Smart City systems capture and exchange information with the aim to improve public services. Particularly, healthcare data could help emergency services to plan resources and make life-saving decisions. However, the delivery of healthcare information to emergency bodies must be balanced against the concerns related to citizens' privacy. Besides, emergency services face challenges in interpreting this data; the heterogeneity of sources and a large amount of information available represent a significant barrier. In this paper, we focus on a case study involving the use of personal health records to support emergency services in the context of a fire building evacuation. We propose a methodology involving a knowledge engineering approach and a common-sense knowledge base to address the problem of deriving useful information from health records and, at the same time, preserve citizens' privacy. We perform extensive experiments involving a synthetic dataset of health records and a curated gold standard to demonstrate how our approach allows us to identify vulnerable people and interpret their particular needs while avoiding the disclosure of personal information.

Keywords: Health Records, Smart City, Emergency Services, Privacy, Knowledge Engineering, ConceptNet.

1 Introduction

The Smart City paradigm has been adopted to deliver technology-driven solutions, designed and built to enhance the management of city services, such as transportation, energy and water supply, health and emergency management, among others [1]. Smart City systems are designed as distributed cyber-physical systems in which the data exchange across different enterprises is of paramount importance to the success of their proposition. Generally, in the Smart City environment data is gathered by different means and from different sources; it could be very detailed and collected in real-time. An area of application in Smart Cities pertains to the use of health information to support emergency events. Just like smart systems for traffic management can help in reducing emergency services response time [2], an intelligent healthcare system could also continuously gather physiological signs (e.g., heart rate, body temperature) from patients [3], thus making data immediately available to hospitals and emergency medical services [4]. Besides, there is a promising trend towards fast, agile access to health records, for instance, the Emergency Care Summary (ECS) system implemented by the Scotland Government [5] aims to provide patient's useful information to healthcare staff. However, this summary is accessible only under the patient express consent and does not include detailed information of the patient. For example, recent diseases, surgeries or disabilities are not part of the summary. Also, ECS is an opt-out scheme; it means that not all the patients will have an ECS if they decide not to participate.

In recent years, research highlighted significant obstacles to effective data sharing between organisations and emergency services [6]. For example, a report from the UK government referring to the emergency response to the 7 July 2005 London Bombings points out that the *"Limitation on the initial collection and subsequent sharing of data"* was due to the concerns on sharing personal data [7]. The issues related to privacy that hamper the effective reuse of data can be summarised as follows:

- Disclosure or dissemination of sensitive information (such as health conditions, disabilities, sexual orientation, location, among others).
- Use of data for purposes other than the one stated initially (such as advertising).
- The exchange/sharing of personal data with other parties (insurance companies, the government, including emergency bodies) [8].
- Breaches of regulations, such as the EU General Data Protection Regulation (GDPR) and the UK Data Protection Act, leading to unlawful personal data exchange during emergency response.

Therefore, emergency responders must assess how to handle personal data just as any other organisation [9]. However, emergency response is exceptional in nature. Let us consider the role of health records in the following scenario. In a large organisation, employees use their access cards to enter the building and visitors must register as they enter or leave the premises. A fire starts on the fourth floor of the building, and emergency services are alerted. Having information about people in the building can help emergency services. However, additional information about vulnerable people could assist emergency responders to intervene and make effective decisions promptly. Crucially, this information can be retrieved from Health Records of the national health service. However, there are two significant problems. First, a person's health record can contain a large amount of very specific information. Therefore, finding a way to detect relevant information is essential. Second, health records contain very sensitive information and, therefore, the exchange of such data constitutes a privacy violation. Preventing the disclosure of personal data while providing emergency services with usable information is an important and difficult problem [10]. In this work, we focus on the following research questions:

- RQ1: How to use health records to support emergency services to identify *who* is in need of special assistance during an evacuation?
- RQ2: How to process health records in order to derive information about *why* the person needs assistance?

To answer these questions, we propose an approach based on knowledge engineering, semantic technologies, and the use of a common-sense knowledge base (ConceptNet

[11]). First, we analyse the regulations that large organisations in the UK are required to apply in relation to vulnerable people during a fire emergency. We performed our experiments relying on a synthetic healthcare dataset, encoded using the healthcare standards, such as, the Fast Healthcare Interoperability Resources (FHIR), for the exchange of electronic health records [12] and the Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT) [13]. Next, we analyse the data schema and annotate it according to its relevance and sensitivity. From the resulting dataset, we review features related to the description of health conditions (represented with SNOMED CT) and the time-validity of data. To enable querying over the data schema, we take a Linked Data approach and use RDF to characterise the information and SPARQL to query a schema-less representation of the data source. By doing this, we achieve a significant reduction of the data points and identify the persons with current medical conditions, therefore, potentially in need of special assistance. To answer the second research question, we match the identified data points with a categorisation of different types of disabilities relevant to building evacuation, according to the governmental guidelines of the UK [14], with the aid of a common-sense knowledge base (ConceptNet). The output of our system is a list of persons requiring assistance and the reason for their needs, without disclosing sensitive information. Our contributions[1] are:

- A novel approach to developing a data pipeline that allows the use of personal health records to derive relevant information and support emergency services;
- A synthetic dataset of annotated FHIR schema elements, according to their sensitivity and utility with respect to a fire evacuation emergency;
- A gold standard dataset developed on the healthcare dataset for evaluating systems in deciding who needs assistance and the reason for it;
- Extensive experiments to demonstrate the effectiveness of our method and to define a baseline for further research on the topic.

The remainder of this paper is organised as follows. We begin by presenting the scenario analysed in this paper in Section 2. After describing the related work in Section 3, we present the proposed methodology in Section 4 and its application in Section 5. Section 6 describes the implementation of the system. In Section 7, we present the results. Finally, we discuss future work and conclusions in Section 8.

2 Scenario

We consider a fire event in a large organisation, analysing the case of The Open University in the UK. The employees use their access cards to enter the building, and visitors must register as they enter or leave the premises. As stated in the organisation's procedures, all employees should inform the Health and Safety Department (HSD) if they have a long-term condition or a temporary disability. Following this notification, the HSD must assure that each employee has an emergency evacuation plan tailored to

[1] https://doi.org/10.5281/ZENODO.3862336

their needs. To record the evacuation plan, the HSD follows governmental guidelines and internal regulations.

Generally, a person designated by the HSD interviews the employee and evaluates his/her capacity to perform the plan. Typically, factors to take into account and negotiate a suitable evacuation plan are a) type of disability, b) the employee's capacity to perform a plan, and c) the means of escape available in the building. For the elaboration of a plan, the UK governmental guidelines [14] provide a comprehensive list of disabilities and recommended options for escape as well as important guidance for assessing and arranging the appropriate means of evacuation for the employee. Once identified the type of assistance required, the following action is to register a tailored Personal Emergency Evacuation Plan (PEEP). The PEEP collects employee's identity information (e.g., name, telephone number, staff number), a description of the conditions and disabilities that may affect his/her ability to evacuate the building, according to the categories identified in the governmental guidelines. Additionally, a step by step description of the actions an employee must perform in case of evacuation, as well as any aid equipment or assistance needed. This information must be shared with the nominated *fire wardens*.

A fire starts on the fourth floor of a building, it is spreading quickly, and emergency services are alerted. The HSD may be able to identify people with special needs by retrieving the PEEP record. However, a number of issues reduce the effectiveness of this approach. In the absence of digital infrastructure, PEEP files may be impossible to retrieve efficiently. But also, in the case of a database, the completeness and accuracy of the data are questionable. Compiling the PEEP requires the sharing of health information that could be considered very sensitive by employees. Many people may not want to share this type of information with the line manager or the colleagues appointed as fire wardens. For example, anxiety or other mental health conditions can be typically hard to disclose. In addition, the information included in the PEEP may be outdated. Crucially, visitors may not be included in the records. Having precise information about vulnerable people could help emergency services react promptly and take the right decisions when planning resources. In this context, accessing the health records of the National Health Service (NHS) by a Smart City system constitutes a substantial opportunity to retrieve up to date information and recognise accurately people requiring support. However, obtaining such amount of fine-grained and specialised data could be overwhelming for firefighters and fire wardens because:

- Healthcare data is highly specialised and may be difficult to interpret by the personnel involved in supporting the evacuation (e.g., firefighters).
- A large amount of data makes it difficult to find relevant information.
- Exchange of sensitive information might put citizens' privacy at risk.

Therefore, it is imperative to find a solution that can access healthcare data, filter out the relevant information and process it to deliver meaningful, *fit for purpose* summaries, while preserving citizens' privacy. In principle, an Intelligent System could act as a mediator between the healthcare data provider and the emergency services to balance the trade-off between utility and sensitivity.

3 Related work

We consider related work in the areas of intelligent systems for emergency response, with particular attention to the use of healthcare data. A considerable amount of literature has been published on the use of healthcare data and personal health records [15, 16] in Smart Cities. In recent years, attention has also focused on the use of health records to assist emergency services [17]. For instance, solutions facilitating confidential access to health records during emergency events [18]. Another example is the implementation of smart home solutions that monitor the elderly's health and provide emergency services with accurate information [3].

The information represents a life-saving resource for first responders, who require data that is well structured, pertinent to their needs and readily available [19]. In this context, information management to support emergency services embraces diverse approaches, for example, decision support systems for data integration and utilisation of provenance data [20], and the use of semantics for the integration of heterogeneous knowledge [21].

Although these works focus on supporting emergency services by using healthcare-related data, they only focus on solving issues related to heterogeneous semantic data integration and organisation. For instance, in [22], the authors propose a similar scenario of a fire evacuation in a University, for which they developed a solution that queries different data sources (such as, an employee management system). The system allows emergency responders access to fire event-related information (hazardous materials, building information, among others) and employees' 'medical status' which is limited to indicate whether an employee has or not a disability. However, they do not consider the use of healthcare data or provide details about the type of disability. In our work, we focus on using health records and rely on the use of semantic web technologies to extract relevant information facilitating the interpretation of health data while providing emergency services details about vulnerable people and the type of assistance required.

Studies also raise privacy concerns when using healthcare data [23–25]. Research to date presents different approaches to tackle these concerns; for instance, a proposed framework for deriving security and privacy requirements [24]. Other solutions propose protocols to enable anonymous data exchange between stakeholders in cloud environments [25]. To the best of our knowledge, none of these approaches undertake the problem of optimising the trade-off between sensitivity and utility while accessing health records during emergency events, hence minimise data sensitivity before it is exchanged.

Several studies apply Knowledge Graphs as a solution for heterogeneous data integration in domains such as disaster management [20, 26] and health monitoring [21, 27]. In our work, Semantic Web technologies are used as part of the approach to managing healthcare data, in particular, representing a synthetic healthcare record dataset [28]. As healthcare records are increasingly becoming digitised, we use FHIR to structure and standardise its content. Our proposed solution uses annotations to identify relevant and sensitive data within the health records dataset. To make these resources available, we use RDF which addresses the requirements to perform meta-queries over the schema, using the utility and sensitivity annotations.

In summary, healthcare data is undoubtedly a paramount source of information for emergency services [17, 18]. Different applications make use of healthcare data implementing approaches concentrated on semantic integration of heterogeneous data sources [15, 16, 22]. Although integration is relevant, it is equally important to devise a method which extracts meaningful data to meet emergency services' requirements. Ultimately it will lead to exchange only useful data, minimising the amount of personal information and protecting citizens' privacy.

4 Methodology

As stated in Section 1, our research focuses on assisting emergency services to make use of healthcare data while preserving citizens' privacy. This methodology has the aim of supporting a data engineer in developing a privacy-aware pipeline for effective reuse of health records. Therefore, the methodology (**Fig. 1**) is generic and portable across similar scenarios where data sources contain extremely sensitive data. Throughout this paper, the term 'datapoint' refers to the smallest piece of information, and it is associated with one or more dimensions of the data schema—for example, a cell in a spreadsheet.

The first activity of our methodology is to identify the data requirements according to the emergency. The task is to represent the knowledge requirement in terms of a closed Competency Question (CQ), therefore formalising the information needs and facilitating the identification and extraction of required data. For instance, if the CQ is "Would the person be able to manipulate small objects?", then the relevant information constitutes the attributes that answer this question (e.g., conditions and procedures related to upper limbs).

Step 2 concentrates on identifying structured or unstructured data sources that could help to answer the CQ. These may include observations as well as reference taxonomies or domain ontologies.

Step 3 performs an exhaustive analysis and annotation of the data sources. The objective is to have a clear understanding of the role that each one of them may have in the pipeline. For example, analysing the content, its data schema (attributes and relationships) and identifying criteria for selecting useful information (for example, filtering out outdated information). Specifically, we inspect the data schema and annotate the properties according to two dimensions: *utility and sensitivity*. By assigning annotations, it is possible now to filter the data points that do are not useful to answer the CQ. The final result is a reduced collection of health records, leading to the extraction of relevant data.

The fourth Step in the methodology takes as input the subset of data points resulting from the analysis in Step 3. Data identified as useful, but not sensitive can be exchanged or used directly. On the other hand, data considered somehow sensitive has to be processed to reduce its degree of sensitivity. Building on these considerations, in Step 4 the objective is for the data engineer to transform the data, by applying privacy-preserving techniques, for example, using standard classification systems in substitution of the specific data point. The final output is a set of tailored information that satisfies the CQ, hence enabling effective use of personal health records.

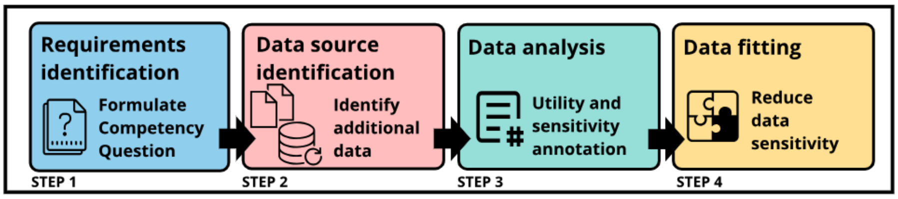

Fig. 1. Step by step methodology

5 Application of the methodology

To illustrate the use of the methodology, we use the scenario set up from Section 2; we focus on a fire event in a large organisation and use a healthcare data source.

5.1 Requirements identification

From the scenario in Section 2, we define the knowledge requirements of the emergency services. Specifically, firefighters need to be informed if any of the occupants of the building need assistance evacuating the premises and the type of support they may provide. Following our methodology, we start by formulating the CQ:

Who among the occupants in the building is in need of assistance in case of a fire evacuation, and what type of need?

5.2 Data source identification

To satisfy the need for information, we review the CQ formulated previously and identify specific or additional information that could contribute to answering the CQ. Specifically, the notification of a disability or a temporary condition triggers the creation of a PEEP. Therefore, we explore data sources and regulations about diseases and chronic conditions as well as types of assistance or disabilities regarding an impediment to performing an evacuation plan. About disabilities and risk for disabled people, the guidelines of the UK Government [14] provide a comprehensive list of types of disabilities and means of escape for people with special needs. Our synthetic dataset of health records uses SNOMED CT, a standard terminology for clinical content in electronic health records. Additionally, the NHS website[2] is a useful source of information for non-experts; it provides details about the impact of diseases and recovery times.

[2] https://www.nhs.uk/conditions/

Health dataset. Health-related data is considered highly sensitive information. Hence, to prevent any disclosure of private information, in our research we make use of synthetic healthcare data. Synthea [28] is an open-source software that generates synthetic electronic health records. The software models the medical history of synthetic patients. The health record of each patient is generated independently and simulates the health registers from birth to death through modular representations of various diseases. The synthetic electronic health records are deep and extensive as they include demographic data, appointments, patient conditions, procedures, care plans, medication, allergies, and observations. We decided to represent the data using FHIR (Fast Healthcare Interoperability Resources) standard specification for exchanging healthcare information electronically. Specifically, we took a Linked Data approach and used the FHIR RDF ontology as the most suitable option to describe the dataset. From the generated dataset, we identify 155 attributes grouped into 14 different types of information, and we focus on these in the subsequent analysis.

5.3 Data Analysis

This step of the methodology is dedicated to analysing the content of the dataset, its data schema and to the production of annotations related to its sensitivity and utility.

We inspect the content of our health record dataset and its data schema to reason over the features of the data that can help to answer the CQ. Crucially, we want to distinguish data features that are useful and their degree of sensitivity. Particularly, we observe that data points describing health conditions have a temporal validity, since a condition may be valid for a specific amount of time. Therefore, we produced a set of annotations to identify the temporal validity of SNOMED CT codes. Separately, two of the authors annotated the time validity of the 417 SNOMED CT codes of our synthetic health dataset, using as support the NHS public information. For example, the NHS web page specifies that 'Pneumonia' disease may require six months to recover; thus, the annotations include the stated recovery time and the source of information. Next, they discussed each annotation and agreed on the time validity representation, including comments to describe the condition where possible.

Utility and sensitivity annotation. First, we define two custom RDF predicates that represent *utility* and *sensitivity*. Then, we annotate each attribute manually according to its utility to answer the CQ. For the sensitivity assessment, the task is to identify the attributes in the data schema that are considered personal data. To annotate attributes as sensitive, we use regulations that govern personal data (for instance, GDPR, data protection act) and define personal data and its impact on privacy. The result is a dataset annotated according to its utility and sensitivity. Applying a Linked Data approach allows us to perform *meta-queries* in SPARQL hence making it easy to filter data points by means of the annotations on their properties. For example, one data point of the health record could be represented in the following query (see **Fig. 2**). By using the annotations, we extract relevant data points and distinguish sensitive information. As shown in **Table 1**, this already translates into a significant reduction of the data to be processed. However, not all relevant data points describe a health condition nor a disability useful to answer our CQ. For instance, a data point describing an appointment, or a general procedure may not be relevant. Instead, one describing a recent fracture of the ankle certainly will. Thus, the next task is to derive the valid data points from the subset of relevant information.

property	subproperty	value	utility	sensitivity
fhir:Condition.codeableConcept	fhir:CodeableConcept.coding	58150001	epront:useful	epront:sensitive
fhir:Condition.codeableConcept	fhir:CodeableConcept.text	Fracture of clavicle	epront:useful	epront:sensitive
fhir:Condition.period	fhir:Period.start	2018-03-12T00:00:00.000Z	epront:useful	epront:sensitive
fhir:Condition.period	fhir:Period.end	2018-09-12T00:00:00.000Z	epront:useful	epront:sensitive
snomed:Validity.months		6	epront:useful	epront:sensitive

Fig. 2. Data point representation

Table 1. Reduction of data to be processed. One-person example

	All data points	Only relevant data points	Relevant & sensitive
# of data points	32,608	9,326	6,399

Time-validity annotations. In our synthetic dataset, the description of any situation (for instance, care plans, appointments, procedures, allergies) gives us an idea of possible illness, disabilities and current health condition of a person. Therefore, to extract the valid data points, we use the time validity annotations assigned to each SNOMED CT code according to the following considerations (see **Fig. 3**):

- Never valid - for SNOMED CT codes that do not describe specific conditions or refer to general procedures
- Specific time validity - for SNOMED CT codes that describe the recovery time in months and could range from one to several months according to the condition.
- Always valid - for SNOMED CT codes that describe long-term conditions.

snomed code	description	validity	comment	source
82078001	Take blood sample	0	General procedure, never valid	https://www.nhs.uk/conditions/blood-tests/
232717009	Coronary artery bypass grafting	3	Up to 3 months to recover	https://www.nhs.uk/conditions/coronary-artery-bypass-graft-cabg/
180030006	Amputation of foot	*	Long term condition, always valid	https://www.nhs.uk/conditions/amputation/

Fig. 3. Time-validity annotation examples

5.4 Data fitting

We have now a reduced collection of data points only including useful and valid information, classified according to their sensitivity. Specifically, we observe how all data points usable for answering our CQ are sensitive data! In addition, we should take into consideration how healthcare data is generated and read by health professionals; this means that the interpretation of such information may represent a challenge for emergency services (e.g., fire wardens). To solve the problems of sensitivity and interpretation of health records, we use the categories that represent disabilities according to [14]. In order to bridge the gap between the categories in the classification and the description of the health records, we use a common-sense knowledge base: ConceptNet. First, for each category in the list of disabilities, we find a key concept in ConceptNet that represents it (see **Table 2**). Second, to match the health record data points with the most related type of disability, we use the ConceptNet API[3]. The API compares two terms and returns a 'relatedness value' indicating how connected the two terms are; the higher the value, the more related each pair of terms are. Hence, we query the API to obtain the relatedness value between each valid data point and each key concept. After comparing all valid data points against the types of disabilities, our system calculates the average score, and this allows us to deliver a ranked list of the most related types of disabilities associated with the time-valid condition extracted from the health record. The result is a ranked list of possible reasons for assistance, answering the second part of our CQ (see **Table 3**).

Table 2. Types of disabilities and correspondent Key Concept

Category Description	Key Concept
Electric wheelchair and wheelchair user	*wheelchair_user*
Mobility impaired person	*movement_disorder*
Asthma and breathing issues	*respiratory_disease*
Visually impaired person	*visual_impairment*
Dyslexic and orientation disorders	*disorientation*
Learning difficulty and autism	*learning_difficulty*
Mental Health problems	*mental_health_problem*
Dexterity problems	*indexterity*
Hearing impaired person	*hearing_impaired*

Table 3. Ranked top 3 reasons for assistance

rank	category	score
1	Asthma and breathing issues	0.368
	/relatedness?node1=/c/en/respiratory_disease&node2=/c/en/injury_of_tendon_of_the_rotator_cuff_of_shoulder	0.126
	/relatedness?node1=/c/en/respiratory_disease&node2=/c/en/pulmonary_emphysema	0.610
2	Electric Wheelchair and wheelchair user	0.201
	/relatedness?node1=/c/en/wheelchair_user&node2=/c/en/injury_of_tendon_of_the_rotator_cuff_of_shoulder	0.371
	/relatedness?node1=/c/en/wheelchair_user&node2=/c/en/pulmonary_emphysema	0.031
3	Mobility impaired person	0.198
	/relatedness?node1=/c/en/movement_disorder&node2=/c/en/injury_of_tendon_of_the_rotator_cuff_of_shoulder	0.103
	/relatedness?node1=/c/en/movement_disorder&node2=/c/en/pulmonary_emphysema	0.293

6 System

In order to apply the proposed methodology, we developed a system which takes as input the annotated data source. Then it processes the data points following our approach to finally deliver the number of people requiring assistance and the type of help required. In what follows, we describe in detail the implementation of the system.

Input: as input, our system uses the types of disabilities, taken from the UK government guidelines for fire evacuation and tailored for our use case. The main data input is the annotated health record dataset according to its utility and sensitivity.

Process: first, our system identifies the people in the building at the moment the fire starts. For each person in the building, our system queries only the data points annotated

[3] https://github.com/commonsense/conceptnet5/wiki/API#relatedness-of-a-particular-pair-of-terms

as relevant. From this collection of relevant data points, our system now identifies the valid data points; this implies an evaluation of each data point according to its temporal validity. Each data point that represents a health condition has a time validity according to the type of disease, and its recovery time, thus, a data point is valid if the time validity and the data point start date overlap. If a person has at least one valid data point, it means that this person requires assistance.
On the contrary, if the person has no valid data points, then no assistance is required. Hence, our system identifies all the people that require assistance and the health records that support this result. Next, the system evaluates the type of disability, which defines the reason for the assistance. Our system uses the ConceptNet API to query the relatedness between each data point and each key concept of the type of disabilities. The query returns a 'relatedness value', the higher the value, the more related the pair of terms are. Thus, for each type of disability, our system registers an average score that allows us to obtain a ranked list of the most related types of disabilities.
Output: As exemplified in **Fig. 4**, the system returns a list of people requiring help and the best matching type of assistance needed according to their medical conditions.

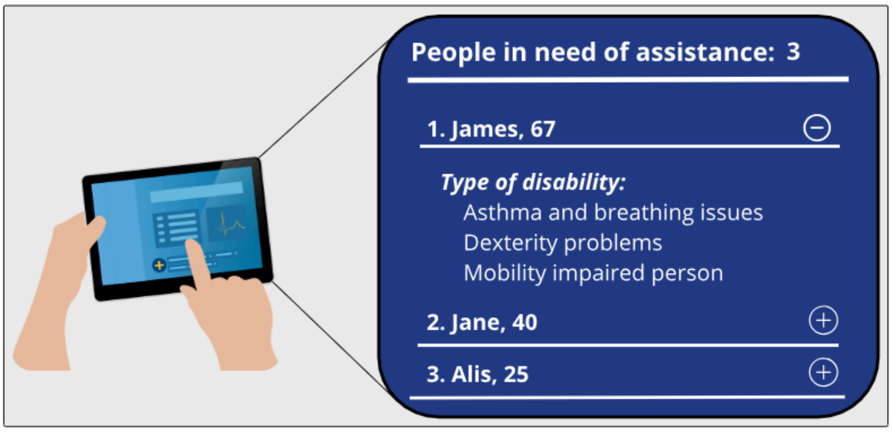

Fig. 4. Information provided by the system

7 Evaluation

In this section, we present the results of our system. We developed a well-curated Gold Standard Dataset that is the point of reference to measure the validity of our results.

Experiment settings. For our study, we generated a sample of 10,000 patients, intending to create a large sample of health records and including as many diseases as possible. From the experiments, we focus on randomly selected 1,012 patients' health records. The age of the patients ranges from 20 to 80 years old, as we try to simulate the ages of employees of a large organisation such as The Open University.

7.1 Gold Standard Dataset

To evaluate the results of our experiments, we developed a Gold Standard Dataset (GSD) based on a collection of annotated patients health records answering the following questions:

- Q1: Who needs special assistance in case of a fire evacuation?
- Q2: What type of assistance the person needs?

The GSD was developed by two of the authors independently, and they are referred as the reviewers. It is worth mentioning that the authors are members of a large organisation (The Open University), and their competence is comparable to that of a fire warden.

To support the reviewers in building the GSD, we developed a web interface that for each sample displays: *a)* the question to be answered, *b)* the patient's details (name, last name, age) and *c)* a section with the whole patient's health record (description, reason, type of record, start date, end date). For GSDQ1, we present the reviewer with the option to answer 'Yes' or 'No' to the question if the person needs assistance. The reviewer should read the health records and detect any condition that could reveal that a person has an impediment to evacuate the building. For GSDQ2, we display the list of the type of disabilities and ask the reviewers to choose at least one item from the list. It is essential to mention that we used the same list of disabilities as our system. Additionally, GSDQ2's sample is composed only of the samples annotated as 'Yes' in GSDQ1. The GSD was initially built by two of the authors, using the following process:

- Annotate the GSD individually.
- Identify discrepancies by reviewing the differences between their answers.
- Discuss each difference, explanations and evidence for answering 'Yes' or 'No', including external sources, such as the NHS website.
- Take a motivated decision: evaluate the evidence and reach an agreement.
- Annotate the reasons for the agreement: write down any comments and reasons to ensure consistency across decisions.

The resulting GSD is an account of how a person typically involved in supporting a fire evacuation may interpret the content of health records, having sufficient time and resources.

7.2 Baselines

We compared the approach described in Section 5, with several baselines developed considering alternative hypotheses.

Baseline (1M) - One-month time frame validity. Baseline 1M applies the hypothesis that most recent health records reflect conditions that affect an individual's capacity to perform an evacuation plan. Therefore, to find valid conditions, we experiment with a time frame of one month.
Baseline (2M) - Two-months' time frame validity. Baseline 2M applies the assumption that valid data points occur in a two-month time frame.

Baseline (1M+BLC) – Block-list for non-descriptive conditions. Usually, the health records include data points that do not represent a health condition, for example, 'Medical Reconciliation'. Hence, this Baseline applies a block-list to filter out non-descriptive SNOMED CT codes, in combination with the 1M time validity.
Baseline (1M+BLC+CHR) - List of chronic conditions. We detect that long-term conditions usually are excluded when considering short time frames (1M and 2M). For example, amputation of foot or heart conditions are not identified. By including long-term conditions, our system may correctly identify more people. This Baseline uses an allow-list of long-term conditions from SNOMED CT valid at any time.

7.3 Results

In what follows, we present the evaluation of the two research questions formulated in Section 1.
Research Question 1(RQ1). For RQ1, the objective is to identify who requires special assistance during an evacuation. To measure the performance of our system, we use the following metrics:

- We use accuracy to evaluate our system as a boolean classifier and measure its ability to distinguish whether an individual needs assistance or not.
- Precision, to measure the percentage of people identified as vulnerable that were correctly classified.
- Recall, to measure the percentage of actual people in need of assistance that were correctly classified. Recall is a particularly relevant measure for our system as we want to minimise the risk of missing a person in need.
- F-Measure, for measuring the performance of the system considering both precision and recall.

We compare the decisions of our system against the GSDQ1; the results obtained from its analysis are summarised in **Table 4**.

Table 4. Identification of people requiring assistance (RQ1)

Experiment	Accuracy	Precision	Recall	F-score
Our approach	**0.91**	**0.69**	**0.82**	**0.75**
Baseline (1M)	0.85	0.54	0.35	0.43
Baseline (2M)	0.82	0.44	0.48	0.45
Baseline (1M+BLC)	0.86	0.65	0.32	0.43
Baseline (1M+BLC+CHR)	0.81	0.43	0.47	0.45

Our approach reported accuracy of 0.91; therefore, our system correctly identifies 91% of the people that either need or not assistance. Precision and Recall prove to be significantly better results compared with alternative approaches. The main aim of our system is to maximise the possibility of identifying people in need of help, and thus in our study, we consider Recall the most important indicator. We managed to identify 82%

of the cases as people that actually need help. There was a significant difference between the F-score of our approach and the other hypotheses. Our system outperforms the baselines in all the measures.

Research Question 2 (RQ2). For RQ2, we focus on *why* people need assistance. We compare the results of our system against the results from GSDQ2. In order to evaluate the capacity of our system of providing a precise ranking of the most relevant categories of disabilities, we use Precision at K. The results obtained are summarised in **Table 5**. The results show that overall our approach attains a high precision on identifying the first three more likely reasons concerning a disability. In **Table 5,** we show the results also using the other baselines with the purpose of demonstrating how a more accurate selection of relevant data points leads to a better-quality classification. It is also important to mention that for the GSDQ2, we asked participants to select at least one type of disability and the one they consider most important. Therefore, our approach also should find a way to give each type of disability a degree of impact or a level of importance besides finding the most related types of disability according to a person's health records.

Table 5. Classification of type of assistance (RQ2). Precision at 3.

Experiment	1st category	2nd category	3rd category
Our approach	**0.47**	**0.52**	**0.73**
Input from Baseline (1M)	0.11	0.13	0.32
Input from Baseline (2M)	0.16	0.17	0.40
Input from Baseline (1M+BLC)	0.10	0.01	0.18
Input form Baseline (1M+BLC+CHR)	0.13	0.18	0.35

8 Discussion and conclusions

In this paper, we have introduced a methodology designed to make use of sensitive data in the context of a fire building evacuation. We developed a system that follows the proposed methodology and uses synthetic healthcare data to answer our research questions. We demonstrate that our approach allows us to identify people that require special assistance during a fire evacuation without the need of disclosing personal information. Specifically, we applied a knowledge engineering approach and used a common-sense knowledge base to categorise health conditions and transform the data for the convenience of non-expert users, for example, fire wardens and emergency responders.

Although results show a considerable high accuracy and recall, there is still work to be done in order to improve the precision. Possible directions could be reasoning over the combination of conditions and procedures. This hypothesis opens interesting challenges in relation to analysing and annotating large knowledge bases such as SNOMED CT in order to fit specific needs such as those of the emergency services.

Another important point is related to the experiment setting. Synthetic data is very accurate with respect to statistical considerations (e.g., the number of persons with a specific condition). However, there are also limitations. For example, we recognise that the synthetic dataset we used did not include explicit descriptions on the use of aid equipment, although these are contemplated in SNOMED CT. This information could definitely help on defining better strategies to automatically respond to why a person is in need of special assistance.

One aspect we consider relevant to explore concerns the different methodologies to support the implementation of systems compliant with data regulations, especially important for emergency services that should exchange sensitive data during exceptional situations. Further experiments could usefully explore the application of our proposed methodology in a different use case and include other sensitive datasets such as location or biometric data. Additionally, we recognise that a natural progression of this work will explore further the use of common-sense knowledge in order to support the interpretation of health records for timely emergency response in the Smart City.

References

1. Caird, S.: City approaches to smart city evaluation and reporting: case studies in the United Kingdom. Urban Res. Pract. 11, 159–179 (2018). https://doi.org/10.1080/17535069.2017.1317828.
2. Rego, A., Garcia, L., Sendra, S., Lloret, J.: Software Defined Network-based control system for an efficient traffic management for emergency situations in smart cities. Future Gener. Comput. Syst. 88, 243–253 (2018). https://doi.org/10.1016/j.future.2018.05.054.
3. Hussain, A., Wenbi, R., da Silva, A.L., Nadher, M., Mudhish, M.: Health and emergency-care platform for the elderly and disabled people in the Smart City. J. Syst. Softw. 110, 253–263 (2015). https://doi.org/10.1016/j.jss.2015.08.041.
4. Gharaibeh, A., Salahuddin, M.A., Hussini, S.J., Khreishah, A., Khalil, I., Guizani, M., Al-Fuqaha, A.: Smart Cities: A Survey on Data Management, Security, and Enabling Technologies. IEEE Commun. Surv. Tutor. 19, 2456–2501 (2017). https://doi.org/10.1109/comst.2017.2736886.
5. Scottish Government, S.: Your Emergency Care Summary: What does it mean for you?, https://www.gov.scot/publications/emergency-care-summary-mean/.
6. Yüksel, B., Küpçü, A., Özkasap, Ö.: Research issues for privacy and security of electronic health services. Future Gener. Comput. Syst. 68, 1–13 (2017). https://doi.org/10.1016/j.future.2016.08.011.
7. HM Government: Addressing lessons from the emergency response to the 7 july 2005 London bombings. (2016).
8. Bertino, E.: Data Security and Privacy: Concepts, Approaches, and Research Directions. In: 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC). pp. 400–407. IEEE, Atlanta, GA (2016). https://doi.org/10.1109/COMPSAC.2016.89.
9. HM Government: Data protection and sharing guidance for emergency planners and responders, https://www.gov.uk/government/publications/data-protection-and-sharing-guidance-for-emergency-planners-and-responders, (2007).
10. Morales, A., Motta, E., Daga, E.: Towards a privacy aware information system for emergency response. In: Proc. 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM). p. 4 (2019).
11. Speer, R., Chin, J., Havasi, C.: ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In: AAAI 31. p. 8.
12. HL7.org: HL7FHIR, https://www.hl7.org/fhir/overview.html.

13. SNOMED International, http://www.snomed.org/snomed-ct/five-step-briefing.
14. HM Government: Fire safety risk assessment: means of escape for disabled people, https://www.gov.uk/government/publications/fire-safety-risk-assessment-means-of-escape-for-disabled-people, (2007).
15. Majumder, S., Aghayi, Emad., Noferesti, M., Memarzadeh-Tehran, H., Mondal, T., Pang, Z., Deen, M.: Smart Homes for Elderly Healthcare—Recent Advances and Research Challenges. Sensors. 17, 2496 (2017). https://doi.org/10.3390/s17112496.
16. Boyi Xu, Li Da Xu, Hongming Cai, Cheng Xie, Jingyuan Hu, Fenglin Bu: Ubiquitous Data Accessing Method in IoT-Based Information System for Emergency Medical Services. IEEE Trans. Ind. Inform. 10, 1578–1586 (2014). https://doi.org/10.1109/tii.2014.2306382.
17. Abu-Elkheir, M., Hassanein, H.S., Oteafy, S.M.A.: Enhancing emergency response systems through leveraging crowdsensing and heterogeneous data. In: 2016 Int Wireless Communications and Mobile Computing Conference (IWCMC). pp. 188–193. IEEE (2016).
18. Thummavet, P., Vasupongayya, S.: A novel personal health record system for handling emergency situations. In: 2013 International Computer Science and Engineering Conference (ICSEC). pp. 266–271. IEEE (2013). https://doi.org/10.1109/icsec.2013.6694791.
19. Nunavath, V., Prinz, A., Comes, T.: Identifying First Responders Information Needs: Supporting Search and Rescue Operations for Fire Emergency Response. Int. J. Inf. Syst. Crisis Response Manag. IJISCRAM. (2016). https://doi.org/10.4018/ijiscram.2016010102.
20. McNeill, F., Bental, D., Missier, P., Steyn, J., Komar, T., Bryans, J.: Communication in Emergency Management through Data Integration and Trust: an introduction to the CEM-DIT system. In: Proc. of ISCRAM. p. 12. , Valencia, Spain (2019).
21. Shi, L., Li, S., Yang, X., Qi, J., Pan, G., Zhou, B.: Semantic Health Knowledge Graph: Semantic Integration of Heterogeneous Medical Knowledge and Services. BioMed Res. Int. 2017, 1–12 (2017). https://doi.org/10.1155/2017/2858423.
22. Nunavath, V., Prinz, A.: LifeRescue: A web based application for emergency responders during fire emergency response. In: 2016 3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM). pp. 1–8. IEEE (2016).
23. Kaelber, D.C., Jha, A.K., Johnston, D., Middleton, B., Bates, D.W.: A Research Agenda for Personal Health Records (PHRs). J. Am. Med. Inform. Assoc. 15, 729–736 (2008).
24. Oladimeji, E.A., Chung, L., Jung, H.T., Kim, J.: Managing security and privacy in ubiquitous eHealth information interchange. In: Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication. p. 1. ACM Press (2011).
25. Rahman, Sk.Md.M., Masud, Md.M., Hossain, M.A., Alelaiwi, A., Hassan, M.M., Alamri, A.: Privacy preserving secure data exchange in mobile P2P cloud healthcare environment. Peer--Peer Netw. Appl. 9, 894–909 (2016). https://doi.org/10.1007/s12083-015-0334-2.
26. Chehade, S., Pothin, J.-B., Matta, N., Cogranne, R.: Situation Representation and Awareness for Rescue Operations. In: Proc. 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM). p. 12 (2019).
27. Gyrard, A., Gaur, M., Shekarpour, S., Thirunarayan, K., Sheth, A.: Personalized Health Knowledge Graph. In: ISWC2018 Contextualized KG Workshop. p. 7 (2018).
28. Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Dube, K., Gallagher, T., McLachlan, S.: Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J. Am. Med. Inform. Assoc. 25, 230–238 (2018). https://doi.org/10.1093/jamia/ocx079.