

Article

# Assessing the Need for Semantic Data Integration for Surgical Biobanks—A Knowledge Representation Perspective

Mathias Brochhausen <sup>1,\*</sup> , Justin M. Whorton <sup>1</sup>, Cilia E. Zayas <sup>1</sup> , Monica P. Kimbrell <sup>2</sup>, Sarah J. Bost <sup>3</sup> , Nitya Singh <sup>4,5</sup>, Christoph Brochhausen <sup>6</sup>, Kevin W. Sexton <sup>1,7,8,9</sup>  and Bernd Blobel <sup>10,11,12</sup>

- <sup>1</sup> Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; jmwhorton@uams.edu (J.M.W.); czayas@uams.edu (C.E.Z.); kevin.sexton@uams.edu (K.W.S.)
- <sup>2</sup> Trauma Performance Improvement Coordinator, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA; mpkimbrell@uams.edu
- <sup>3</sup> Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL 32611, USA; sarah.bost@ufl.edu
- <sup>4</sup> Department of Animal Sciences, University of Florida, Gainesville, FL 32611, USA; nitya11@epi.ufl.edu
- <sup>5</sup> Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA
- <sup>6</sup> Central Biobank, Institute of Pathology, University and University Clinic of Regensburg, 93053 Regensburg, Germany; christoph.brochhausen@klinik.uni-regensburg.de
- <sup>7</sup> Department of Surgery, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA
- <sup>8</sup> Institute for Digital Health & Innovation, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA
- <sup>9</sup> BioVentures LLC, Little Rock, AR 72205, USA
- <sup>10</sup> Medical Faculty, University of Regensburg, 93053 Regensburg, Germany; bernd.blobel@klinik.uni-regensburg.de
- <sup>11</sup> eHealth Competence Center Bavaria, Deggendorf Institute of Technology, 94469 Deggendorf, Germany
- <sup>12</sup> First Medical Faculty, Charles University, 11636 Prague 1, Czech Republic
- \* Correspondence: mbrochhausen@uams.edu



**Citation:** Brochhausen, M.; Whorton, J.M.; Zayas, C.E.; Kimbrell, M.P.; Bost, S.J.; Singh, N.; Brochhausen, C.; Sexton, K.W.; Blobel, B. Assessing the Need for Semantic Data Integration for Surgical Biobanks—A Knowledge Representation Perspective. *J. Pers. Med.* **2022**, *12*, 757. <https://doi.org/10.3390/jpm12050757>

Academic Editor: Theodoros Papaioannou

Received: 22 March 2022

Accepted: 30 April 2022

Published: 7 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** To improve patient outcomes after trauma, the need to decrypt the post-traumatic immune response has been identified. One prerequisite to drive advancement in understanding that domain is the implementation of surgical biobanks. This paper focuses on the outcomes of patients with one of two diagnoses: post-traumatic arthritis and osteomyelitis. In creating surgical biobanks, currently, many obstacles must be overcome. Roadblocks exist around scoping of data that is to be collected, and the semantic integration of these data. In this paper, the generic component model and the Semantic Web technology stack are used to solve issues related to data integration. The results are twofold: (a) a scoping analysis of data and the ontologies required to harmonize and integrate it, and (b) resolution of common data integration issues in integrating data relevant to trauma surgery.

**Keywords:** surgical biobank; post-traumatic arthritis; osteomyelitis; semantic data integration; system theory; biomedical ontologies; knowledge representation

## 1. Introduction

Trauma is the leading cause of death and disability for patients less than 45 years old [1]. Recently, the need to decrypt the post-traumatic immune response to improve patient outcomes has been identified [2–5]. One strategy proposed to achieve this is large fluidics biobanks [6]. Along with other researchers [7], we advocate for more expansive surgical biobanks, including tissue. One of the core issues when building a trauma-oriented surgical biobank is integrating patient data from other healthcare providers. This is a generic data management problem regarding biobanking [8].

Implementation, maintenance, and use of a surgical biobank are mandatory, among other things, for a better understanding of the post-traumatic immune response, for instance, in patients with multiple injuries. We aim at improving secondary data analysis to

assess factors contributing to complications. Since patients with numerous injuries are frequently transferred from lower acuity facilities to a more specialized or higher-level trauma center, the data management activities of a surgical biobank need to include collecting and harmonizing data from all trauma care providers of the specimen donors. Collecting data from lower acuity facilities to trauma centers is crucial to identify potentially specific pathophysiological or new prognostic factors.

In this paper, we aim to achieve two goals:

Propose system architecture considerations for a surgical biobank against the background of crucial data needs to address post-surgical arthritis and osteomyelitis.

Demonstrate how semantic analysis guided by the generic component model (GCM) informs harmonization of clinical and clinically relevant data regarding post-traumatic arthritis and osteomyelitis in a surgical biobank. Thereby, we highlight the importance of definition, representation, and integration of the underlying concept spaces (ontologies) of the different domains involved, from which the data can be derived, represented, and implemented, but also the current context [9].

In the first step of our analysis, we will assess the patterns of complex collaborations in trauma patient triage and care using a systems theory approach. This approach will assist us in determining the scope of data collection and data harmonization across multiple healthcare providers. This research can aid in the development of IT infrastructure to support the implementation, maintenance, and management of a surgical biobank. To provide actionable information, we will suggest current domain ontologies that are applicable to our use cases.

In the second step, we will illustrate how Semantic Web technologies can resolve common data collection challenges related to surgical trauma: (a) issues created by restrictions specified in local information models, (b) issues created by the need to integrate data from heterogeneous sources, and c) issues created in the process of data entry.

## 2. Data Needs of Surgical Biobanks

Our research focuses on musculoskeletal injuries, one of the most common injury patterns leading to disability after trauma. These injuries are almost exclusively repaired surgically, making this the ideal population for creating a surgical biobank. As part of the fracture repair process, we can easily collect samples of bone, skin, subcutaneous tissue, small veins, and muscle without impact on the patients' recovery. Additionally, joint replacement is a common procedure allowing for age-matched control sample collection of similar tissues. Interestingly, the main indication for joint replacement is arthritis, a known complication of orthopedic trauma, thus allowing us to collect samples across the spectrum of life and investigate the mechanisms of post-traumatic arthritis. This is not our only outcome of interest, however, as infection, such as osteomyelitis or soft tissue infection, is the most common early complication of fracture repair. Diagnosis of both post-traumatic arthritis or osteomyelitis occurs after initial treatment for a fracture, and the reasons for the occurrences are not completely understood. Due to this, treatment for these diagnoses frequently occurs at different trauma care facilities from the initial treatment. In this context, it is desirable to establish predictive and prognostic factors for future therapeutic strategies to avoid numerous repetitive surgical interventions or amputations in complicated trauma patients. For this purpose, surgical biobanks collecting relevant biomaterials and comprehensive clinical and laboratory data are needed. The complete monitoring of laboratory, clinical, and imaging data as well as the historical information of primary care, transportation times, and duration of treatments are relevant. To reach this goal, data from different sources should be brought together with the biospecimens to curate data-rich biobank specimens.

The lack of a unified coding schema and of structured data for information relevant to address the pathogenesis of individual cases creates additional difficulty for data harmonization in this domain. While there are emerging classification systems, no widely accepted standard exists. In all likelihood, any standards set quickly, would continue

to change and evolve as our understanding of precision medicine in trauma increases. Thus, the problem of lack of comparable data for mid- and long-term studies would be proliferated for years to come. Since the existence of a well-developed understanding is a requirement for creating stable standards, it is obvious that we cannot use standards to collect, curate, and maintain the data we are collecting with the goal of developing our understanding of precision medicine in trauma. Challenges in harmonizing data are frequently found regarding treatment and procedure information, demographic information, problem lists, medication lists, and organizational structures of trauma centers [10–12] or healthcare providers involved in the initial treatment.

### **3. Integration Challenges Created by Restrictions of Local Information Models**

A common practice is that each trauma patient is admitted to the nearest hospital that participates in the regional or state trauma program. However, complications such as infections and prolonged or failed bone or wound healing due to complicated trauma or to known, or even unrecognized, primary diseases, such as diabetes, hypertension, or arteriosclerosis, make the transfer to a more specialized or higher-level trauma center necessary [13]. In these centers, an interdisciplinary team will review previous laboratory as well as clinical tests. Recent and earlier imaging data will be compared, and commonly, a surgical revision is necessary. With respect to the management of specimens acquired during that healthcare process, one data integration problem frequently occurs: The entry number of a specimen is commonly used as the reference number. Over the course of treatment, especially with complications, a patient may have several entries, each receiving a new entry number. These samples are identified and documented in the so-called sample history as follow-up samples for the initial injury. Typically, the sample history assignment is performed using the patient identifier. Many systems enforce that there be only one patient identifier for each patient. If the patient has all samples taken at the same healthcare institution, this does not present a problem. However, if some of the samplings are performed at another healthcare provider, tracing disease progression via the patient identifier is no longer possible because the entry number cannot be matched with the patient identifier at the location of the surgical biobank.

### **4. Integration Challenges Created by the Need to Integrate Data from Heterogeneous Sources**

The University of Arkansas for Medical Sciences (UAMS) Medical Center is the only adult Level 1 Trauma Center with verification by the American College of Surgeons in Arkansas and sees over 3000 trauma-related patients annually, with admission exceeding 2000 trauma-related admissions. The medical center is the only tertiary care center in the state and has 24 trauma and surgical intensive care unit beds. In collaboration with the surgical services, the surgical critical care teams manage critically ill patients yearly across trauma surgery, burn surgery, emergency general surgery, surgical oncology, pancreas-biliary surgery, bariatric surgery, neurosurgery, orthopedics, obstetrics and gynecology, vascular surgery, otolaryngology, urology, plastic surgery, and transplantation surgery. Being the only Level 1 trauma center in the state of Arkansas, UAMS is an excellent example of a trauma care facility, receiving 560 orthopedic trauma patients annually. Additionally, they are a referral center for orthopedic patients with complications, including post-traumatic arthritis and osteomyelitis. The UAMS Trauma Database strives to collect, maintain, and store clinical data on patients in the acute care phase as well as readmission of their population within 30 days of discharge. For 2021, we had 18 readmissions within 30 days in our ortho population. This excludes patients with acute care at another hospital, who are subsequently referred for post-acute complications. Acute-phase trauma information on this selected population of Arkansas trauma patients with complications is limited to what is in the transferring records or the patient's history and physical notes. These patients are not included in the UAMS Trauma Database, thus restricting orthopedic studies to those with initial acute trauma care at UAMS.

The Arkansas Trauma Registry (ATR) might be a potential source for tying all Arkansas trauma cases together. Challenges with using data from the ATR include variability in the collection of National Trauma Data Bank (NTDB) data elements and access to the data. In Arkansas, only Level 1 and Level 2 trauma center personnel receive annual training as part of the Trauma Quality Improvement Program (TQIP) [14]. Level 3 and Level 4 programs are encouraged to use the NTDB data dictionary for applicable NTDB data elements, but training is limited to a review of data dictionary changes at the Annual ATR Conference.

Even with TQIP training, variability of data collection is suspected to be present among Level 1 and Level 2 facilities due to different interpretation of definitions. To mitigate variability, the National Trauma Data Bank (NTDB) Data Dictionary [15] work group vets the dataset annually. This process includes refinement of data element terminology, plus revision and/or additions to definitions. Additionally, TQIP provides annual education update through their annual conference, and via web training modules/quizzes. This is why the educational opportunities offered by TQIP annually are relevant to data quality.

Even with this, there remains some variation in data elements such as “unplanned OR”, which has been refined from “unplanned return to the OR” to “unplanned visit to the OR” (OR stands for operation room). Another refinement includes adding exclusions to the definition. The current definition is as follows:

“Patients with an unplanned operative procedure OR patients returned to the operating room after initial operative management of a related previous procedure” [15].

This definition leaves much open to interpretation. Some users interpret this as a surgery that is unexpected or due to some untoward event and not potentially expected events such as visits to the OR due to a failed limb salvage and similar clinical situations. This is an example of a clinical situation where the possibility that the patient may require surgery is inherited. There is anecdotal evidence that there is a tendency to subsume limb salvage failure and similar examples under the unplanned OR category. This highlights the potential for different interpretations and the lack of shared understanding of the data elements in the NTDS Data Dictionary.

## 5. Integration Challenges Created during Data Entry

Besides the accurate resolution of identifiers and the ability to semantically integrate data, information about the organizational structure of trauma care providers give important insight into potential effects on patient outcomes. Curating data about organizational structures in trauma care, in particular with controlled vocabularies and ontologies, is a relatively new area of research and development. The NIH-funded CAFÉ (Comparative Assessment Framework for Environments of Trauma Care (R01GM111324)) provides a controlled vocabulary to describe and assess organizational structures of trauma care providers [11]. In addition, the project created a web-based architecture to collect, manage, and store semantically rich information about trauma care organizations via a user-friendly online questionnaire with the Ontology of Organizational Structures of Trauma centers and Trauma systems (OOSTT) automatically representing the answers in a computer-interpretable language [10,12].

For validation purposes, the CAFÉ project invited stakeholders to enter data about trauma centers and trauma systems. In the first round of this activity, we collected data on two Level 1 trauma centers and one Level 2 trauma center. Table 1 shows the extracted results for two CAFÉ questions.

The answers from trauma centers A and C looked unlikely based on the CAFÉ team’s experience. Our interpretation of the situation is that the persons who entered the data for those two questions for those two centers were not certain of the difference between these two categories or they did not have access to data that provided a differentiation between board-eligible and board-certified emergency physicians. While OOSTT has a representation of both “board-certified emergency physician role” and “board-eligible emergency physician role”, only the first is formally fully defined, so that at the current

state of semantic representation, a disambiguation was not possible. One factor contributing to the lack of clarity is that all board-certified emergency physicians were at some point (and remain to be) board-eligible. However, the intention behind this question is to find out how many emergency physicians are board-eligible, but have not yet been board-certified, as opposed to those who are already board-certified.

**Table 1.** Example answers to CAFÉ questionnaire.

CAFÉ Question	Trauma Center A (Level 2)	Trauma Center B (Level 1)	Trauma Center C (Level 1)
Number of emergency physicians who are board-certified in emergency medicine.	21	29	23
Number of emergency physicians who are board-eligible in emergency medicine	21	2	23

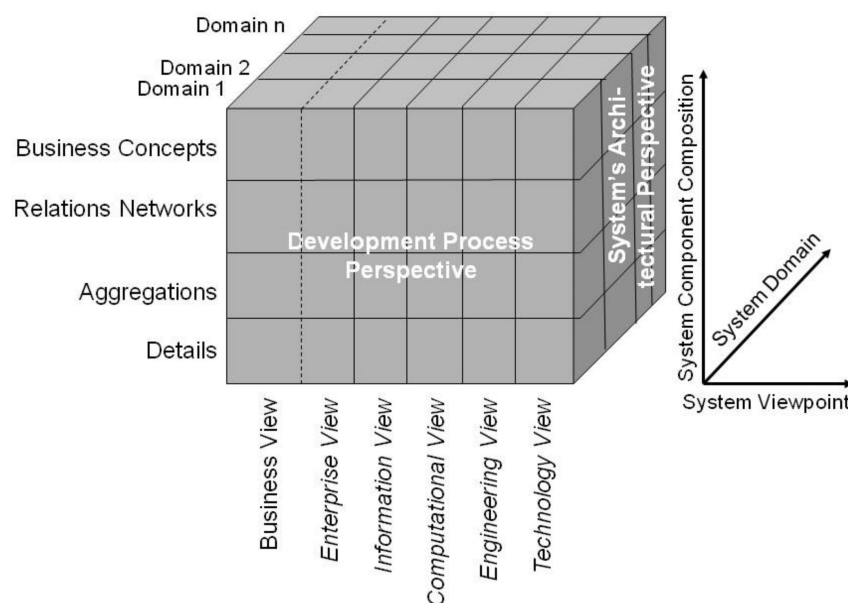
## 6. Methods

### 6.1. System Architecture Methodology: The Generic Component Model

The GCM is a top-level architectural model for any multi-domain system, formally representing the system’s components, their functions, and interrelations structurally and behaviorally by a cube with three dimensions: (a) specific aspects/perspective (domains) of/on the system forming domain-specific sub-systems; (b) generic granularity levels of the system’s elements enabling the composition/decomposition of the system; (c) the view-points within the system’s development process (Figure 1). It is described and specified in ISO 23903 Interoperability and integration reference architecture—Model and framework [16,17]. For each business case, the subsystem components and their functions and interrelations are instantiated by naming and representing them using the specific terminologies and ontologies of the domains involved in that business case. For this reason, the GCM specifies a business view in addition to the five views defined by the ISO 10746 Open distributed processing—Reference model (RM-ODP) as starting point for a system development process [17]. The views prescribed by the RM-ODP are enterprise view (purpose, scope, and policies of the system), information view (information processing, semantics of information), computational view (functionality of the system, functional decomposition), engineering view (implementation, distribution of processing performed by the system), and technology view (choice of technology for the system), all represented using information and communication technology (ICT) ontologies [17,18]. Furthermore, ISO 23903 [16] introduces generic granularity levels for correctly representing and interrelating compositions/decompositions of elements, that way enabling integration of, and interoperability between, elements of different complexity. As only elements at the same granularity level can be interrelated, elements at different granularity level first have to be harmonized by composing or decomposing them. By adding the business view and granularity levels, the GCM enables the correct representation and management of multi-domain real-world systems, including supporting ICT solutions.

Applying the methods to the use case at hand first requires considering the role of the information models and the database schemata according to the views of the GCM. Both belong to the information view since their primary focus is to guide information processing within the system. Hence, they do not provide real-world knowledge, such as the fact that one person can have more than one patient ID and that sampling of tumor progression is frequently performed by different healthcare providers. More generally, due to the limitations of IT grammars, the correctness of representations of, and relations between, elements cannot be justified within this viewpoint. Notably, those aspects lie outside each electronic health record (EHR) system and are ill-fitted to be represented within the information view. However, from our database example, not accounting for those aspects of the business view may lead to errors in the system.





**Figure 1.** The generic component model. From: [18].

## 6.2. Semantic Integration Methods

Our challenge is to provide a semantically rich representation of clinical and clinically relevant data for a surgical biobank. This approach has been shown to be useful for integrating heterogeneous clinical data in an international cancer imaging repository, The Cancer Imaging Archive [19–21]. To create a semantically rich representation for biomedical data, we follow the recommendations outlined by Brochhausen et al. [22] The basic strategy is to use the Semantic Web stack [23] along with realism-based ontology development [24] as the general methodology to inform knowledge representation. Their approach is based on experiences with multiple implementations in biomedical informatics, which have influenced the evolution of their methodology from the beginning [25–28].

In our project, we are transforming clinical and clinically relevant data into the resource description framework (RDF) language before we load it into our RDF-based data management system of the biobank. RDF [29] is a Semantic Web standard that allows the representation of information in a machine-interpretable way. For each entity, RDF provides a unique identifier. RDF data can be annotated and used along with domain descriptions provided in the Web Ontology Language (OWL) [30]. We are using RDF representations and OWL ontologies following the methodology described by Smith and Ceusters [24] and the best-practice principles of the OBO Foundry [31,32].

In addition, we will use the GCM (Section 6.1) to assist knowledge representation. The GCM proves to be a fitting tool to help with selecting optimal knowledge representation for incoming data and guide the process of filling gaps in existing ontologies, based on the medical information models used to curate the data at the initial healthcare providers. While IT-oriented information models fulfil a crucial role in planning, defining, and describing the operational behavior of an IT system, such as an EHR system or a biobank information system, there is a lack of semantic capabilities of these models [21]. Due to the high level of abstraction and expressivity of information models, so rarely being complete and decidable, the resulting representation, especially regarding relationships, is frequently focused on data in that system alone and not on what the data represents, in our case the medical world, e.g., the patients, encounters, and prescriptions described in an EHR or the specimens, donors, and storage properties in a biobank, thereby including context and common-sense knowledge. Building an ontology or selecting from existing ontologies can lead to a mismatch between the system data model and the more holistic ontologies. In our approach, we use the GCM to overcome the implementation gap between healthcare-

provider-specific information models and preexisting domain ontologies needed for the semantically rich representation of surgical biobank data.

In this paper, we follow the approach of Uribe et al. [33] used to represent a GCM-based generic model of a Type 2 diabetes mellitus care system. In parallel to their approach, we analyze the following three domains of the system: medical domain, policy domain, and resource domain.

- *Resource Domain*—represents all agents (humans, organizations, devices), means, and equipment to carry out activities (in our example, trauma care activities) in the system [33].
- *Policy Domain*—represents rules, regulations, and guidelines relevant to the system [33] (e.g., clinical guidelines, trauma system regulations, EHR regulations).
- *Medical Domain*—represents the medical and biological entities relevant to the system [33] (anatomical entities, treatments, diagnoses, etc.).

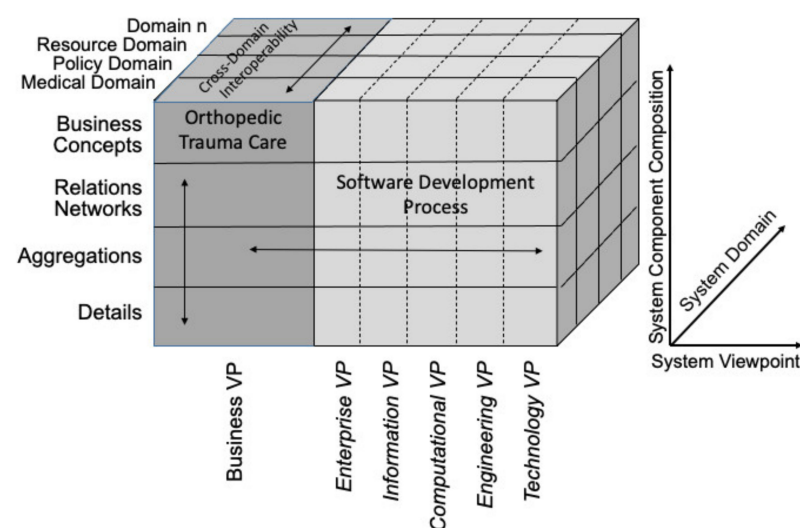
In this paper, we are not applying the GCM for a software development and implementation process, but restrict our considerations to the business view, following the approach by Uribe et al. [33].

## 7. Results

### 7.1. System Architecture Results

Using the GCM and applying it to analyze a medical care system for orthopedic trauma from the business view yields several interesting results. The findings we describe can help to inform the development of better communication and data management methods to close communication gaps that currently exist, particularly when it comes to the need for additional therapy, for osteomyelitis and post-surgical arthritis. Regarding our goal of establishing a surgical biobank to fill knowledge gaps regarding these two diagnoses, this analysis provides a survey regarding the relevant agents (healthcare providers), facilities, and guidelines. This enables a comprehensive assessment of the data needs of such a biobank.

Figure 2 shows the domains relevant to orthopedic trauma care, in accordance with the GCM. Table 2 provides a list of examples of entities relevant to the GCM business view of orthopedic trauma care. These entities are sorted into the resource domain, policy domain, and medical domain. In addition, we provide a list of OBO Foundry [31,32] ontologies that provide a semantically rich representation of entities in those domains. This analysis allows for correct modeling of the inter-domain relationships to ensure cross-domain interoperability.



**Figure 2.** The GCM applied to analyze the business VP of orthopedic trauma care, following the approach used by Uribe et al. [33] for diabetes mellitus. Adapted from [33].

**Table 2.** Domains of orthopedic trauma care with examples of relevant entities and potential domain ontologies.

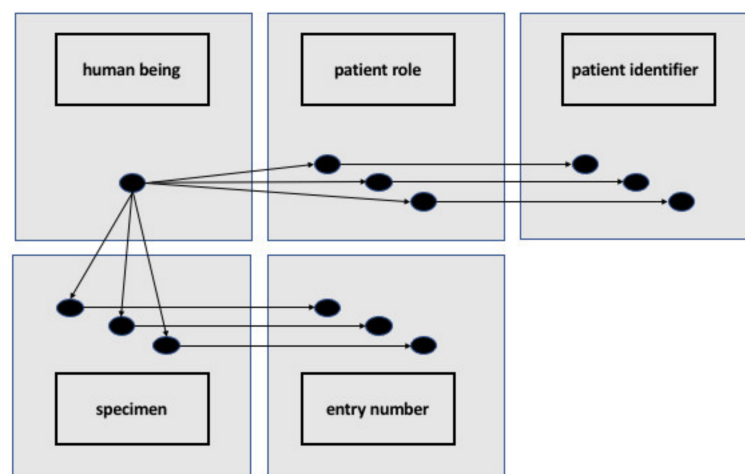
Domain	Types of Entities in Trauma Care	Potential Ontologies
Resource Domain	<b>Organization:</b> Trauma center, trauma system, emergency medical services (EMS), trauma team	OOSTT [10–12], OMRSE [34]
	<b>Human Individual:</b> Trauma patient, trauma medical director, trauma registrar, trauma surgeon, EMS personnel, plastic surgeon, infectiologist, microbiologist, surgical pathologist, endocrinologist, radiologist	
	<b>Facilities:</b> EMS vehicle, emergency room, trauma biobank, trauma registry	
Policy Domain	Resources for optimal care for injured patients; [35] EAST Practice Management Guidelines, [36] trauma system policies; triage plans; clinical guidelines	
Medical Domain	<b>Specimens:</b> Bone specimen; skin specimen; subcutaneous tissue specimen; small vein specimen, muscle specimen	OBIB [37], OBI [38], OMRSE [34], SNOMED-CT
	<b>Diagnosis:</b> Osteomyelitis, post-operative arthritis, pseudarthrosis,	
	<b>Treatment:</b> Resection, reconstruction, prosthesis, amputation, chronic injection therapy, nerve blocks for chronic pain	
	<b>Pathological process:</b> Inflammation, systemic inflammatory response syndrome, infection, necrosis, wound healing, bone regeneration, stress-induced hyperglycemia,	

### 7.2. Semantic Integration Results

#### 7.2.1. Overcoming Issues Created by Local Information Models—Identifiers

Applying the methods to the use case of integration issues created by restrictions on identifiers in local information models requires consideration of these information models from the perspective of the GCM (see Section 6.1).

Figure 3 shows an RDF representation of specimen donor, specimen, patient role, and the relevant identifiers. There is no restriction on the numbers of identifiers, and the graph representation allows one to trace each identifier to what it identifies, and each patient’s role and each specimen to the person who is the specimen donor. The different classes and instances in our solution are modeled based on the best practice of realism-based ontology as described in Smith and Ceusters [24].



**Figure 3.** Representation of RDF individuals and OWL classes representing a human being having multiple patient roles, corresponding to multiple patient IDs and multiple specimens derived from that human being corresponding to multiple entry numbers. From: [39].



### 7.2.2. Integration of Heterogeneous Data Sources

This problem represents the classical use case for Standard Widget Toolkits (SWTs) using the native graphical elements of the computer's operating system: existing data models, common data elements, or dictionaries are heterogeneous or ambiguous. Solving this issue using SWTs rests on the idea of using computer systems to help disambiguate the data from heterogeneous sources. Bona et al. pointed out that this rests on an agreed understanding of what it means for a computer to *understand* the content [22]. Ultimately, the level of understanding proposed by Bona et al. is the ability to sort data points into data elements using SWTs [22].

The solution we propose here is to start with providing an RDF-style representation of "unplanned visit to the OR" that formally defines whether or not ER visits such as OR due to a failed limb salvage are an instance of an unplanned visit to the OR, or similar surgical procedures, or not. The disagreement is whether or not OR visits that might not have been initially planned, but that were always a final option, belong in this category. We propose to create a new OWL class for "surgical encounter" as a subclass of "health care encounter" (OGMS\_0000097) from the Ontology for General Medical Science (OGMS) [40]. For the "surgical encounter" class, we propose the following textual definition: A healthcare encounter, where the goal is to provide a patient with surgical treatment.

In addition to expanding a class from OGMS, we also propose to expand the class "plan specification" (IAO\_0000104) from the Information Artifact Ontology (IAO) [41] by creating a subclass "treatment plan specification". For this class we propose the following textual definition: A plan specification that defines the objectives, actions, and healthcare encounters to treat the medical condition of one patient.

Now we can create a subclass of "surgical encounter" named "unplanned surgical encounter", which we plan to represent "unplanned visit to the OR", excluding visits such as OR due to a failed limb salvage. For this new class, we propose the following definition: A surgical encounter that does not realize the concretization of a plan specification which is part of a treatment plan.

We propose the following equivalence axiom for "unplanned surgical encounter":  
"surgical encounter"

and not realizes some ('realizable entity' and concretizes some ('plan specification' and 'part of' some 'treatment plan'))

This axiom would ensure that any surgical procedures that are performed during healthcare encounters without being part of a treatment plan are not pulled into the class "unplanned surgical encounter". In the example of failed limb salvage, the amputation of the limb under treatment is already a medical option in the treatment, although certainly not the most desirable outcome. Hence, any case of failed limb salvage would not be subsumed under an unplanned surgical encounter. This would disambiguate the current situation and resolve the lack of clarity, using a computational method.

### 7.2.3. Resolving Problem Integration Issues Created during Data Entry

The solution to the lack of clarity between board-eligible emergency physicians and board-certified physicians we detected during the data entry for the CAFÉ project begins with OOSTT. The knowledge representation requirements of the CAFÉ project did not force the previous development of a class representing the individual emergency physicians, but the following classes were sufficient: "board-eligible emergency physician role" (OOSTT\_00000173) and "board-certified emergency physician role" (OOSTT\_00000171). To resolve the issue surrounding potential ambiguities during data entry, we created two novel classes in a new branch of the OOSTT ontology: "board-certified emergency physician" (OOSTT\_00010000) and "board-eligible emergency physician" (OOSTT\_00010001). These two classes have the following equivalence axioms:

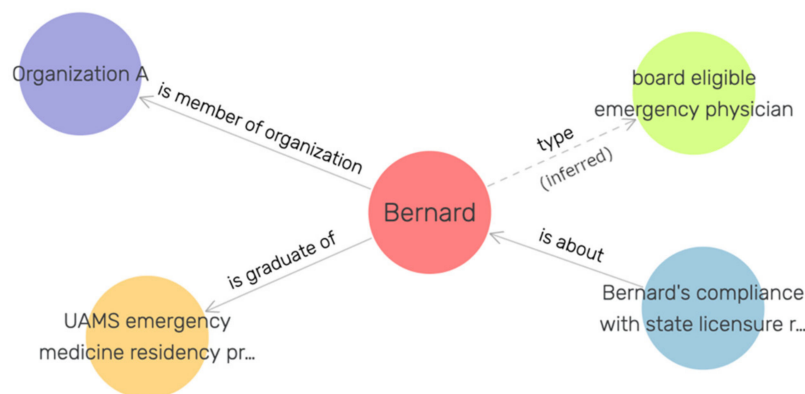
"board certified emergency physician" (OOSTT\_00010000):

"Homo sapiens"

and 'bearer of' some "board-certified emergency physician role"

“board-eligible emergency physician’ (OOSTT\_00010001):  
 “Homo sapiens”  
 and (is\_specified\_output\_of some “emergency medicine residency program”)  
 and (inverse (“is about”) some ‘compliance with state licensure requirement information content entity’).

After making these changes to the class structure of OOSTT, we created 100 test individuals to represent emergency physicians in 28 virtual trauma centers, some being board-eligible, some being board-certified. Figure 4 shows an example of the RDF representation of a board-eligible emergency physician. We loaded the resulting ontology and the individuals in a GraphDB triple store using ROBOT [42] to materialize class assertions. We queried this triple store with the following SPARQL queries to query out the board-certified individuals vs. the board-eligible individuals that are not (yet) board-certified.



**Figure 4.** Example individual representing a human being, Bernard, who is board-eligible, but not yet board-certified.

### 8. Discussion

In applying a system architecture approach to orthopedic trauma care using the GCM, we follow Uribe et al. [33] to order the involved domains in the following sequence: medical, policy, and resource (Figure 2). This allows us to analyze and to represent the relationships between orthopedic trauma care (e.g., procedures) and the policy domain (e.g., clinical guidelines, state regulations). In placing the resource domain behind the policy domain, the focus of our analysis demonstrates how the resources (e.g., trauma centers, trauma personnel) relate to the policies, meaning which actions, requirements, and credentials the policies specify for trauma care resources. This order of the domain prioritizes the role of policy in the trauma care process, and the roles of the resources in those policies for our initial approach. This order is consistent with the accreditation practices of American College of Surgeons and state trauma systems. It is important to note that the ordering of the domains when analyzing a system is arbitrary and can be chosen according to the focus of the study. To analyze and represent how the trauma care personnel relates to the procedures, we will reorder our domains to focus on the cross-domain interoperability between the resource domain and the medical domain.

To resolve problems of data integration due to restrictions in local information models, we have applied both the GCM and methods of a realism-based ontology development. Notably, the successful deployment of ISO 23903 [17] for integrating different domains and knowledge spaces, including their specific models, has been demonstrated for the integration of HL7 privacy and security specifications [43] in ISO 13606 EHR communication [44], the harmonization of concepts from ISO 12967 [45] and ISO 13940 [46], or the mapping of open EHR (ISO 13606) archetypes [47], ISO 13972 clinical models [48], and HL7 FHIR resources [49].

The limitations of this study and the three use cases demonstrated lie in the fact that the work presented addresses issues from a knowledge representation perspective.

The solutions presented are readily available to be implemented in a data management system, but they have not yet been implemented. Thus, in this study we do not assess effects that are created by using our knowledge representation in a running medical information system, such as a biobank system.

The next step from the knowledge representation perspective is to develop computational ways to leverage the GCM system analysis and its results to immediately develop knowledge representation solutions such as OWL ontologies and RDF data. In addition, we plan to explore the deployment of the GCM to transform medical data captured with information models into computer-interpretable RDF data sources for secondary use in medicine and vice versa.

From the clinical perspective, the next step is to build a small proof-of-concept surgical biobank data management system following the knowledge representation principles demonstrated in this paper. This step, which we are currently working on, will allow validation of the functionality and usability of the proposed solutions in a clinical research setting.

## 9. Conclusions

From the results presented in this paper, we concluded that SWT is required to disambiguate and integrate data for surgical biobanks. While surgical biobanks with exclusively local data might not have the same level of need for semantic integration as those aiming to integrate data from multiple trauma care providers, we hold that integrating data from the entirety of the trauma care process is a necessity.

In addition, this paper demonstrates that the GCM is a useful tool to assess the data scope of a surgical biobank and identify relevant domain ontologies. The GCM even assists in resolving specific semantic integration problems, supplementing the SWTs.

**Author Contributions:** Conceptualization, M.B., C.B., K.W.S. and C.B.; use-cases, C.E.Z., J.M.W., M.P.K., S.J.B., N.S., C.B., K.W.S.; methodology, M.B., B.B.; virtual data set and formal analysis J.M.W.; writing—original draft preparation, M.B.; C.E.Z., J.M.W., M.P.K., C.B., K.W.S., B.B.; writing—review and editing, M.B.; C.E.Z., J.M.W., M.P.K., S.J.B., N.S., C.B., K.W.S., B.B.; visualization, M.B., J.M.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** Parts of the research presented in this article are funded by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM111324.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Injuries and Violence Are Leading Causes of Death. 2021. Available online: <https://www.cdc.gov/injury/wisqars/animated-leading-causes.html> (accessed on 24 February 2022).
2. Janicova, A.; Relja, B. Neutrophil Phenotypes and Functions in Trauma and Trauma-Related Sepsis. *Shock Augusta Ga.* **2021**, *56*, 16–29. [[CrossRef](#)] [[PubMed](#)]
3. Skelton, J.K.; Purcell, R. Preclinical models for studying immune responses to traumatic injury. *Immunology* **2021**, *162*, 377–388. [[CrossRef](#)] [[PubMed](#)]
4. Recknagel, S.; Bindl, R.; Brochhausen, C.; Göckelmann, M.; Wehner, T.; Schoengraf, P.; Huber-Lang, M.; Claes, L.; Ignatius, A. Systemic inflammation induced by a thoracic trauma alters the cellular composition of the early fracture callus. *J. Trauma Acute Care Surg.* **2013**, *74*, 531–537. [[CrossRef](#)] [[PubMed](#)]
5. Kerby, J.D.; Griffin, R.L.; MacLennan, P.; Rue, L.W.I. Stress-Induced Hyperglycemia, Not Diabetic Hyperglycemia, Is Associated with Higher Mortality in Trauma. *Ann. Surg.* **2012**, *256*, 446–452. [[CrossRef](#)]
6. Yanik, E.L.; Keener, J.D.; Lin, S.J.; Colditz, G.A.; Wright, R.W.; Evanoff, B.A.; Jain, N.B.; Saccone, N.L. Identification of a Novel Genetic Marker for Risk of Degenerative Rotator Cuff Disease Surgery in the UK Biobank. *J. Bone Joint Surg. Am.* **2021**, *103*, 1259–1267. [[CrossRef](#)]
7. Hojat, A.; Wei, B.; Olson, M.G.; Mao, Q.; Yong, W.H. Procurement and Storage of Surgical Biospecimens. *Methods Mol. Biol. Clifton NJ.* **2019**, *1897*, 65–76. [[CrossRef](#)]

8. Spjuth, O.; Krestyaninova, M.; Hastings, J.; Shen, H.-Y.; Heikkinen, J.; Waldenberger, M.; Langhammer, A.; Ladenvall, C.; Esko, T.; Persson, M.-Å.; et al. Harmonising and linking biomedical and clinical data across disparate data archives to enable integrative cross-biobank research. *Eur. J. Hum. Genet. EJHG* **2016**, *24*, 521–528. [CrossRef]
9. Blobel, B.; Ruotsalainen, P.; Oemig, F. Why Interoperability at Data Level Is Not Sufficient for Enabling pHealth? *Stud. Health Technol. Inform.* **2020**, *273*, 3–20. [CrossRef]
10. Utecht, J.; Judkins, J.; Otte, J.N.; Colvin, T.; Rogers, N.; Rose, R.; Alvi, M.; Hicks, A.; Ball, J.; Bowman, S.M.; et al. OOSTT: A Resource for Analyzing the Organizational Structures of Trauma Centers and Trauma Systems. *CEUR Workshop Proc.* **2016**, *1747*. Available online: [http://ceur-ws.org/Vol-1747/IT504\\_ICBO2016.pdf](http://ceur-ws.org/Vol-1747/IT504_ICBO2016.pdf) (accessed on 4 May 2022).
11. Utecht, J.; Ball, J.; Bowman, S.M.; Dodd, J.; Judkins, J.; Maxson, R.T.; Nabaweesi, R.; Pradhan, R.; Sanddal, N.D.; Winchell, R.J.; et al. Development and validation of a controlled vocabulary: An OWL representation of organizational structures of trauma centers and trauma systems. *Stud. Health Technol. Inform.* **2019**, *264*, 403–407. [CrossRef]
12. Brochhausen, M.; Ball, J.W.; Sanddal, N.D.; Dodd, J.; Braun, N.; Bost, S.; Utecht, J.; Winchell, R.J.; Sexton, K.W. Collecting data on organizational structures of trauma centers: The CAFE web service. *Trauma Surg. Acute Care Open* **2020**, *5*, e000473. [CrossRef]
13. Bradley, N.L.; Garraway, N.; Bell, N.; Lakha, N.; Hameed, S.M. Data capture and communication during transfers to definitive care in an inclusive trauma system. *Injury* **2017**, *48*, 1069–1073. [CrossRef]
14. American College of Surgeons. *Trauma Quality Improvement Program (TQIP)*; American College of Surgeons: Chicago, IL, USA; Available online: <http://www.facs.org/quality-programs/trauma/tqip/center-programs/tqip> (accessed on 15 February 2022).
15. American College of Surgeons. *NTDS 2021 Data Dictionary*; American College of Surgeons: Chicago, IL, USA; Available online: <http://www.facs.org/quality-programs/trauma/tqip/center-programs/ntdb/ntds/data-dictionary> (accessed on 15 February 2022).
16. *ISO 23903:2021; Health Informatics—Interoperability and Integration Reference Architecture—Model and Framework*. ISO: Geneva, Switzerland. Available online: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/73/77337.html> (accessed on 17 February 2022).
17. *ISO/IEC 10746-3:2009; Information Technology—Open distributed processing—Reference Model: Architecture—Part 3*. ISO: Geneva, Switzerland. Available online: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/57/55724.html> (accessed on 10 March 2022).
18. Blobel, B. Standardization for Mastering Healthcare Transformation—Challenges and Solutions. *Eur. J. Biomed. Inform.* **2017**, *13*, 9–15. [CrossRef]
19. Prior, F.W.; Clark, K.; Commean, P.; Freymann, J.; Jaffe, C.; Kirby, J.; Moore, S.; Smith, K.; Tarbox, L.; Vendt, B.; et al. TCIA: An information resource to enable open science. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Osaka, Japan, 3–7 July 2013; pp. 1282–1285. [CrossRef]
20. Sharma, A.; Tarbox, L.; Kurc, T.; Bona, J.; Smith, K.; Kathiravelu, P.; Bremer, E.; Saltz, J.H.; Prior, F. PRISM: A Platform for Imaging in Precision Medicine. *JCO Clin. Cancer Inform.* **2020**, *4*, 491–499. [CrossRef]
21. Bona, J.P.; Prior, F.W.; Zozus, M.N.; Brochhausen, M. Enhancing Clinical Data and Clinical Research Data with Biomedical Ontologies—Insights from the Knowledge Representation Perspective. *Yearb. Med. Inform.* **2019**, *28*, 140–151. [CrossRef]
22. Brochhausen, M.; Bona, J.; Blobel, B. The Role of Axiomatically-Rich Ontologies in Transforming Medical Data to Knowledge. *Stud. Health Technol. Inform.* **2018**, *249*, 38–49.
23. The Semantic Web Layer Cake. Available online: <https://www.w3.org/2007/03/layerCake.svg> (accessed on 16 February 2022).
24. Smith, B.; Ceusters, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Appl. Ontol.* **2010**, *5*, 139–188. [CrossRef]
25. Hogan, W.R.; Hanna, J.; Joseph, E.; Brochhausen, M. Towards a Consistent and Scientifically Accurate Drug Ontology. *CEUR Workshop Proc.* **2013**, *1060*, 68–73.
26. Ceusters, W.M.; Spackman, K.A.; Smith, B. Would SNOMED CT benefit from realism-based ontology evolution? *AMIA Annu. Symp. Proc. AMIA Symp.* **2007**, *2007*, 105–109.
27. Lin, Y.; Xiang, Z.; He, Y. Brucellosis Ontology (IDOBURU) as an extension of the Infectious Disease Ontology. *J. Biomed. Semant.* **2011**, *2*, 9. [CrossRef]
28. Grenon, P.; Smith, B.; Goldberg, L. Biodynamic ontology: Applying BFO in the biomedical domain. *Stud. Health Technol. Inform.* **2004**, *102*, 20–38.
29. Resource Description Framework (RDF) Primer. Available online: <https://www.w3.org/TR/rdf11-concepts/> (accessed on 16 February 2022).
30. OWL 2 Web Ontology Language Primer (Second Edition). Available online: <https://www.w3.org/TR/owl2-primer/> (accessed on 16 February 2022).
31. Jackson, R.; Matentzoglou, N.; Overton, J.A.; Vita, R.; Balhoff, J.P.; Buttigieg, P.L.; Carbon, S.; Courtot, M.; Diehl, A.D.; Dooley, D.M.; et al. OBO Foundry in 2021: Operationalizing open data principles to evaluate ontologies. *Database J. Biol. Databases Curation* **2021**, *2021*, baab069. [CrossRef]
32. Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L.J.; Eilbeck, K.; Ireland, A.; Mungall, C.J.; et al. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **2007**, *25*, 1251–1255. [CrossRef]

33. Uribe, G.A.; Blobel, B.; López, D.M.; Schulz, S. A generic architecture for an adaptive, interoperable and intelligent type 2 diabetes mellitus care system. *Stud. Health Technol. Inform.* **2015**, *211*, 121–131.
34. Hicks, A.; Hanna, J.; Welch, D.; Brochhausen, M.; Hogan, W.R. The ontology of medically related social entities: Recent developments. *J. Biomed. Semant.* **2016**, *7*, 47. [[CrossRef](#)]
35. American College of Surgeons and Committee on Trauma. *Resources for Optimal Care of the Injured Patient*; American College of Surgeons: Chicago, IL, USA, 2014.
36. EAST Practice Management Guidelines. Available online: <https://www.east.org/education-career-development/practice-management-guidelines> (accessed on 15 February 2022).
37. Brochhausen, M.; Zheng, J.; Birtwell, D.; Williams, H.; Masci, A.M.; Ellis, H.J.; Stoeckert, C.J. OBIB—a novel ontology for biobanking. *J. Biomed. Semant.* **2016**, *7*, 23. [[CrossRef](#)]
38. Bandrowski, A.; Brinkman, R.; Brochhausen, M.; Brush, M.H.; Bug, B.; Chibucos, M.C.; Clancy, K.; Courtot, M.; Derom, D.; Dumontier, M.; et al. The Ontology for Biomedical Investigations. *PLoS ONE* **2016**, *11*, e0154556. [[CrossRef](#)]
39. Brochhausen, M.; Bost, S.J.; Singh, N.; Brochhausen, C.; Blobel, B. Understanding the Gap between Information Models and Realism-Based Ontologies Using the Generic Component Model. *Stud. Health Technol. Inform.* **2021**, *285*, 159–164. [[CrossRef](#)] [[PubMed](#)]
40. Ontobee: OGMS. Available online: [http://www.ontobee.org/ontology/OGMS?iri=http://purl.obolibrary.org/obo/OGMS\\_0000029](http://www.ontobee.org/ontology/OGMS?iri=http://purl.obolibrary.org/obo/OGMS_0000029) (accessed on 31 December 2021).
41. Information Artifact Ontology (IAO). The Foundry, OBO. Available online: <https://obofoundry.org/ontology/iao.html> (accessed on 4 May 2022).
42. Jackson, R.C.; Balhoff, J.P.; Douglass, E.; Harris, N.L.; Mungall, C.J.; Overton, J.A. ROBOT: A Tool for Automating Ontology Workflows. *BMC Bioinform.* **2019**, *20*, 407. [[CrossRef](#)] [[PubMed](#)]
43. HL7, HL7 Standards Product Brief—HL7 Version 3 Domain Analysis Model: Composite Security and Privacy, Release 1 | HL7 International. Available online: [http://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=536](http://www.hl7.org/implement/standards/product_brief.cfm?product_id=536) (accessed on 17 February 2022).
44. *ISO 13606-1:2019*; ISO: Health Informatics—Electronic Health Record Communication—Part 1: Reference Model. ISO: Geneva, Switzerland. Available online: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/78/67868.html> (accessed on 10 March 2022).
45. *ISO 12967-1:2020*; ISO: Health Informatics—Service Architecture (HISA)—Part 1: Enterprise Viewpoint. ISO: Geneva, Switzerland. Available online: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/10/71037.html> (accessed on 10 March 2022).
46. *ISO 13940:2015(en)*; Health Informatics—System of Concepts to Support Continuity of Care. ISO: Geneva, Switzerland. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:13940:ed-1:v1:en> (accessed on 17 February 2022).
47. OpenFoundation, EHR. Available online: <https://www.openehr.org/> (accessed on 16 February 2022).
48. *ISO 13972*; ISO: Health Informatics—Clinical Information Models—Characteristics, Structures, and Requirements. ISO: Geneva, Switzerland. Available online: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/94/79498.html> (accessed on 17 February 2022).
49. HL7 International. HL7 Fast Healthcare Interoperability Resources (FHIR). Available online: <http://www.hl7.org/fhir/> (accessed on 16 February 2022).