MS in Healthcare Management
University of Macedonia

Information Systems in Clinical Research: Categorization and evaluation of information systems and development of a guide for choosing the appropriate information system.

Advisor: Professor Konstantinos Tarabanis
Student: Lydia N. Zachariadou

Thessaloniki, September 2019

Abstract

The development of information systems used in clinical research is constantly increasing, as their advantages are widely acknowledged. Although many researchers have introduced information systems which can be used during a clinical study's process, a scarcity of information systems accommodating the complete process has been detected. Based on this finding, twenty-three (23) information systems and ontologies used in clinical research were retrieved, based on certain criteria. The information systems and ontologies were then categorized and evaluated based on categorization and evaluation tools. Finally, the result was the synthesis of the eligible-for-evaluation information systems and the development of a guide for choosing the appropriate information system during each step of a clinical trial; the data provided by each information system were identified. Unfortunately, some information systems and ontologies were excluded from the synthesis due to lack of information regarding the evaluation criteria. Therefore, future research should proceed with retrieving this information and developing a guide which will consider more information systems, especially for conducting observational studies.

Table of Contents

Table of Figures

Table of Tables

Introduction

Information technology is widely used in the healthcare sector and the clinical research field has made many attempts in order to be part of this innovation. New Clinical Research Information Systems (CRIS), i.e., information systems supporting the dataflow during a clinical study's process (Richesson and Andrews, 2019), are continually developed, as their obvious advantages intrigue any clinical research organization to implement them. Some of these advantages are improved data quality, low costs and more effective management. Moreover, Clinical Trial Management Systems (CTMS) are computerized systems responsible for auditing, supporting and reporting clinical trials. CTMS should be developed according to specific guidelines introduced by the U.S. Food and Drug Administration (FDA), the European Agency for the Evaluation of Medicinal Products (EMA) or each country's equivalent administration (Leroux, McBride and Gibson, 2011). The difference between CRIS and CTMS is that CTMS are used specifically in clinical trials and support the whole process of a clinical trial. A clinical research information system can be a part of a CTMS.

Clinical Information Systems (CIS), such as Electronic Health Records (EHRs) and Hospital information Systems (HIS), are systems used in a health care environment, especially a hospital. CIS aim to gather reliable information, use this information for supporting decision-making and increase and process this piece of information while it is shared between different clinical areas of expertise. Their more important goal is to optimize patient care (Geissbuhler, 1998). In many cases, clinical research organizations use Electronic Health Records (EHRs) and Hospital Information Systems (HIS) for collecting data which will be used in clinical studies. Therefore, a CIS can become part of a CTMS. Although this choice could seem efficient due to its low cost (especially in comparison with commercial CRIS), these systems do not fulfil all the prerequisites for importing and exporting data in the structure, quality and accuracy necessary for the clinical research regulatory requirements (Schreiweis et al., 2014). Therefore, these two systems cannot stand alone in managing clinical research data, but they could be the bases for creating new CRIS.

Moreover, commercial CRIS are mainly used by pharmaceutical companies. Commercial CRIS provide data capture, flow and monitoring based on Good Clinical Practice (GCP) Guidelines, automatic reporting and, in some cases, clinical trial management tools (Oliveira and Salgado, 2006). Although their advantages make them appealing, they are, also, characterized by some technical and functional problems, such as high cost (acquisition and maintenance), high training-time or even need for specialized personnel, inability to share data freely with each other (because of their patented architecture and their incompatible data) and low usability ("cumbersome user interfaces") (Oliveira and Salgado, 2006). Therefore, although pharmaceutical companies provide clinical research organizations with commercial systems, the clinical studies personnel avoid using them (Oliveira and Salgado, 2006). As the clinical studies personnel showed their preference, researchers proceeded with developing information systems for supporting clinical studies which will be more usable, scalable, accessible and less expensive. In the beginning of this development, CRIS could only support one function of the clinical research process, such as recruitment or reporting of adverse events. But as the technology evolved, more complicated and multifaceted systems were created (Richesson and Andrews, 2019). However, even though the existing CRIS might be able to support more than one function of the clinical research process, they still seem unable to support this process end to end.

This scarcity of a complete CRIS led to the goal of this thesis, which is the development of a guide that will assist clinical study management teams in choosing the appropriate CRIS for conducting a complete clinical study. First, information regarding the clinical research process and its architecture is presented. Then the dataflow in this process is pointed out. Afterwards, a literature review of 23 information systems and ontologies used in clinical research are analyzed. Finally, categorization and evaluation of these systems are conducted and the CRIS guide for carrying out a clinical study is synthesized.

Methodology

Twenty-three (23) information systems and ontologies used in clinical research were retrieved via a literature review. The search engine used was “PubMed”. The search-keywords were: (“clinical trial*” OR “clinical research”) AND (“information system*” OR ontology). The search was restricted by searching only the “Title” for the keywords in order for the results to be more specific and accurate. Moreover, in order to collect more up-to-date information, the search was restricted to papers published in a specific timeframe: 2000-2019.This search led to 379 results, but 344 were papers presenting clinical trials’ reports and therefore, they were excluded. Afterwards, the following exclusion criteria were implemented: information systems for translational science (because this thesis is not focused on that phase of a medical or pharmaceutical innovation) and information systems which consider patient satisfaction and patient-reporting (because the evaluation of the CRIS analyzed in this thesis is based on more technical characteristics). From the 35 remaining results and their references and citations, 23 information systems and ontologies were included in this thesis’ analysis and synthesis. For the evaluation and categorization of the information systems, another literature review was conducted (keywords: clinical information systems; evaluation tool; categorization), and evaluation and categorization tools were selected.

The Process of Clinical Research

The Purpose of Clinical Research

During the last decades, many innovations in the field of healthcare have been introduced and clinical trials emerged as the cornerstone of this progress. The results of clinical trials show that clinicians and healthcare professionals can use them as a vehicle for improving public health and, in the long run, the condition of healthcare systems. Due to the advantages that clinical research provides, more information regarding its purpose and process should be discussed.

Definition of Clinical Studies

Clinical studies are based on research which uses human volunteers/participants and their goal is to offer more knowledge to the medical field (ClinicalTrials.gov, 2019). Clinical studies can be observational studies and interventional studies or clinical trials.

The two main goals of *observational studies* are the testing of the allocation of predictors and outcomes in a population (descriptive) and the description of the associations between these variables (analytic) (Hulley et al., 2013). Observational studies most commonly take place in cases that an investigator cannot apply a randomized controlled clinical trial, e.g., due to ethical issues, in rare diseases, etc. (Mann, 2003). Moreover, an investigator might choose to conduct an observational study prior to a clinical trial (Hulley et al., 2013).

Interventional studies (or clinical trials) are conducted in a way that the researcher is allowed to intervene during the study (Thiese, 2014). Friedman et al. defined clinical trial as "a prospective study comparing the effects and value of intervention(s) against a control in human beings" (2015). According to Shankar et al., the purpose of conducting a clinical trial is to test if a drug or procedure is safe and effective. Moreover, the process of a clinical trial usually begins after the laboratories studies have shown promising results (2006).

For a better definition of a clinical study, its life-cycle (steps), phases and designs are presented below.

The Architecture of a Clinical Study

The life-cycle of a clinical study

According to Sim et al. (2014), a clinical study can be ideally described by five steps/phases (Fig.1): (i) refine a scientific question by reviewing and interpreting results of previous studies; (ii) design a new study; (iii) carry out the study; (iv) report results; (v) interpret the results and apply them to clinical care and policy. These steps complete a circle (life-cycle), because clinical practice (step 5) will provide information for refining a new scientific question.

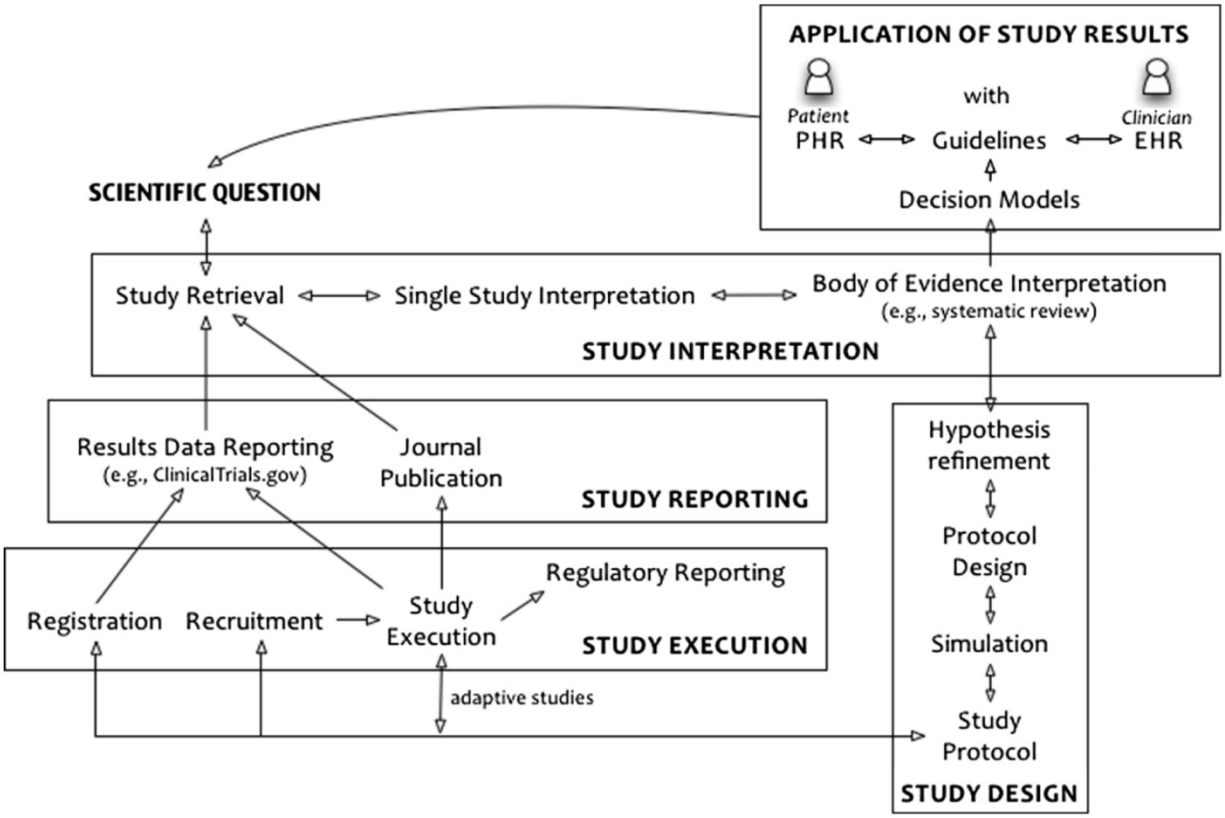

Figure 1. Idealized scientific lifecycle of a human study within a learning health system (Sim et al., 2014).

(i) Conceiving the research question

The first step of a clinical research project is the conception of the right research question, as it specifies the methodology and structure of the research study. (Thabane et al., 2008). According to Thabane et al. (2008), the best approach for conceiving the research question is based on the PICOT approach (Population, Intervention of interest, Comparator intervention, key Outcomes and Time frame) and consists of the following steps: *First*, the investigator should collect data via the scientific literature, e.g., a systematic review made in the research area of the investigator's interest. It is crucial for an investigator to be up to date with the recent discoveries and published papers. *Second*, the PICOT framework should be followed in order to identify what is missing from the existing literature, i.e., the researcher should decide on (a) the population of interest; (b) the intervention of interest; (c) the comparison with other interventions; (d) the outcome of interest; (e) the time needed for the study to be concluded. *Third*, the PICOT framework is followed in order for the research question to be appropriately modelled. *Fourth*, the investigator should estimate if the research question is characterized by the FINER criteria, i.e., the research question ought to be feasible, interesting, novel, ethical and relevant; a well-proposed research question must be characterized by these epithets (Thabane et al., 2008; Riva et al., 2012; Hulley et al., 2013).

(ii) Designing a new study

During this step, decision support tools are used, and researchers ought to make some decisions regarding the appropriate study design type (which will be discussed below), the appropriate study population and the identification of biases. As shown in Figure 1, during this step, the study protocol is composed. The planned study protocol consists of the activities that should be carried out during execution and analysis, while the executed study protocol consists of the activities that actually happened (Sim et al., 2014; Hulley et al., 2013).

The study protocol summarizes and presents the scientific clinical study design (Sim et al., 2014) and can be considered an official paper which demonstrates an agreement between the clinical investigator conducting the study, the participant/volunteer and the scientific community (Friedman et al., 2015). In other words, the study protocol can describe each step that the

investigator needs to complete in order to design a clinical study. Friedman et al. (2015) suggested the contents of a typical study protocol based on the SPIRIT 2013 Statement, which developed guidelines for a clinical trials protocol development (Chan et al., 2013). Therefore, a *clinical study protocol's outline* as proposed by Chan et al. (2013) and Friedman et al. (2015) can be:

I. Administrative information

1. Title

The title must be descriptive and ought to contain the study design, population, interventions and trial's acronym (if applicable). (Chan et al., 2013).

2. Trial registration

The trial's identifier, the trial's registry number and the "Trial Registration Data Set" (Version 1.3.1.) as defined by the World Health Organization (WHO) must be provided (Chan et al., 2013).

3. Protocol version

Contains the date and version identifier of the protocol (Chan et al., 2013).

4. Funding

Presents sources and types of any kind of support (financial, material, etc.) (Chan et al., 2013).

5. Roles and responsibilities

This part contains the names of the protocol contributors and study sponsors and the roles and responsibilities of every professional participating in the trial (Chan et al., 2013).

II. Introduction

1. Background of the study

This part provides information regarding the research question and the reasons for conceiving it, published and unpublished relevant studies and benefits and adverse events of the intervention (Chan et al., 2013).

Adverse events: According to the International Conference of Harmonisation (ICH) of technical requirements for registration of pharmaceuticals for human use, adverse event is "*any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment*",

i.e. an adverse event can be an unplanned and discouraging reaction (sign, disease, symptom) related to the administration of a treatment or a drug (1994).

2. *Objectives*

The objectives or hypotheses of the trial are specified (Chan et al., 2013).

3. *Design of the study*

In this part the study design type is provided (Chan et al., 2013) (the study design typology is analyzed below).

III. Methods

1. Study setting(s) and country(-ies) collecting data for this particular clinical trial (Chan et al., 2013).
2. Study population
    a) Inclusion criteria

Inclusion criteria issue the basic features of the target population which relate to the research question, i.e., age, nationality and race. Many changes might occur while introducing the inclusion criteria regarding the geographic and temporal features of the accessible population (Hulley et al., 2013).

    b) Exclusion criteria

Exclusion criteria define the population groups which will not participate in a study, because of their inability to adhere to the study's requirements (e.g., follow-ups) or because they are prone to developing adverse effects (Huller et al., 2013).

3. Sample size

The number of participants needed for conducting the trial and for achieving its goals/objectives is estimated (Chan et al., 2013).

4. Recruitment/Enrollment of participants
    a) Assessment of eligibility based on the selection criteria (inclusion and exclusion) (Chan et al., 2013).
    b) Baseline examination/assessment

Baseline examination is designed for evaluating a patient's eligibility for recruitment in a clinical study. This examination is the first contact between a physician and a patient/possible volunteer

and provides baseline data, which will be used in the next steps of the clinical research process Moreover, it should be mentioned that baseline characteristics are the participants' demographic characteristics and baseline measures and data. In general, baseline is the participants' initial status, before the beginning of a clinical study. Therefore, baseline measures can be blood pressure measures, cholesterol levels, etc. which were measures before the first intervention (Friedman et al., 2015).

c) Intervention allocation (e.g., randomization method)

5. Intervention(s)
    a) Interventions for every group participating are described, along with the time that these interventions were administered, the method used and the professionals administering them. This part is crucial, because it provides the ability to repeat the same intervention, if needed.
    b) Criteria for changing or terminating an intervention for a trial participant (e.g. changes in the drug dose due to the presence of adverse events).
    c) Strategies for monitoring adherence to protocol and for improving protocol adherence, if needed (Chan et al., 2013).
6. Follow-up visit description and schedule

Follow-up visit: Follow-up visits can take place after the end of the "original" clinical trial (post-trial follow-up) and can be continued for decades, as new reactions/effects to a drug or treatment can be observed many years after the trial has ended (Llewellyn-Bennett, Bowman and Bulbulia, 2016). Complete follow-up visits should be promoted (Chan et al., 2013) and follow-up reports containing follow-up data ought to be composed (ICH Expert Working Group, 1996).

7. Data collection methods

Clinical trial data entry and its quality are planned, along with the study instruments used (e.g., laboratory tests). If some data is not presented in the study protocol, instructions/reference for finding the collection forms demonstrating them can be found in the protocol (Chan et al., 2013).

8. Data Management

In this part of the study protocol, strategies for "data entry, coding, security and storage" can be found. Data quality of these processes is promoted (Chan et al., 2013).

9. Assessment of Adverse Events

    a) Type and frequency

There can be a variety of adverse events; they can be minor symptoms, medium reactions or major complications. Therefore, adverse events are usually classified based on their severity (Hulley et al., 2013).

    b) Instruments (interviews, questionnaires, etc.)

    c) Reporting

The adverse events that should be reported are unexpected adverse events, expected adverse events with an extreme increase in their rate of occurrence, an important risk (e.g. if a treatment for a life-threatening disease is proved to be inefficient), etc. (ICH Expert Working Group, 1994)

10. Data analysis/Statistical Methods

    a) Interim monitoring, including data monitoring committee role (Chan et al., 2013).

    b) Final analysis: analysis of primary and secondary outcome (Chan et al., 2013). For this analysis primary outcome/response variables will be used which will answer the research question. Moreover, secondary outcome/response variables will provide helpful and supporting information for answering the research question (Vetter and Mascha, 2017).

    c) Methods for additional analyses, if needed (Chan et al., 2013).

    d) Statistical methods for managing missing data (Chan et al., 2013).

11. Data monitoring

Monitoring of data is a crucial process during a clinical study. Every piece of information entered in the study's system ought to be checked for completeness, consistency (internal and external) and currency. If data is inconsistent, it should be corrected. The most error-prone data are the dates and times. Moreover, missing data is a typical problem and systems for minimizing this problem are necessary (Friedman et al., 2015).

A Data Monitoring Committee (DMC) is composed and information regarding its role and its relation to a sponsor (or not) is provided in the study protocol. In case DMC is not necessary for a trial, explanation regarding this decision is presented (Chan et al., 2013).

12. Auditing of the way that the clinical trial is conducted

Audits can be random routine audits, structured audits and audits based on claims about scientific misconduct (Friedman et al., 2015).

IV. Ethics and dissemination

1. Research ethics approval

Institutional Review Board (IRB) approval should be administered and therefore, plans for receiving the approval are presented in this part of the study protocol (Chan et al., 2013).

2. Communication of crucial protocol changes to relevant individuals and organizations (Chan et al., 2013).
3. Participants consent/ Informed consent

A clinical trial's participants must be aware of their voluntary participation, their role, the possibility of unknown risks and discomforts, the possible absence of benefits during and after a clinical trial and the way that the protocol works. In other words, participants ought to understand that, during a clinical trial, they are not considered a "patient" but a "subject of research". (FDA, 2018).

Individuals responsible for collecting the consent and individuals authorized to have access to participants data are presented (Chan et al., 2013).

4. Confidentiality

Ways for collecting and sharing participants data in a confidential way are described (Chan et al., 2013).

5. Declaration of interests

Principal investigators (i.e., the professionals conducting a clinical trial) ought to declare financial or other personal and competing interests for the clinical trial or the study settings (Chan et al., 2013).

6. Access to data

Professionals with access to the post-trial data are officially presented. Official agreements from investigators with limited data access are, also, presented (Chan et al., 2013).

7. Post-trial care

In case of harm, post-trial care is offered to the participants suffering (Chan et al., 2013).

8. Sharing policy

Instructions for sharing data after the end of the trial are demonstrated. In case the study protocol can be publicly published, plans for doing so are presented (Chan et al., 2013).

V. Organization

1. Participating investigators
    (a) Statistical unit or data coordinating center
    (b) Laboratories and other special units
    (c) Clinical center(s) (Chan et al., 2013).
2. Study administration
    (a) Steering committees and subcommittees
    (b) Monitoring committee
    (c) Funding organization (Chan et al., 2013).

VI. Appendices

1. Informed consent materials

Official documentation for informed consent (informed consent form) is provided to participants (FDA, 2018). This part of the protocol contains model consent forms which should be given to participants (Chan et al., 2013).

2. Biological specimens

Methods for collection, assessing and storing biological specimens for analysis are presented. These methods are important, because they can be used in current and future studies researching the same treatment or drug (Chan et al., 2013).

(iii) Study execution

The tasks included in this step are:

- Institutional Review Board (IRB) application

The Food and Drug Administration (FDA) or the European Medicines Agency (EMA) (or each country's responsible agency) establish some specific requirements regarding clinical research process and how to conduct one. Based on these requirements, an Institutional Review Board

(IRB) is authorized to approve, disapprove or modify a clinical study. The IRB's goal is to protect humans participating in clinical research (FDA, 2019).

- Set-up study registration

In order to register a clinical study, five steps ought to be followed:

a) Define the responsible individual or organization for registering the clinical study and which Protocol Registration and Results System (PRS) account should be used.
b) Be aware of submission requirements. For example, ClinicalTrials.gov allows the submission of studies researching biomedical and health outcomes.
c) Login to PRS (an account with password is used).
d) Insert the required and optional data elements, which are presented in the PRS and are related to the clinical study's objectives.
e) Preview, inspect and submit the record and then, verify that the submission is completed (ClinicalTrials.gov, 2019).

- Recruitment and enrollment of patients (according to eligibility criteria)

According to Friedman et al., "*successful recruitment depends on developing a careful plan with multiple strategies, maintaining flexibility, establishing interim goals, preparing to devote the necessary effort, and obtaining the sample size in a timely fashion*". During this process, it is important to find an adequate number of study subjects in a short period of time. The enrolled group of subjects (subject samples) is a subgroup of the study population (i.e., individuals sharing a condition or some characteristics related to the clinical study's objectives– inclusion criteria. Moreover, the success of the recruitment process is based on the systems used in order to identify and select subjects for the participants pool (2015).

- Protocol execution

Clinical investigators and other professionals working in a specific clinical study ought to follow the instructions provided by the study protocol (as presented above). Therefore, the used methodology, definitions and diagnostic criteria are predefined and specific (Friedman et al., 2015).

- Site Monitoring and reporting of adverse events (pharmacovigilance) (Sim et al., 2014).

Pharmacovigilance: According to the World Health Organization (2002), "pharmacovigilance is the science and activities relating to the detection, assessment, understanding and prevention of adverse events or any other possible drug-related problems".

(iv) Report results

Not only the results must be reported, but the complete study design that brought these results must be reported, as well. Moreover, positive and negative results ought to be reported and selective reporting should be avoided because it might lead to biased publications. For ethics purposes, this phase ought to be unbiased and transparent, as recommended by the Declaration of Helsinki (Sim et al., 2014; WHO, 2019; World Medical Association, 2001). The World Health Organization has established some requirements regarding the process of reporting clinical trials results. First, certain timeframes ought to be followed. More specifically, the first results report must be submitted within 12 months after the completion of the trial and, also, their publication ought to be public. If public access is denied, specific reasons for this decision ought to be presented and the results must become publicly available within 24 months after the completion of the trial. This report should contain main findings and key outcomes, such as adverse events, baseline characteristics, participant flow (i.e., the participants' data throughout the clinical study, ClinicalTrials.gov, 2018), primary and secondary outcome measures, etc. (WHO,2015).

(v) Interpret the results and apply them to clinical care and policy

In order for the results to be applied to clinical care and policy, they have to be effectively and correctly available. For this requirement to be successful, all study results must be collected in certain databases (e.g., journal publications or trial registers) (Sim et al., 2014) and therefore, be readily retrieved for further research and policy application. Finally, researchers should be able to orchestrate a high-quality systematic review, if needed (Sim et al., 2014). Furthermore, clinicians can follow seven steps in order to decide whether they will apply the results of a study to clinical care. These steps are:

- Understand the clinical study's hypothesis.
- Recognize biased outcomes.
- Confirm that differences between the two groups (control and treatment) are unrelated to other prognostic factors.
- Confirm that the analytical methods used were performed correctly and at an appropriate time (once or continually).
- Confirm that the results were not the product of chance variation.
- Ensure that the main result demonstrates the superior treatment based on its benefits, toxicity, cost and convenience.
- Clarify the future patients to whom this treatment/drug can be administered (Elwood, 1980).

It should be mentioned that in some clinical studies, specimens need to be stored for future use or in some cases, specimens from previous studies are used for a new one. Furthermore, most clinical studies are carried out in multi-site centers and therefore, some requirements need to be met. For example, investigators need to have access to follow-up data and patients' medical examinations. Finally, a universal medical and drug terminology is needed for minimizing the possibility of "translation" errors between multiple sites (Hulley et al., 2013).

Study design typology

Clinical studies are characterized by different study design types (Fig. 2). The selection of the study design is based on several factors: cost, access, the nature of the participant's exposure to the treatment or drug in trial, required epidemiologic measures and currently published data related to whatever is researched (Thiese, 2014). The two main categories are observational and interventional studies.

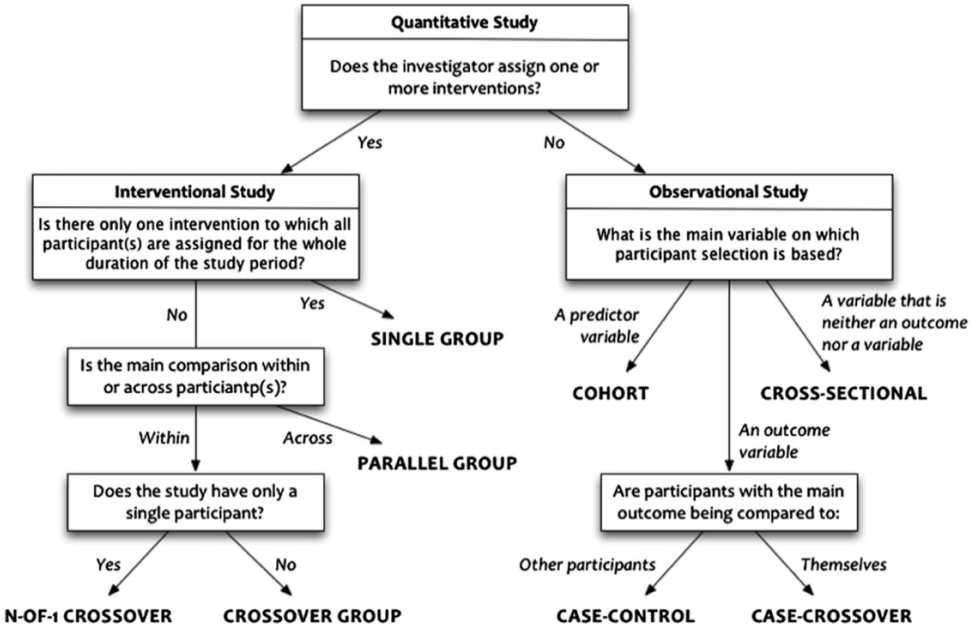

Figure 2. The study design typology (Sim et al., 2014).

Clinical research phases

Clinical studies can also be described by four phases; the quality of every phase depends on the quality of the previous one and influences the quality of the next one (Friedman et al., 2015).

- *Phase I Studies*

According to the Food and Drug Administration (FDA), during phase I studies, a new drug or treatment is tested for the first time in a small group of people (2019). For the first phase of a clinical trial, healthy volunteers should be recruited. Another option is the recruitment of patients who have the condition tested and have unsuccessfully tried out the already known and used treatments. During phase I, researchers aim to find more information over the biopharmaceutical aspect of the new treatment/drug, e.g., estimations and data regarding pharmacokinetics, pharmacodynamics, tolerability, bioavailability, body distribution and feasibility and safety of the delivery systems used. However, the most important finding which ought to be discovered during phase I is the "maximally tolerated dose". The amount of drug safely distributed in the body is controlled and decided with several methods (Friedman et al., 2015).

- *Phase II Studies*

In phase II, the group of people participating in the study increases. The criteria used for selecting participants are highly specific and the whole process is done with excessive attention to detail. During this phase, the appropriate dosing levels of a drug are established and the effectiveness of a drug is tested and therefore, researchers decide if the development of this treatment (or drug or device) will continue. In other words, the aim is to assess whether phase III has the prospect to be successful or not; this success is the outcome of many factors, e.g., beneficial and adverse events (Friedman et al., 2015).

- *Phase III/IV Studies*

The goal during Phase III is the estimation of the benefit that the treatment in trial will provide to clinical practice and the discovery of the adverse effects occurred by this treatment. Moreover, risks of the treatment or drug in trial are evaluated and compared with its benefits and effective dosing levels are established. The duration of phase III clinical studies may be long. Researchers should consider the importance of the follow-ups after the end of this phase; although many tests occur during a trial, a treatment should be investigated for many years afterwards. This follow-up investigation is called Phase IV and a large population participates (Friedman et al., 2015). Phase IV studies are post-marketing studies and the drug or treatment tested is already approved by the FDA, European Medicines Agency (EMA) or each country's responsible agency (FDA, 2019). The outcome from a phase IV study will be more accurate because of larger populations, actual dosing, longer exposure and long-term follow-ups. Usually, phase IV studies are observational studies (Antoniades et al., 2012).

Information Needs in Clinical Research

Clinical studies are considered to be information-intensive processes, as much information is necessary for the study protocol to be designed and executed. However, attention to the amount of information collected should be drawn, as unnecessary data is usually collected and therefore, the cost of a clinical trial increases without the corresponding benefits. Moreover, excessive data collection can lead to a decrease in data quality (Friedman et al., 2015).

According to Tran et al. (2011), the clinical research data life-cycle includes three phases:

- *Data specification*: The method(s) for storing and collecting data is determined.
- *Data collection* via data collection forms, such as the Case Report Form (CRF). These forms aim to capture information regarding a patient's eligibility, status, medical history, biochemical data, etc.

A Case Report Form (CRF) is a significant document used in a clinical research process. The data collected in this form is patient data (Bellary, Krishnankutty and Latha, 2014). According to the Guideline for Good Clinical Practice, a CRF is "*a printed, optical, or electronic document designed to record all of the protocol required information to be reported to the sponsor on each trial subject*" (ICH Expert Working Group, 1996).

- *Data reviewing, reporting and analysis*: data entry is checked, and the study's results are reported and analyzed. Moreover, real-time collected data, such as adverse events, are reported (Tran et al., 2011).

Based on the protocol's needs, Friedman et al. suggested that the data needed for every clinical study should include:

- Baseline information (e.g., selection criteria for determining the study population)
- Measures of the participants' adherence to the interventions being tested
- Crucial concomitant interventions, which refers to other treatments or medications than the one in trial that the participants are submitted to (e.g. concomitant medications during the trial).

- Primary outcome/response variables
- Secondary outcome/response variables (only the important ones)
- Adverse events (2015).

In order for the data collection process to be successful, several sources ought to participate in it, such as Electronic Health Records (EHR), paper patient charts, investigator and patient report outcomes. Hence, problems during this process can arise. The four basic problems in data collection are: missing data, incorrect data, excess variability and delayed submission (Friedman et al., 2015). One of the solutions to these problems, as proposed by Friedman et al. (2015), is the "Electronic Source Data"; data should be stored in electronic health information systems, which will have the ability to integrate. Via this solution, data from various sources will be collected, integrated and managed more successfully.

Gouveia-Oliveira and Salgado suggested that the data needed in a clinical study are related to:

- Research plan (or study protocol)
- Clinical Trial Monitoring
- Laboratory results
- Treatment allocation
- Study variables
- Baseline variables
- Adverse events
- Outcome
- Case Report Form (CRF)
- Lack of adherence (or discontinuation)
- Protocol deviations
- Concomitant treatment/medication (1999; 2006).

Clinical Trial Monitoring: Clinical trial monitoring goals are to confirm that human rights are respected, the reported data are complete and accurate and the study is being conducted according to the protocol's instructions, Standard Operating Procedures (SOP), Good Clinical

Practice (GCP) and the regulatory requirements. Clinical trial monitoring includes study/site/monitoring visits (ICH-GCP, 2019; European Medicines Agency, 2006).

Study visits are conducted by an investigator or a clinical research coordinator (CRC). During the first study visit, the participant is usually informed for the clinical study's characteristics and requirements and signs the Informed Consent Form. Moreover, selection criteria, demographics, medical history, concomitant medications, etc. are checked. During the next study visits, vital signs, adverse events and blood examinations are usually checked. More generally, during a study visit, the investigator or CRC checks for anything important for the clinical study's process and outcome. Furthermore, the investigator or CRC overviews the study and confirms that the protocol is being followed and that the personnel is correctly trained according to the protocol's requirements (Friedman et al., 2015; ICH-GCP, 2019).

Considering the information above regarding the process of observational studies and clinical trials and the information needs as suggested by Tran et al. (2011), Friedman et al. (2015) and Gouveia-Oliveira and Salgado (1999; 2006), conclusions over the dataflow in this process can be drawn. The data needed in every clinical research process can be divided in the following categories:

a) data for selection criteria, sampling and recruitment
b) data from stored specimens, images, etc. from previous studies
c) data for adverse events, especially in multi-centered trials
d) data regarding medical clinical terms and drug terms used globally (standardized "dictionaries")
e) data found in publications regarding the trial
f) data from the process and outcomes of old trials
g) follow-up data
h) data from the participants' medical examinations during the trial
i) data for the statistical analysis (integrated information might be needed).

As patient privacy and safety is of great importance, all data ought to be managed and accessed according to legal and ethical requirements (ICH Expert Working Group, 1996). Moreover, during

a clinical trial, information is shared between many procedures, organizations and people of interest, e.g., investigators, Institutional Review Boards (IRB), monitoring visits, etc. (Gouveia-Oliveira and Salgado, 1999). Furthermore, according to Gouveia-Oliveira and Salgado, clinical trial monitoring includes detailed documentation regarding Good Clinical Practice (GCP) guidelines and study status, records of compliance to the protocol and administrative procedures (1999) and finally, the existence of patients' Case Report Forms (CRFs), which are developed based on the trial's protocol and present all the data recorded and collected during the study (trial activity information) (Salgado and Gouveia-Oliveira, 2000).

For all this information to be collected and stored, Clinical Trial Management Systems (CTMS) ought to be used. CTMSs support the clinical research process by collecting and integrating data from several sources; therefore, the number of errors during a clinical study is decreased and communication between clinical research professionals (and participants) is enhanced (Park et al., 2018). According to Park et al., CTMSs are responsible for "inputting data, receiving an interface through a different system and automatically calculating" (2018).

Therefore, an information system which will be used in clinical research should be able to input, analyze, store and report the clinical study's data presented above. If the Clinical Research Information System (CRIS) cannot support the clinical research information needs, then it should be altered, integrated or improved.

Information Systems in Clinical Research

According to Richesson and Andrews (2019), *"Clinical Research Information Systems (CRISs) are software applications intended to handle one or more aspects of supporting clinical research".* Nowadays, many CRISs are able to accommodate more than one aspect of clinical trials (Richesson and Andrews, 2019). However, despite this progress (in the past a CRIS could only support one aspect of a clinical study), in order for a clinical process to be completed, more than one information systems might be needed. In this chapter, information systems used in clinical research are presented and analyzed based on the information needs in clinical research, as presented in the previous chapter.

- **Tools in a clinical information system supporting clinical trials**

Weisskopf, Bucklar and Blaser (2014) proposed that in order for a clinical information system to support clinical trials, some tools should be developed and applied. These tools are:

- *Clinical information system implemented trial registry and patient-trial-assignment*

Clinical trials ought to be characterized by: study title (complete, short), start date, end date (planned), Institutional Review Board (IRB) reference number, principal investigator responsible, contact person, alert at hospitalization, short description of the trial, link to the trial, link to trial documents, inclusion and exclusion criteria, attachment container for PDF files. Regarding the patient-trial-assignment part, in a scenario that a patient is enrolled in another trial, a notification appears.

- *Medical record templates for trial documentation*

For sufficient documentation, the clinical information system provides a trial note form, a screening visit form, a follow-up visit form and a form for confirmation that the participant's best interest is sought (the form of confirmation is provided by an independent physician). Data from

the registry is integrated in the record templates. The reports are created in a flexible way and therefore, they can be used in all types of trials. Moreover, the drug name is reported and complete information regarding this drug is included in the report.

- *Trial and trial subject queries*

Queries regarding the trial can be performed by the management and participants lists of all trials, including the participants medical information, can be found via a trial subject query.

- *Access to electronic medical records*

This access is authorized and only care providers can access the system with their name and password. Many levels of access are available, and each level is role-based.

- *Hospital admission alerts*

If a non-authorized clinician proceeds with a change in a patient's chart, a notification will be sent, because only specific personnel can have full access to a patient's chart. This notification system is important for the patient's privacy, but also for his/her safety, because a change in the chart might be the introduction of an adverse event.

- *Trial feasibility checks and subject recruitment*

Feasibility checks for the assessment of the number of the eligible patients and subject recruitment are accomplished by performing queries in the clinical information system and the clinical data warehouse.

- *Order sets*

Many types of orders are included in these sets. The orders can be simple tasks or complex ones. The mix of tasks in each order set can be very detailed or vague and can relate to one another. The order sets are named based on the trial or study visit (Weisskopf, Bucklar and Blaser, 2014).

After the implementation of these tools, a clinical information system should be able to support clinical research. Even though results of the study show that the number of users of this system increases, this tool-based clinical information system lacks in scalability. The reason for this occurrence is that unless hospitals use the same clinical information system as Weisskopf, Bucklar and Blaser used to demonstrate these tools, this tool-based clinical information system is not always compatible. Moreover, this tool-based CIS is not automated and patient data safety and quality need evaluation (Weisskopf, Bucklar and Blaser, 2014). This tool-based information system will be able to support data for recruitment, data for adverse events, data from stored specimens, images, etc., data from the medical examinations during the trial and follow-up data.

- **Integration of a Clinical Data Warehouse (CDW) and a Clinical Information System (CIS)**

Clinical Data Warehouses (CDW) emerged to be extremely useful in the process of integrating heterogenous data from different data sources. However, CDWs are not widely used in the healthcare sector and therefore, Zapletal et al. (2010) suggested a method for applying a CDW integrated with a hospital's Clinical Information System (CIS). For developing the CDW framework, four domains were considered important:

- *The technical realm:* The process of data anonymization was the first principle guaranteed; the anonymization should be able to be reversed. Structured data were extracted from the Hospital Information System, while unstructured data were extracted from the Electronic Health Record (EHR) database.
- *The data realm:* After security was guaranteed, real patient data were integrated into the CDW in three steps: (i) Identification of data sources, (ii) Mapping of data sources into the target schema (target schema: PATIENT, PROVIDER, VISIT, CONCEPT AND OBSERVATION), and (iii) Building of dedicated datamarts (for users who need a higher level of data granularity).
- *The restitution realm:* Users with different profiles are able to interact with other components of the CIS.

- *The administration realm:* The hospital's Help Desk became responsible for more functionalities, such as the ability to register for a dashboard creation (Zapletal et al., 2010).

According to Zapletal et al. (2010), the CIS and the CDW should work together and in full collaboration; the ClS is responsible for providing the data and the CDW is responsible for using this information in order to create new data, such as reports. As Zapletal et al. claim, ultimately, this method will benefit: "quality management for auditing or for outcome studies, population follow-up, clinical investigations or cases studies and intervention studies (before/after studies or controlled trials)" (2010).

- **A Database System** for integrating Clinical Trial Management, Control, Statistical Analysis and ICH-Compliant Reporting

According to Salgado and Gouveia-Oliveira (2000), a clinical trials cycle consists of Patient Screening and Registration, Protocol Implementation, Trial Monitoring, Analysis and Publication phase. Based on that opinion, Gouveia-Oliveira and Salgado (1999) developed a system responding to the International Conference of Harmonization (ICH) guidelines, which includes:

- COATI (Control, Assessment and Tracking of Therapeutic Investigations): a system built for the monitoring and management of clinical trials. This system/model is based on patient's study visits and is time-oriented (Fig. 3). As the system's database contains all the appropriate data for a study (as presented in the second chapter), analytical tools (e.g., DART and PANDA) can be used in order to introduce more applications to the system (Oliveira and Salgado, 2006).
- Synthesis of the several ontologies needed for coding the outcomes (ICD for related diseases and medications (WHO, 2018), WHO-ART for adverse events (WHO, 1984)) (Oliveira and Salgado, 1999).
- Data Dictionary (Oliveira and Salgado, 1999).

- Statistical Analysis Program (PANDA, DAta ANalysis Package) – PANDA selects patients and observations, categorizes the information in the central database, develops data marts and creates virtual forms (e.g., tables) which will be used for study reports. “PANDA will lead the way for research on stored data in less time and with less effort” (Gouveia-Oliveira and Salgado, 1999).
- Report generator used for controlling the trial and the reports needed during its process (DART, Data Analysis and Reporting Tool) – DART imports data from PANDA, decides over the statistical analysis plan and exports its output to PANDA for introducing the final report (Gouveia-Oliveira and Salgado, 1999).
- A clinical trials reporting system. Reports regarding the participants adherence or withdrawal, monitoring visits to the clinical trial’s sites, laboratory results and the study’s outcomes are necessary.
- A query generation and management system (Q-GEM) – Q-GEM is a software module which will allow clinical trials data to be invalid. This parameter is important, because the data errors ought to be detected after they have been imported into the system and issued in data discrepancy forms (DDF) or data clarification forms (DCF) Gouveia-Oliveira and Salgado, 2006). The DDFs/DCFs are documents which include questions regarding identified discrepancies in a clinical study (Krishnankutty et al., 2012) and are always accompanied by a text-summary regarding the detected problem. Manual intervention is permitted whenever a user notices an error not detected by the system. First, the DDFs/DCFs are organized by patient and study site, then PDFs are developed and sent via email to the study sites or clinical monitors for finding a solution to the problem and finally, the user is able to reach directly from the DDF to the respective screen form via hyperlinks from Q-GEM to COATI (Gouveia-Oliveira and Salgado, 2006).

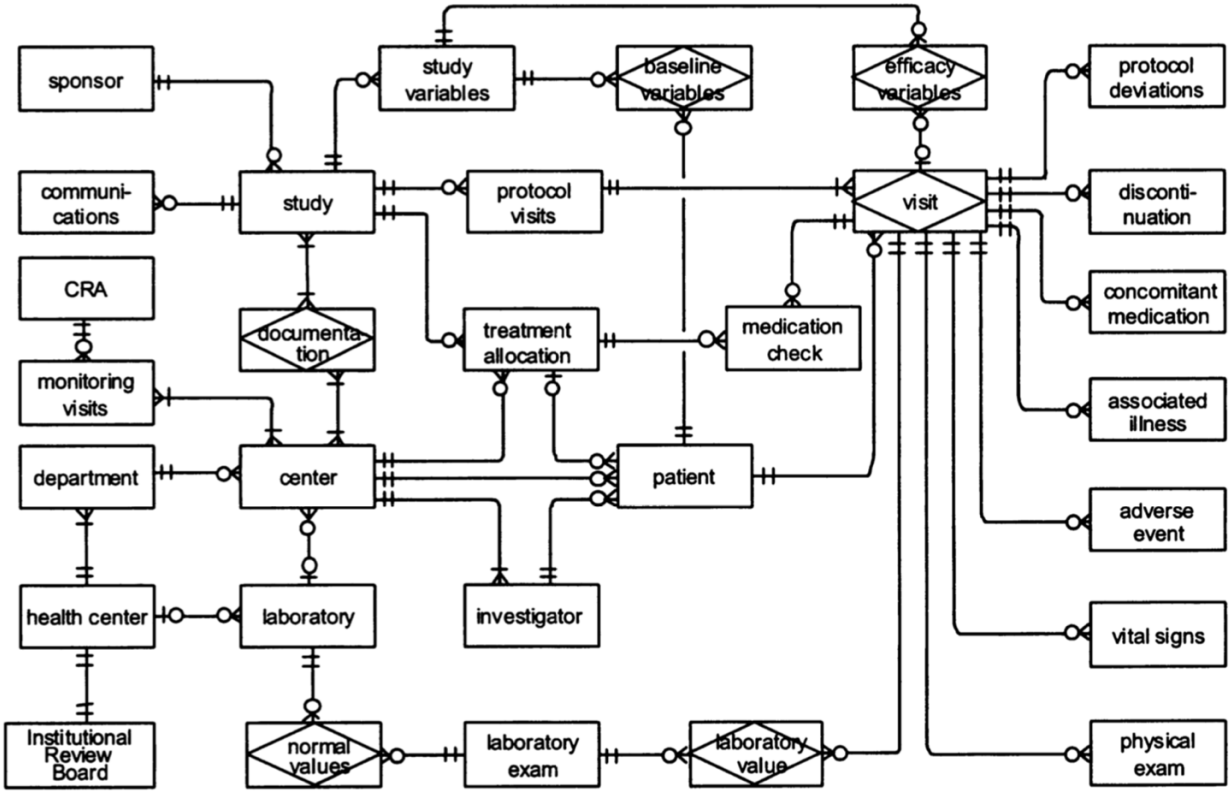

Figure 3. Top-level Entity-Relationship diagram of the data model of a multiple clinical trials management system (COATI) (Oliveira and Salgado, 1999).

Figure 3 represents an Entity-Relationship diagram in which rectangles represent the entities, arcs represent associations, a single dash near the entity represents a one-to-one association, a circle represents lack of association and a three branched arch represents a one-to-many association (Oliveira and Salgado, 1999).

The entities presented are common in all clinical trials, such as previous illness and adverse events. This system is based on a study-centric approach (i.e., process-oriented approach) instead of a data-centric approach. Hence, it can lead to the development of a generic database which will support the data management of any clinical trial. In order for this to happen, each study design ought to be modeled into the system via some parameters which encode the characteristics of the study design, the research plan and the statistical design. These characteristics can be found in the clinical trial protocol and the CRF. The final outcome was that the system could provide information regarding more aspects of a trial, such as eligibility

verification, data-entry, document management, clinical trial monitoring and site management, etc. (Oliveira and Salgado, 2006).

Via this system, the arrangement of a new trial requires an average of 30-45 minutes; what is needed is the definition of the study plan, the eligibility criteria, the study's medications and the baseline and efficacy/outcome variables (Oliveira and Salgado, 1999).

Information for every patient is available in this system; the format used is similar to the case report forms. Afterwards, data collected during the trial for each patient (adverse events, laboratory results, adherence to the treatment, monitoring reports etc.) is registered in the system (Oliveira and Salgado, 1999).

Other features of COATI are related to identification of suitable patients for a trial (based on eligibility criteria), centralized patient registration and randomization and data-entry from remote areas/organizations. Moreover, a great advantage of this system is the access to study reports online in real-time (Oliveira and Salgado, 1999).

In 2000, Salgado and Oliveira, mentioned additional information needed during a trial. For example, the system they proposed should be able to distinguish between continuous and categorical variables and between different data analysis statistical methods, recognize the validity of study visits regarding efficacy and safety analyses and decide over the suitability of participants based on eligibility criteria and statistical methods.

- **HIS-based patient recruitment:** Workflow to improve patient recruitment for clinical trials within hospital information systems.

Patient recruitment is extremely important in the clinical research process. Despite the development of several information systems, eligible patients might not be recruited due to the study professionals' vast amount of responsibilities during patient care. Moreover, some participants tend to withdraw while the study is still in progress. Dugas et al. (2008) identified this problem and suggested the development of a workflow to improve the patient recruitment step in clinical trials, while using the already existing hospital information system (HIS) (Fig. 4).

According to their workflow, a database query (regarding admission, primary and secondary diagnosis, patient age and gender and routine lab values) is created using a HIS's function: the report generator. This query is executed frequently, sometimes once per day, and its outcome is compared with the outcomes of prior queries. Afterwards, an automated email is sent to the study nurses and physicians as a notification in case a patient is distinguished as a potential trial subject. The email does not contain any information regarding the patient's name, for privacy matters. As instructed by the email, the study professional ought to fully access the system and select any additional eligibility information; the study professional must be authorized to do so. If the patient is eventually considered an eligible participant, s/he is contacted by a study physician in order to provide his/her full consent. The actual inclusion and exclusion criteria for each patient should be recorded and a trial management tool is responsible to arrange each clinical trial's technical parameters (trial title, data query for each trial, contact information for email notification) (Dugas et al., 2008).

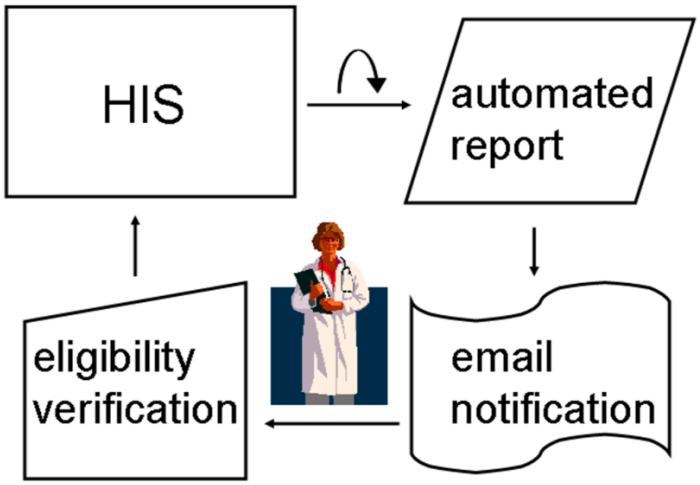

Figure 4. HIS-based patient recruitment: an overview (Dugas et al., 2008).

The results of this case study seem positive. Not only eligible patients were not neglected, but the training process of the new workflow was quick, because the users were already familiar with the HIS's functions. Finally, the health information system enabled study professionals to preselect patients to participate in specific trials according to their disease status and individual features (Dugas et al., 2008).

- **Single Source Information System**

Even though many proposals have been published for using Hospital Information Systems (HIS) and their data in order to enrich clinical research data, the information system used is still a dual source one (Fig. 5). In a dual source information system, patient care data is collected within the Hospital Information System and clinical research data is collected separately in electronic Case Report Forms (eCRF) (Dugas et al., 2009).

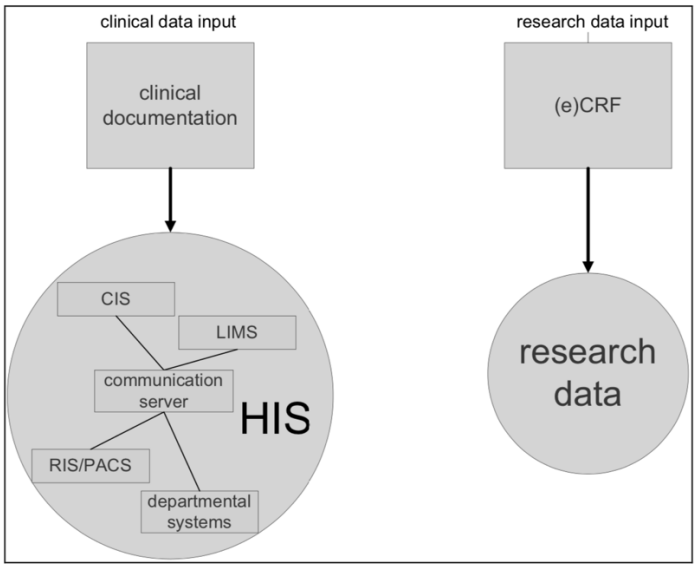

Figure 5. Dual source information systems architecture (Dugas et al., 2009).

As shown in Figure 5, HISs contain information from Clinical Information Systems (CIS), Laboratory Information Management Systems (LIMS), Radiologic Information Systems (RIS), etc., while research data are collected via electronic Case Report Forms (CRFs). This system leads to a vast amount of paperwork and documentation tasks and therefore, Dugas et al. (2009) proposed that this system should be developed as a single source information system (Fig. 6). In this system, patient care data and clinical research data are both collected in the HIS. However, it is important to mention that clinical research data can only be exported from HIS to a different research database and only the study data management team will have access to this data. This restriction is important, because patient data privacy and security are extremely crucial in a clinical research process. Therefore, Dugas et al. (2009) tested their proposed system in an

observational study, because in this kind of study, documentation procedures do not need validation. Moreover, they tested their system in leukemia trials, but only for patient recruitment and only after being approved by the data protection officer (Dugas et al., 2009).

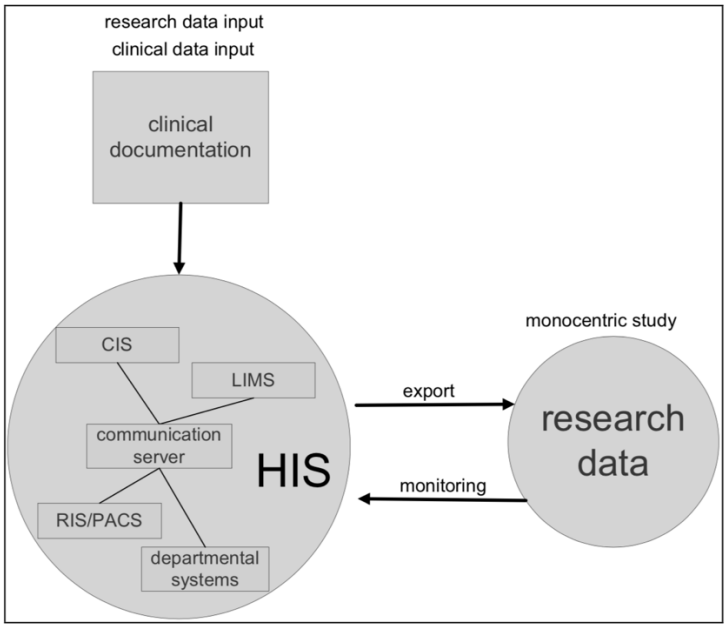

Figure 6. Single source information system architecture (Dugas et al., 2009).

To sum up, this single source information system might emerge to be a very useful tool for the clinical research field. However, patient data privacy, security and quality are important factors in every information system in clinical research and therefore, many changes ought to be made and many approvals to be introduced. Moreover, changes in the HIS in order to accommodate more functionalities should be made (Dugas et al., 2009).

- **Clinical Research Data Warehouse (CRDW)**

Geibel et al. (2015) suggested the Clinical Research Data Warehouse (CRDW) system. As shown in Figure 7, routine data (such as patient data, laboratory results, coded information on diagnoses and procedures) are exacted from the Clinical Information System (CIS) and are unified with patient information. Via an ETL (extract, transform, load) process, the integrated data are loaded into the Clinical Data Storage (CDS), a.k.a., the data warehouse. The ETL process offers three

more advantages: it can extract information from unstructured data, pseudonymize structured and unstructured data and handle information updates. Information in text-form ought to be extracted and therefore, CRDW uses a *linguistic pipeline* to perform this requirement. First, elements from diverse providers must be combined and secondly, the phrases selected from the first step ought to be linked to medical terms ("concept mapping") in a semantic knowledge base and be organized into classes. An advantage of using semantic knowledge is allowing queries for synonymous terms. Moreover, this system enables the user to utilize the web in order to gain access to the database; therefore, clinical trials and their criteria (inclusion and exclusion) are described together and former and current patient eligibility is presented together. Regarding the clinical trials criteria, it is important for them to be translated into the system's language (terminology and structure) and, also, they will eventually pair with the structured patient data and the medical data provided by the linguistic pipeline. According to the case study's evaluation, not many patients were missed during screening; the nurse does not have to check every patient's eligibility status because this is CRDW's responsibility (Geibel et al., 2015).

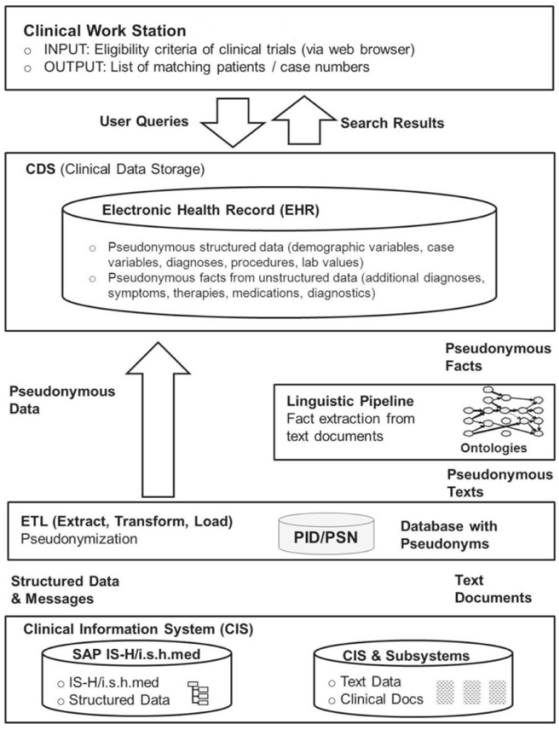

Figure 7. Simplified architecture of the CRDW system (Geibel et al., 2015).

- **ISCO: Information System for Clinical Organizations** (Development of the hospital information system)

Bailey and Urqhart identified the information needs in a clinical trials unit in a hospital in Whales and provided solutions for improving the clinical trials dataflow process (Table 1). The solutions were based on changes and additions in the already existing hospital information system (ISCO). This choice was made because of the ability to use ISCO in multisite trials and the lower cost and training-time needed for this development in comparison with other proposed systems, such as a new stand-alone database. Some of these changes were the development of a tool for automatically importing results into ISCO, the inclusion of notes for nurses for every contact (for better auditing of clinical trials information), inclusion of activity reports, inclusion of trial information into ISCO, inclusion of "required pages" for the data which ought to be collected, etc. The proposed information system will improve trials information availability and clinical trials accrual (2003).

Table 1. Identified problems and their resolutions (Bailey and Urqhart, 2003).

Problem	*Resolution*	*Possible using ISCO*	
Trial nursing notes collected manually	Include new 'page' in ISCO for visit notes	Y	
Difficulty obtaining investigation results	Automatic import of results into ISCO	Y	
Obtaining specific trial information for audit, e.g. number of research nurse patient contacts	Include nursing notes for each contact in ISCO	Y	
Obtaining tumour measurements	Include in radiology system and import into ISCO or enter direct into ISCO	Y	If included in radiology system
Trial activity data collected manually	Include standard activity reports	Y	
Availability of trial information sheets and the use of them on the wards, etc.	Include in ISCO or other hospital wide information system	Y	
For non-trials staff knowing whether a patient is on a trial	Patient flagged as on a trial in ISCO	Y	
Easy access to appropriate information out of hours and knowing which trials nurse to contact	Include trial and contact information in ISCO/hospital wide information system	Y	
The need for more (and more complete) data collection in ISCO e.g. surgery, outcomes	Inclusion of required 'pages'/ISCO training, increasing mandatory data items	Y	Requires consultant input

The data included in ISCO were expressed in a specific data dictionary of items, which is based on ISCO standards, the NHS Data Dictionary (NHS, 2019), the Welsh NHS Data Dictionary (NHS Wales, 2017) and the entity-relationship model (ERM) in Figure 8. A trial's information can be categorized in two types: patient-specific data (e.g., study number, reason for not enrolling, reason for withdrawal, date registered/randomized, etc.) and trial-specific data (e.g., data approved by ethics committee, title, acronym, etc.). The latter can be categorized as trial documentation (stored in HTML) and data required for trial management (stored in ISCO). Afterwards, each user requirement was added in ISCO according to prioritizing criteria, such as the urgency and the workload for this addition. A crucial step was the establishment of reporting, because data warehouse prototypes were already being built. The entity relationship model (ERM) in Figure 8 presents the required reports and data items and the ISCO structure (Bailey and Urqhart, 2003).

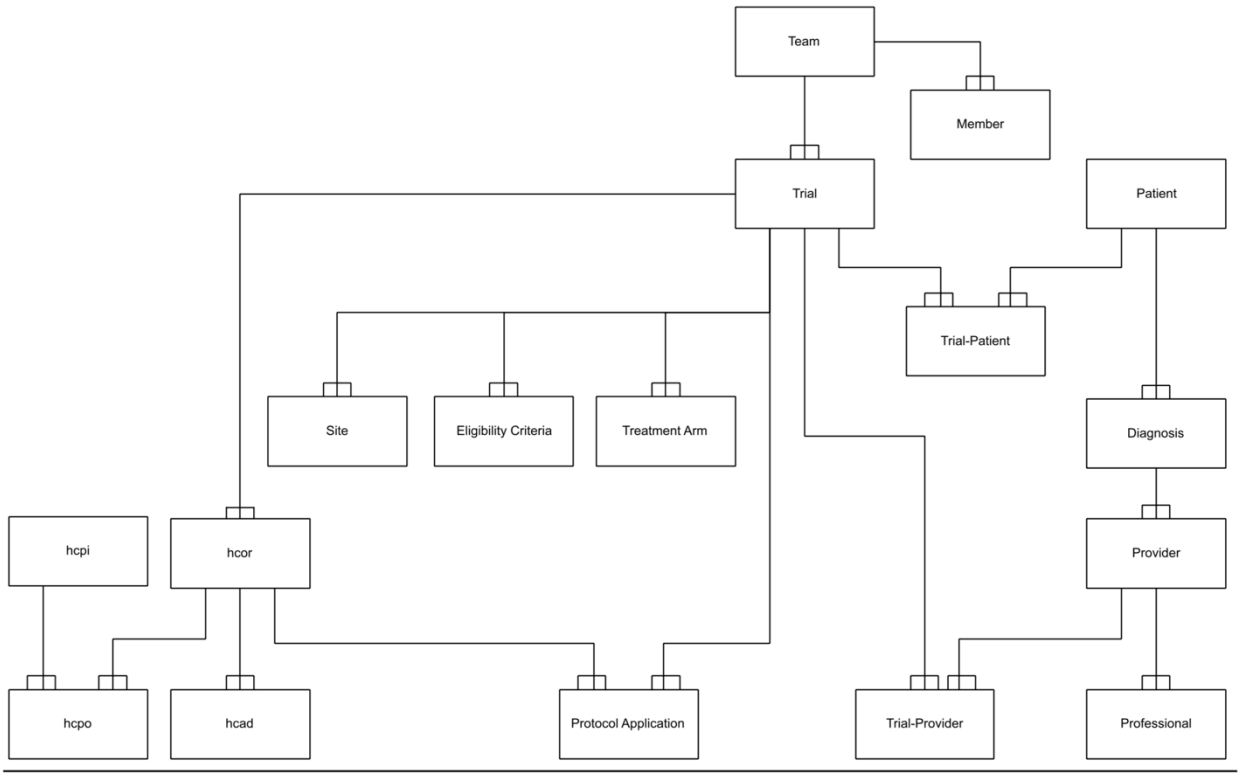

Figure 8. Entity relationship model for clinical trials in ISCO (Bailey and Urqhart, 2003).

[hcor (healthcare organization), hcpo (healthcare professional-organization link table) and hcpi (healthcare professional) tables were included for storing addresses and coding information for healthcare professionals and organizations (Bailey and Urqhart, 2003).]

For the implementation and evaluation of this information system, prototype data entry screens were designed and approved by users, tables within ISCO and trial documentation in HTML were introduced, retrieval of data entry and trial documentation was arranged, user approval was examined and evaluation after the implementation (and benefit analysis) was carried out (Bailey and Urqhart, 2003).

Unfortunately, this suggested information system cannot easily be generalized to other hospitals, as not all clinical trials units are located in specialized hospitals and some of them need administrative data management for multisite clinical studies and therefore, more advanced information systems (Bailey and Urqhart, 2003). However, the inclusions made within ISCO represent some information needs that many clinical studies have. For example, information for the clinical studies being conducted in a hospital is vital for their successful outcome. Moreover, data from study visits should be electronically collected in every clinical study, because manually created documents can be lost or incomplete. Therefore, the information system suggested by Bailey and Urqhart (2003) seems to be very useful for meeting the clinical research information needs.

- **Adverse Event Data Management System (AEDAMS)**

As mentioned above, “pharmacovigilance is the science and activities relating to the detection, assessment, understanding and prevention of adverse events or any other possible drug-related problems” and many regulations and requirements have been established in order to audit pharmacovigilance, especially in clinical studies (WHO, 2002). Richesson et al. (2008) recognized the information needs regarding adverse events (AEs) spawned during clinical trials. While aiming to the correct and trustworthy audit of adverse events in clinical trials, Richesson et al. (2008) introduced the Adverse Event Data Management System (AEDAMS). AEDAMS offers many advantages to the clinical research process, such as a more standardized procedure of reporting noted adverse events in different sites during a multi-sited study and the customization of unique

protocols, while saving time (reporting, transferring to other sites and training time) and money and succeeding in quality controls. This automated approach can be efficient and easily adaptable to different study designs, trial phases and disease areas. This information system is able to collect, track, manage and communicate data regarding AEs. The AEDAMS can be described by its three components (Fig. 9):

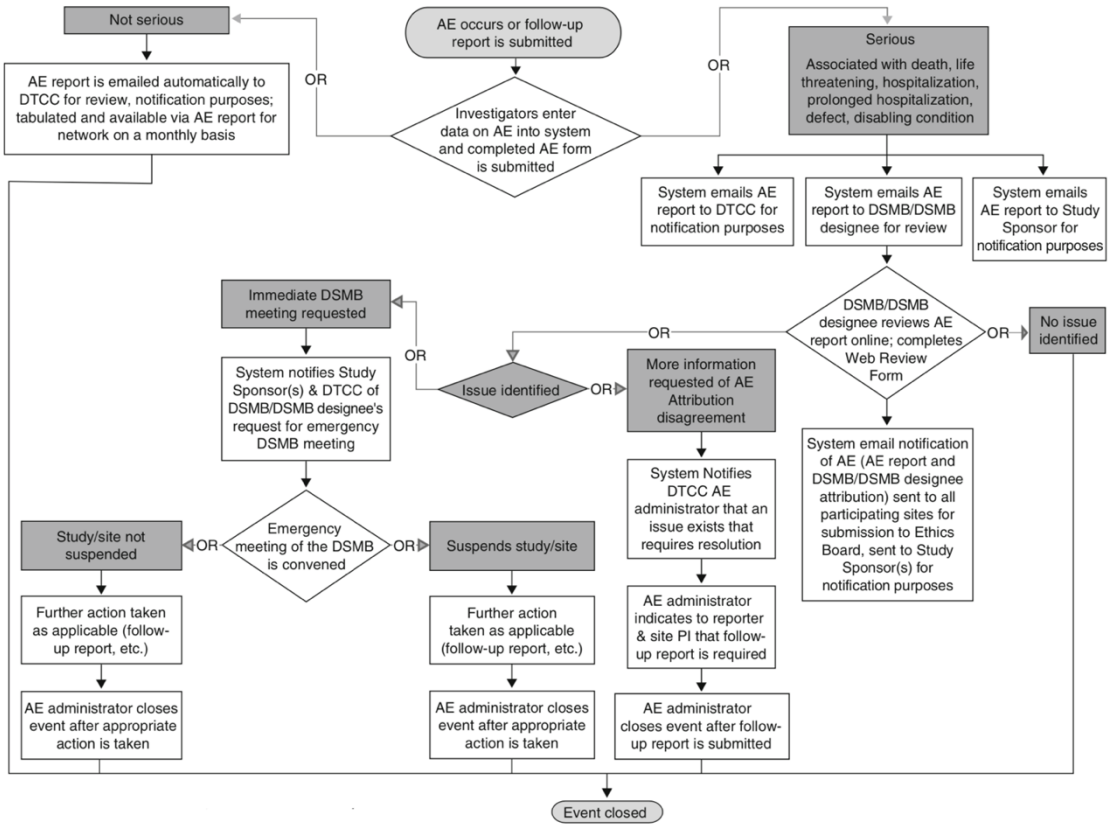

Figure 9. Flowchart of adverse events (AEs) handling in Adverse Events Data Management System (AEDAMS) applied to a clinical research network. DSMB: Data Safety and Monitoring Board; DTCC: Data and Technology Coordinating Center; PI: Principal Investigator (Richesson et al., 2008).

i. Administration and Configuration of the System

The system consists of three main roles:

- the AE administrator, who is responsible for the system's assignments, notifications and access validations,
- the AE reporter, who is responsible for allowing the appropriate research staff to report and update the system with AE information,

- and the AE reviewer (medical monitor), who is informed about reported AEs and any updates regarding AE information; s/he, also, reviews all the data in order to introduce comments and recommendations for a study's course (i.e., changes or even termination).

The AE administrator is offered the flexibility to interfere manually to the assignment of the AE reporter and reviewer roles and change their responsibilities correspondingly to their availability, schedule and possible delay (Richesson et al., 2008).

ii. AE reporting

If an adverse event occurs, the AE reporter accesses the patient's study data and reports the adverse event in a standardized reporting form. The AE reporter provides information regarding the noted AE and either submits the form or saves that piece of information and comes back to it later for editing the data. Because of this function, research staff can make a note on an adverse event and wait until their suspicions are ascertained to declare them. In the above-mentioned form, AEs are categorized corresponding to their origin/field and severity and information for any changes in the AEs report is included; if an AE reviewer edits a reported event, then the reasons for each change and any additional information ought to be presented in the original report (Richesson et al., 2008).

iii. AE review – Monitoring

The review criteria used from AEDAMS for reported AEs are seriousness and expectedness. Information for each patient is presented as a whole and each reported AE is accompanied with a status, such as open events (not assigned yet), events awaiting review (assigned, but not reviewed) and closed events (reviewed). The system offers the convenience of selecting a separate tab button for viewing each status category. After the reviewer has completed his/her part (e.g., assign causality: define the reason for the appearance of an adverse event), the final review will be checked by individuals with the responsibility of the "notification after review" role, such as the study chair, site investigators, study sponsors, etc. AEDAMS secures its database and enables the auditing of any new data related to reported AEs (Richesson et al., 2008).

Finally, Richesson et al. (2008) also mentioned that AE management systems, such as AEDAMS, can be used as electronic records and provide electronic case report forms (e-CRFs). This function will lead to a more efficient audit of data entry, communication and updates.

- **Active Computerized Pharmacovigilance** using NLP, Statistics and EHRs (Drug-AE detection system).

Wang et al. (2009) proposed a system that can detect new adverse events. In comparison with Richesson et al. (2008), this system was based on the already existing Electronic Health Records. As they mention in their article, there are some crucial adverse events (AEs) which are usually found in the unstructured (narrative) parts of EHR reports. After realizing this fact, Wang et al. (2009) developed a framework which will provide automated active pharmacovigilance and is based on natural language processing (NLP), statistics and unstructured clinical data found in EHR systems. Their framework (Fig. 10) can be divided into five phases and its final goal is to detect any drug-AE associations from narrative reports.

i. Phase 1: Data Collection

In the beginning, narrative reports, coded laboratory data and pharmaceutical orders are collected (structured and unstructured data should be collected).

ii. Phase 2: Data Extraction

The next step is to encode clinical entities found in the collected reports. For this procedure, NLP is used. More accurately, in order to analyze discharge summaries and present them as structured data the MedLEE system is used.

iii. Phase 3: Data Selection

For selecting entities which can possibly be AEs, the UMLS codes extracted in phase 2 correspond to the following semantic classes: Finding, Disease or Symptom, Mental or behavioral dysfunction, Sign or symptom and Neoplastic process.

Correspondingly, for selecting medication entities the semantic classes are: Pharmacologic Substance, Antibiotic and Clinical Drug.

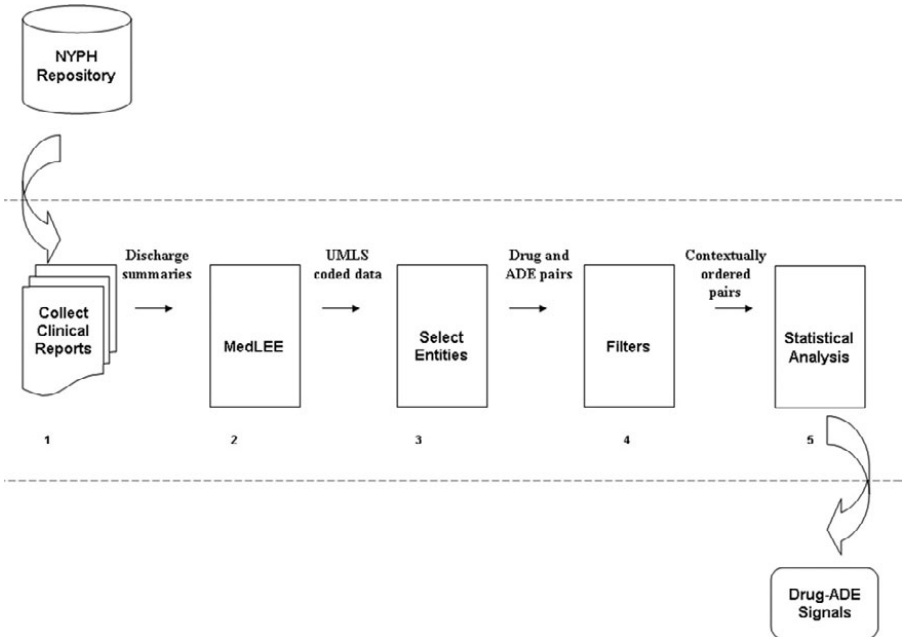

Figure 10. Overview of System Framework (Wang et al., 2009).

iv. Phase 4: Data Filtering

In order to eliminate some clinical entities, two main and one contextual filters are used. Filter No1 eliminates entities associated with modifiers related to some certainty values (e.g., low certainty), past events, or family history events. Filter No2 eliminates chronologically wrong drug-indication timelines. The contextual filter confirms that the order of events is chronologically correct.

v. Phase 5: Drug-AE Association

In order to reveal drug-AE associations, potential drug-AE pairs were developed by employing statistics (Wang et al., 2009).

This information system has the power to automatically detect new adverse events by using only the existing information in EHR systems; only few arrangements are needed (NLP and statistics). In the future and as the creators of this system suggest, it might enable the combination of

structured and unstructured data. However, some precautions should be considered when this system is used. For example, treatment indications should not be confused with adverse events; sometimes a drug is used in the later stages of an illness and the symptoms that might be observed during its administration might only be the result of the illness's progression. Moreover, other limitations are presented in Wang et al.'s paper (e.g., narrative reports were collected only from inpatients or not appropriately defined UMLS codes) and therefore, more research on this system should be done (Wang et al., 2009).

- **National Clinical Trials Registry**

McGray and Ide recognized a "gap" in the literature regarding information systems for clinical trials; the existing systems centralize the need for access to clinicians and researchers, while access to patients and other public groups is neglected (2000). Therefore, designing their system began with guarantying that the patients' needs were identified. Their next steps were to implement a standard syntax and semantics for the data of interest and to realize that the system should be built in phases, as more requirements would surface during the system's development. As the World Health Organization requires clinical studies reports to be publically available (WHO, 2015), one of their primary goals was to build a system understandable to the public. Therefore, they developed an accessible and easily operated Web-based system for patients to express their queries. The first trials to be introduced to the system were trials sponsored by the National Institute of Health (NIH) and McGray and Ide agreed with the trials providers to a common set of data elements (required and optional) which will be used for the clinical trials data (McGray and Ide, 2000). The required and optional data elements are shown in Table 2.

Table 2. Required and Optional Data Elements (McGray and Ide, 2000).

Required Data Elements	Optional Data Elements
Study identification number	NIH grant or contract number
Study sponsor	Investigator
Brief title	Official title
Brief summary	Detailed description
Location of trial	Study start date
Recruitment status	Study completion date
Contact information	References for background citations
Eligibility criteria	References for completed studies
Study type	Results
Study design	Keywords
Study phase	Supplementary information
Condition	URL for trial information
Intervention	
Data provider	
Date last modified	

According to this system, the recruitment status can be: not yet recruiting, recruiting, no longer recruiting, completed, suspended and terminated. Moreover, the study types (9) recognized by this system are: diagnostic, genetic, monitoring, natural history, prevention, screening, supportive care, training and treatment. Whenever is feasible, the intervention and condition studied in each trial is provided via the Medical Subject Headings (MeSH) of the UMLS (Unified Medical Language System); an exception to this rule can be the case of a new drug which is not included in the MeSH yet. Furthermore, the references of publications should be delivered by the study providers with a unique identifier (UI) which will be linked by the system's creators to a MEDLINE citation record (McGray and Ide, 2000).

Every record introduced to the system is characterized by a specific and unique number, which operates as the identifier of the trial. The data from each provider is included in the system's central database at the National Library of Medicine (NLM). McGray and Ide(2000) offered help to institutes that wanted to redesign their databases in order to provide more standardized data, and even to those that did not need assistance, the system's creators provided their services before the final confirmation of the data. Finally, each report ought to be sent to the system in extensible markup language (XML) format correspondingly to the created document type definition (DTD) (McGray and Ide, 2000). The DTD created by McGray and Ide (2000) contains data regarding the study title, study sponsor, a brief summary of the study, the start and end

dates, the intervention type, etc. This choice was made because of the XML's advantage to be understandable by both humans and computers (McGray and Ide, 2000).

The final outcome of this design is the collection of NIH-trials records and their storage in a centered database at NLM. The dataflow during the implementation of this system is shown in Figure 11. The provider regularly sends XML formatted reports to NLM via file transfer protocol (FTP) and the system proceeds with data validations whenever an update is sent from a provider. Afterwards, the validated data is expanded with condition names (from a list of disease categories), literature references, links to MEDLINEplus and NLM's consumer health site. After this process, the data is made available on the Web and therefore, to the patients. Finally, the retrieval engine is responsible for processing users' queries; the engine inspects for spelling errors, expands the query and develops HTML for Web browsers (McGray and Ide, 2000).

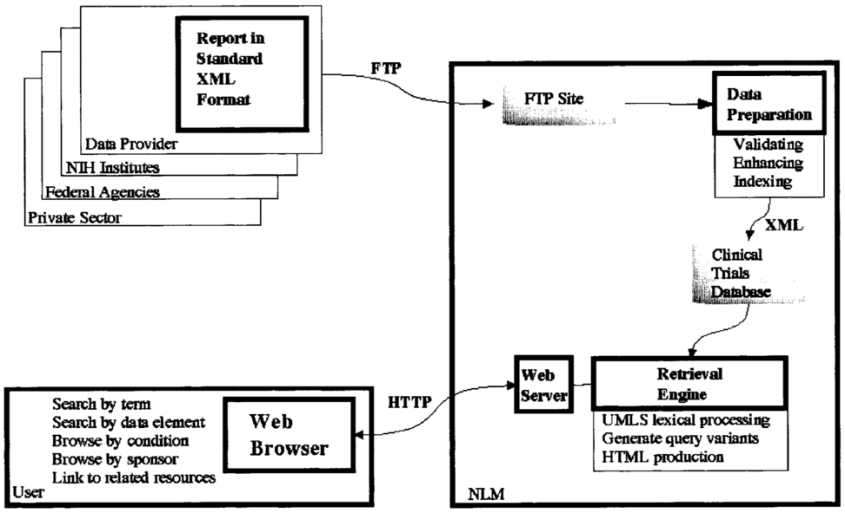

Figure 11. System Design (McGray and Ide, 2000).

During data preparation, the data flow is as shown in Figure 12. The data is received as XML documents ("received area"), they are checked for structural errors ("validated area"), they are enriched in order to enable a "browse-by-condition" process ("enhanced area") and finally, the clinical studies database is built with the collection of clinical trials' data and a vocabulary

collection ("publisher process"). After the end of the data preparation, the database is ready and accessible by the retrieval engine (McGray and Ide, 2000).

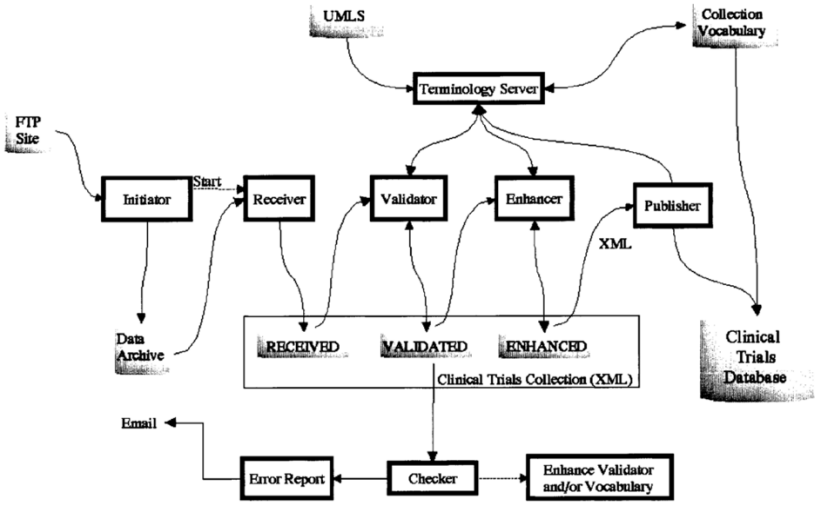

Figure 12. Data preparation subsystem (McGray and Ide, 2000).

- **OpenTrials**

Goldacre and Gray (2016) discussed the OpenTrials project and its Phase I development. Via this project, information regarding clinical trials registered in different databases/registers and diverse structures will be able to be combined, compared and accessible to researchers, academic and healthcare personnel, patients and clinical research related organizations. This information system aims at the creation of an open database and the sharing and analysis of information regarding clinical research. As shown in Figure 13, clinical trials data are extracted from:

- *Registers (industry registers: for trials carried out by a company; national registers: for trials carried out by one regulator or in a specific location)*: different registers might contain different information for the same clinical trial and these differences should be noticed and clarified.

- *Academic journals:* they offer information in several document types (e.g., protocols) and in semi-structured free text form. Also, an ID number can pair an academic journal article with a registry entry.
- *Regulatory Documents:* they present information in a defined structure or in a free text form.
- *Structured Data:* they can be retrieved via registers and manually extracted data from free text reports of systematic reviews (e.g., from SRDR – Systematic Review Data Repository).
- *Trial Paperwork:* information can become available via several forms (blank consent form, blank case report form, patient information sheet, etc.), protocols, lay summaries and statistical analysis strategies (Goldacre and Gray, 2016).

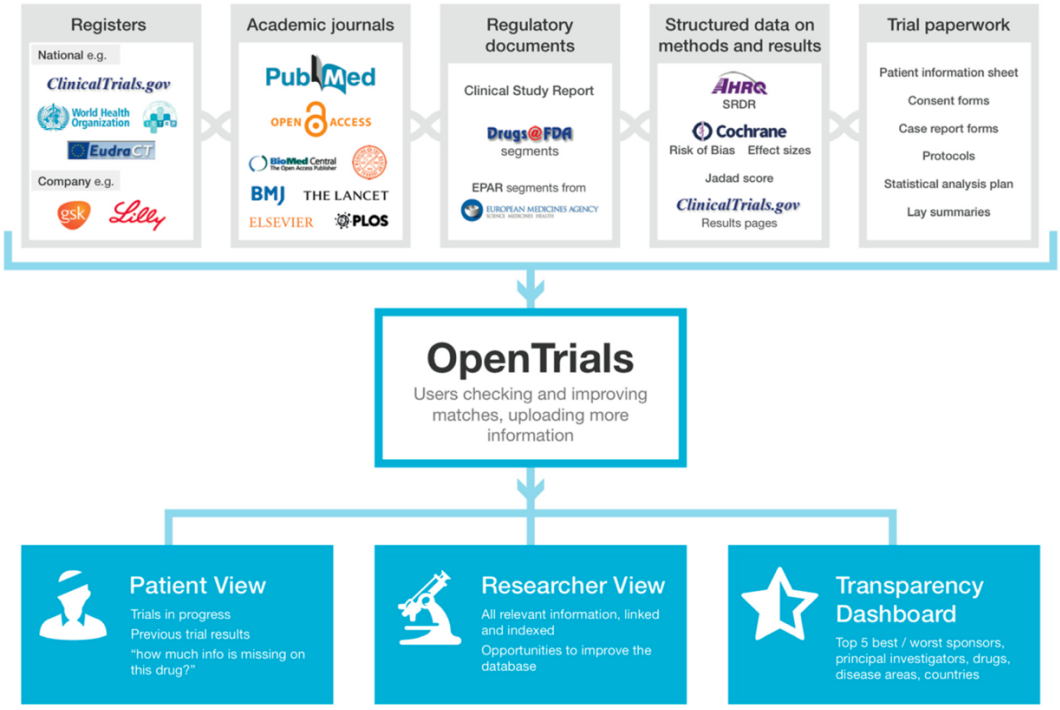

Figure 13. Overview of OpenTrials data schema and information flow (Goldacre and Gray, 2016).

The database needed for this information system can be created by several techniques, such as "importing publicly accessible structured data, web-scraping (for accessible, but not available for download data), record linkage (a matching data technique) and curated, targeted

crowdsourcing and donations of structured data". Unfortunately, the database development (inclusion of data from all clinical trials conducted or being conducted) is a very expensive procedure and the project's funding is limited and therefore, this development will proceed step by step and with the above-mentioned techniques (Goldacre and Gray, 2016).

Despite the advantages that this system provides, problems regarding its application can be noticed, as well. For example, in order for the combination of different data from different sources to be introduced into one database, a common "dictionary" ought to be developed and applied. Goldacre and Gray (2016) decided to solve this problem by introducing wide data-categories based on the list of sources mentioned above (registers, academic journals, regulatory documents, structured data and trial paperwork). By following this approach, different ways of addressing structured data are included and therefore, data with multiple formats can be managed and interpreted. Specifically, what Goldacre and Gray (2016) created is a "thread of documents on a given trial"; a document can be an actual document, such as an informed consent, or structured data, e.g., data from ClinicalTrials.gov, such as reported results of clinical trials, reported adverse events and all the information provided in a clinical study's protocol, as presented in the first chapter (Goldacre and Gray, 2016).

Ontologies and Ontology-based Information Systems in Clinical Research

Clinical research is an information-intensive field and in order for its outcome to be valid, accurate and complete, the information needed should be collected via several sources. This "key requirement" for successful data collection, integration and harmonization is difficult to be met. Hence, the development of ontologies supporting the clinical research, as well as, ontology-based clinical research information systems seems to become more and more necessary. The ultimate goal is to create a connection between the many ontological resources found in health care sites (Richesson and Andrews, 2019).

The use of ontologies in data representation during a clinical study can offer many advantages, such as successful integration of data, increase in the published literature and database and reusability of research data (Smith and Scheuermann, 2011). Ontologies can benefit every step of a clinical study; they can describe a study's design, represent study metadata deriving from several sources (e.g., Electronic Health Records or Clinical Report Forms), enable selected tasks during a study's execution, capture eligibility criteria and identify new patients as eligible participants, help investigators to detect a study's strengths and weaknesses and to interpret them along with the study's results (Sim. Et al., 2014). Therefore, ontologies and ontology-based information systems used in clinical research are analyzed in this chapter.

- **Epoch:** an ontological framework to support clinical trials management

The goal of the creators of Epoch was to build an ontological framework which will be able to follow-up participants throughout the duration of a clinical studiy and clinical specimens tested at trial laboratories. During a clinical study, numerous tests and examinations are conducted. Therefore, following-up participants in order to assure that these tests are complete and to select and report their results is an important part of a clinical study. Epoch (Fig. 14) is developed in

order to be used for the management of multi-site clinical trials carried out at the Immune Tolerance Network (ITN), as it will "collect, manage, integrate and analyze clinical trial and immunoassay data" (Shankar et al., 2006). Although many activities are considered a part of clinical trials management, Shankar et al. (2006) focused only on two of them for building Epoch: (a) participant tracking and (b) specimen tracking.

Their knowledge-based framework was developed based on three types of methods:

A. Development of a number of ontologies for a more accurate specification of knowledge on immune disorders.
B. Ontology-database mapping methods which will merge the study metadata and biomedical knowledge base with stored primary data in the data repository.
C. Concept-driven querying methods via which the data repository will accommodate integrated data management and high-level data analysis projects (Shankar et al., 2006).

Epoch consists of more than one ontology:

i. *The Clinical Ontology,* i.e., a terminology of the clinical and biomedical discipline regarding immune tolerance disorders.
ii. *The Protocol Ontology,* i.e., a model focused on the protocol concepts of participant and specimen tracking; hence, the protocol schema (information for the phases of a trial and their chronological sequence) and the schedule of events (the timing and events of a study visit) are represented in this ontology.
iii. *The Specimen Ontology,* i.e., the modelling of specimens' workflow (collected in a site – transferred to bio-repositories – transferred to core laboratories where results are analyzed– stored and analyzed in a data warehouse).
iv. *The Site Model,* i.e., a structure for storing site-related data (e.g., protocols, participants, study coordinators, etc.).

v. *The Specimen Container Ontology,* i.e., a list with information regarding all specimen containers and each one's characteristics (e.g., size, manufacturer, additives, material, shipping instructions, etc.).

vi. *The Virtual Trial Data Ontology,* i.e., a summary of the collected data during the study, such as specimen workflow records and participants' clinical records (Shankar et al., 2006).

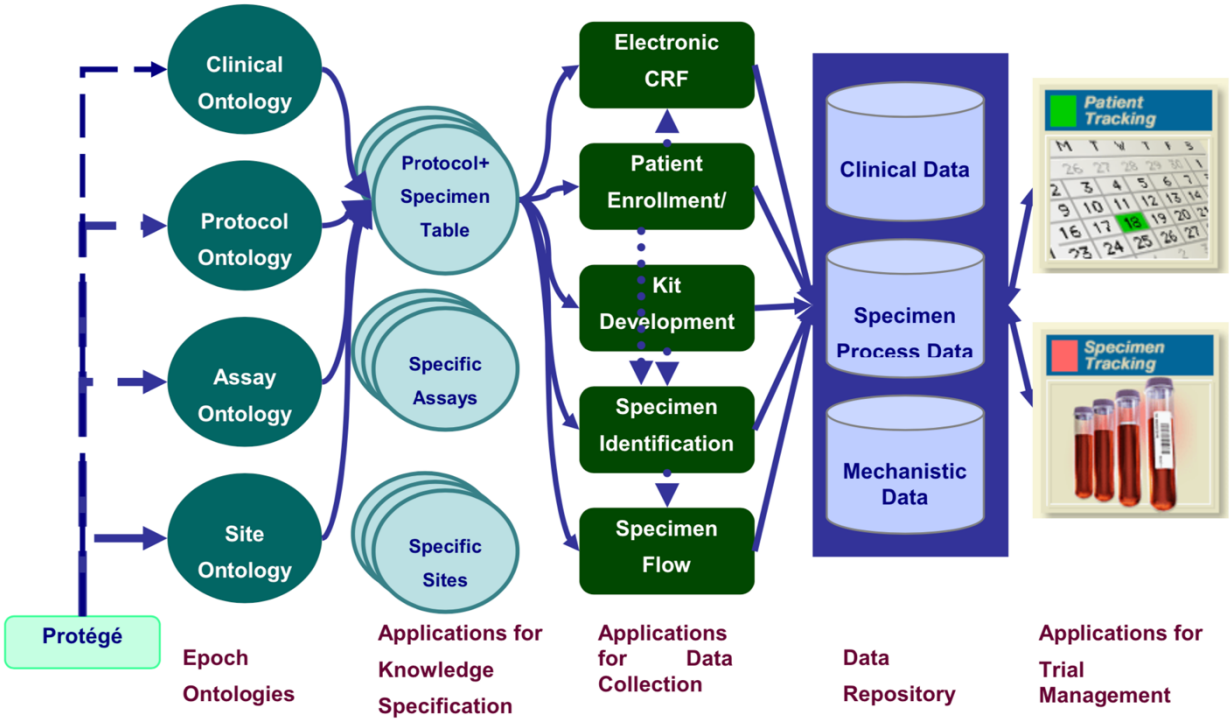

Figure 14. Epoch: An Ontological Framework to support Clinical Trials Management (Shankar et al., 2006).

Other activities needed for the development of Epoch, besides the protocol schema and the schedule of events, are the specimen table and specimen flow. The specimen table provides information regarding the specimens collected for each participant, as well as their analyses. The specimen flow's function is to provide information regarding the processing of the specimens, as mentioned above (Shankar et al., 2006).

For the Epoch ontological framework to provide knowledge-based reasoning and query methods, its development is based on several interacting elements (Fig. 15). The Epoch Knowledge Base provides the ontologies described previously, the Knowledge Base Server provides an application program interface (API) for permitting access to other sections to the ontology repository, the Clinical Trial Database stores data created from the implementation and execution of a clinical trial (included data: participant enrollment data, visits and activities, specimen shipping and receiving data and clinical results) , the Model-Database Mapper, based on a mapping ontology, enables access to the Clinical Trial Database for relational data, the Inference/Rule Engine carries out constraints in Epoch ontologies, the Utility Functions enables import and export of the knowledge base and finally, the Clinical Trials Management Applications (such as "authoring tools, operational plan builders, study site management tools, participant and specimen tracking applications, and trial data analyzers") work with the Epoch components in order to support the clinical trials management (Shankar et al., 2006).

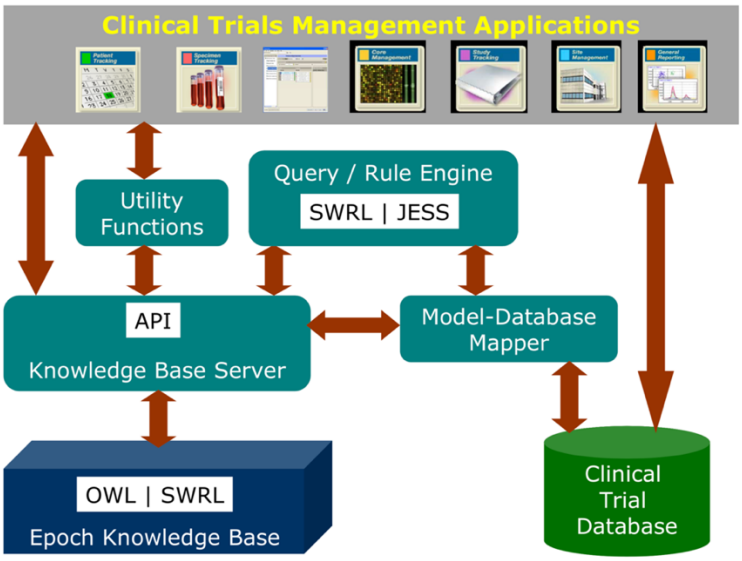

Figure 15. A high-level functional architecture of the Epoch framework (Shankar et al., 2006).

- **OnWARD:** Ontology-driven web-based framework

After setting as a goal to create a secure, high-quality and flexible framework, Tran et al. (2011) proposed OnWARD for a Phase II clinical study called HeartBeat (Heart Biomarker Evaluation in

Apnea Treatment). This Ontology-driven web-based framework is a combination of three components: Data Specification (DS), Data Capture (DC) and Data Exploration (DE); these components are the three phases of a clinical study's data life-cycle and are integrated under a clinical research ontology. Once more clinical studies will use OnWARD, this clinical research ontology will be expanded (Tran et al., 2011). The functionality of each component is described below:

- *Data Specification (OnWARD-DS):* this element provides to the investigator the opportunity to choose the way for storing and collecting data. In order for data specification to be complete, a backend relational database is selected, and each clinical study connects with this database and proceeds with data entry, retrieval, validation, etc. with the help of the other two components.
- *Data Capture (OnWARD-DC):* OnWARD-DC is a "dynamic form generation engine" which will provide to the system a variety of different, flexible and changeable forms. These forms will contain data in several structures (e.g., free-text, numeric responses, etc.), which will be based on the metadata chosen from OnWARD-DS and on the clinical research ontology.
- *Data Exploration (OnWARD-DE):* OnWARD-DE participates in the input validation process, during which data entry is checked. This process also uses the clinical research ontology used by the system and validates data in three pillars: (i) Data type, (ii) Hard range validation and (iii) Soft range validation. Moreover, another function OnWARD-DE provides is *"skip patterns":* data entry is adjusted according to the way that previously asked questions were answered. Finally, OnWARD-DE benefits the clinical study process by enabling researchers to develop reports regarding a clinical study and to search for reports of a study of interest; hence, patient recruitment becomes more efficient (Tran et al., 2011).

The results from the evaluation of this system/framework showed that not only the advantages in finding eligible participants are obvious, technical advantages can be observed, as well. First,

training-time was low, and usability was high. Second, the system will be able to accommodate clinical studies of different sizes (from small to medium). Third, OnWARD can support randomization. Last, this Ontology-driven web-based framework can be characterized by data quality and security (Tran et al., 2011).

- **A semantic interoperability layer**

Alonso-Calvo et al. realized the need for minimizing the heterogeneity of data originating from distributed centers in multi-sited clinical studies. As a solution to the existing, manual and error-prone approach, they proposed a method for automatically integrating heterogenous data by introducing a Common Information Model (CIM), which will be the outcome of the integration and homogenization of information from Clinical Trial Management Systems (CTMS), Laboratory Information Management Systems (LIMS), Electronic Health Record (EHR), etc. The CIM will be a combination of a *Common Data Model (CDM)* and a *Core Dataset (CD)* (Fig. 16) (2015).

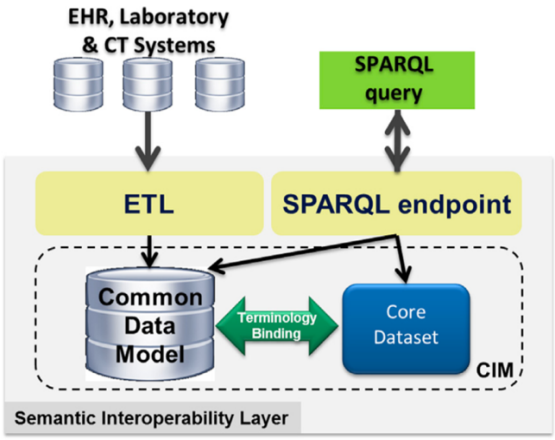

Figure 16. Semantic Interoperability Layer (Alonso-Calvo et al., 2015).

- *Common Data Model (CDM):* The CDM will be responsible for homogenizing information system data models from several institutions or from an institution's different departments
- *Core Dataset:* The CD will be able to provide a vast amount of clinical terminologies found in several data sources and scenarios.
- *Terminology Builder:* Another component of this system is the *Terminology Binding* which will determine relations between CDM and CD. For developing the Terminology Binding,

three steps should be followed: vocabulary integration, binding annotation and disambiguation. Finally, for retrieving data from this system, SPARQL queries are executed (Alonso-Calvo et al., 2015).

This semantic interoperability layer will be used for recruitment and screening of patients based on eligibility criteria and for retrospective analysis of multi-centered clinical trials. This system will eventually be improved by implementing authorized access and therefore, becoming more secure and legal (Alonso-Calvo et al., 2015).

- **The Linked2Safety project:** Using distributed Electronic Health Records (EHR) with different legal and ethical requirements.

According to Antoniades et al. (2012) the Linked2Safety project was initiated because of the growing need of using distributed Electronic Health Records (EHR) with different legal and ethical requirements for several healthcare reasons, e.g., clinical trials recruitment, identification of adverse events, increase in the statistical power of data analyses, etc. Although a combination of EHRs from different organizations is a helpful solution whenever merged data are needed, limitations in this procedure exist due to the legal and ethical implications of patient privacy. In order to overcome these limitations, Linked2Safety was suggested (Fig. 17) (Antoniades et al., 2012).

As the main goal of this system is the preservation of legal and ethical requirements, a "closed-world" room is introduced, in which EHRs are processed in a network-connection-free environment. In this room, only authorized personnel have access. During the aggregation of data, Linked2Safety will provide options for the appropriate ways of analyzing and combining data; because of this function, problems regarding the ethics and laws of aggregating data are minimized. The program proceeds with a quality control, aiming to identify possible ways of connecting the data to a subject or a group of subjects. By doing so, the system provides means to prevent such connections. Finally, only aggregated data will be eligible to be transferred

outside the “closed-world” room and become a part of the rest of the Linked2Safety platform (Antoniades et al., 2012).

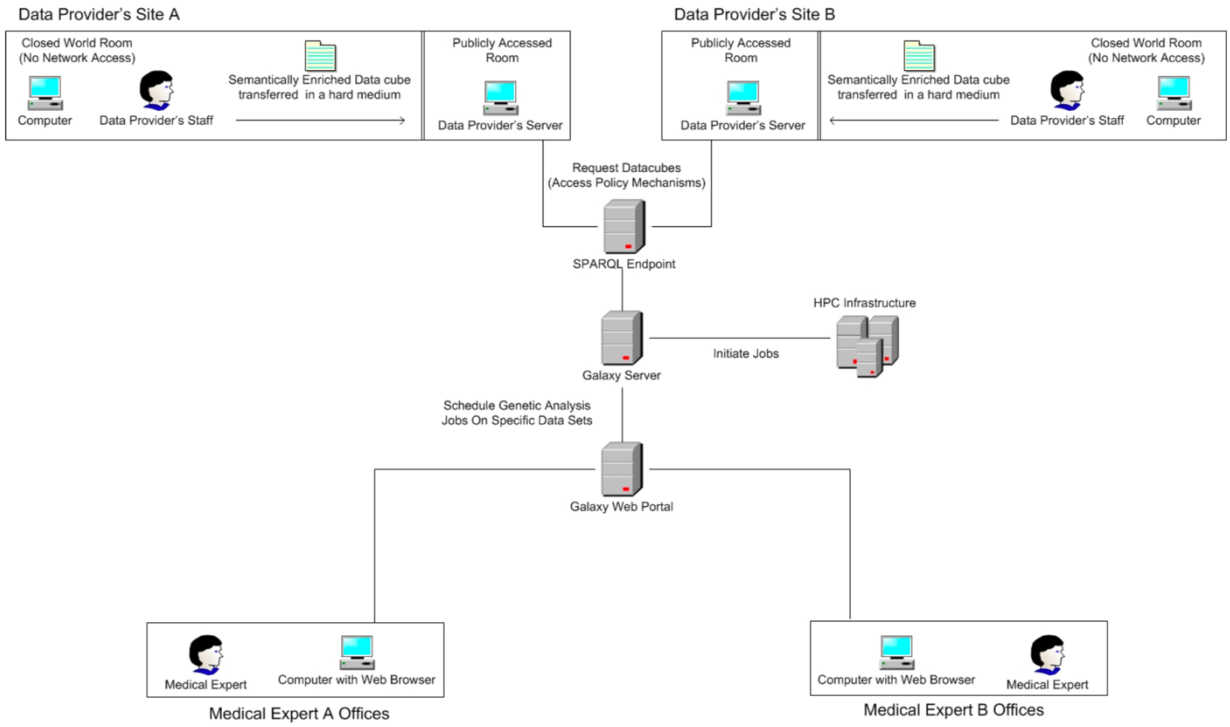

Figure 17. The Linked2Safety Platform (Antoniades et al., 2012).

This procedure will be followed by each site maintaining clinical data. If some definitions overlap, then they are recorded and combined by the Linked2Safety platform and as a result, the statistical power of the outcome increases. By the end of this process, the data will be available to researchers and other healthcare professionals with access to the Linked2Safety platform (Antoniades et al., 2012).

This project will eventually decrease the expenses of clinical trials, because sites and subjects will be readily identified as appropriate for a clinical trial, as described below. Another advantage that the Linked2Safety platform will offer is the ability of analyzing aggregated data with methodologies which were not developed or considered when the data were collected in the first place; therefore, this ability will make the use of old data for new clinical studies easier and

more possible. In order for this ability to be created, multidimensional data cubes will be used. The Linked2Safety project seems to be able to offer its help to many procedures during a clinical study. However, three functions were introduced by Antoniades et al. (2012):

i. Recruitment/Selection of Subjects for Phase III Clinical Trials

This step is a costly and time-consuming procedure. Moreover, it should be mentioned that Linked2Safety will help the recruitment in multi-centered clinical trials and in the scenario that investigators and academic research organizations search for physicians who carry out specific clinical trials with patients meeting specific eligibility criteria.

ii. Post Marketing Surveillance Trials (Phase IV)

During Phase IV trials, Linked2Safety will offer information over adverse events presented in several and distributed sites along with data, such as demographics and data on medication used and treatment performed. Therefore, factors causing these adverse events and challenging safety will be noticed and certified as soon as possible.

iii. Identification of Relations between Molecular Fragments and Specific Adverse Side Effects Categories (Chemoinformatics)

Linked2Safety is able to identify possible relationships between chemical structures of drugs and adverse events. For this process to work, a vast amount of information regarding patients and their drug treatments ought to be available and readily accessible (Antoniades et al., 2012).

- **ODaCCI: Ontology-guided Data Curation for Multisite Clinical Research Data Integration**

Cui et al. introduced ODaCCI as an ontology-guided approach for supporting the Informatics and Data Analytics Core (IDAC) curation strategy of CSR (Center for SUDEP research; SUDEP: Sudden Unexpected Death in Epilepsy). This approach was conducted due to the CSR's challenges, such as data heterogeneity (an Ontology-driven Patient Information Capture system – OPIC – will decrease the heterogeneity), data access restriction, multimodal data linkage and data quality. These challenges are the result of the need for integrating data from the distributed sites of multi-centered clinical studies (2016).

The ODaCCI architecture is shown in Figure 18 and the steps for developing this system are the following. First, identifiable information (patient phenotype data) from each site is inserted into the OPIC system. Second, via an automatic (by OPIC) and manual (by personnel) de-identification process, the information in the OPIC system becomes de-identified. Then, the de-identified data will be available in the CSR central repository. Third, the de-identified data from each site are transferred to a database in the central repository which is connected with the data curation system (the ODaCCI) and therefore, an expert will be able to audit and curate the data; the curated data will be stored in a separate database which will store only curated data (Cui et al., 2016).

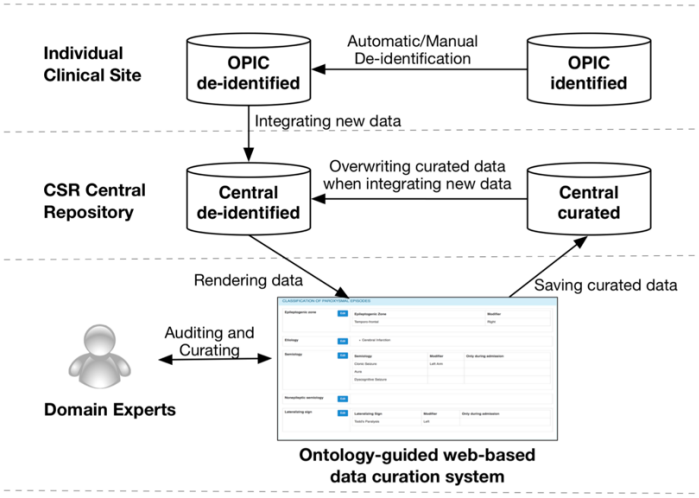

Figure 18. The Ontology-guided Web-based Data Curation System (ODaCCI) architecture (Cui et al., 2016).

The ODaCCI consists of six sections:

- *Common data elements (CDEs):* CDEs are common data found in every clinical site and the data sources were chosen by experts in the epilepsy domain. This data will be integrated.
- *Ontology-based vocabulary:* The ontology used is the Epilepsy and Seizure Ontology (EpSO).
- *CDE to data source mappings:* The CDEs from the different clinical sites might not be the same, because different versions of OPIC might be used in these sites. Hence, the data ought to be mapped to the data dictionary's data tables and columns.

- *Dynamic generation of data curation widget:* This widget is a web-based data curation interface used during the manual curation of data.
- *Dynamic Generation of MySQL statements:* They are used for saving the changes that domain experts make.
- *Data auditing measures:* These measures are actually data quality measures and are *completeness* and *consistency* (Cui et al., 2016).

ODaCCI's aim is to provide integrated, high-quality and secure data from distributed clinical trial sites. As a result, recruitment SUDEP research was increased. Lastly, it should be mentioned that ODaCCI is a scalable system, due to its general design, but its usability has not yet been tested (Cui et al., 2016).

- **Recruit:** Ontology-based information retrieval system

Recruitment in clinical trials faces many challenges as the recruiting target is rarely achieved and even when it is achieved, it is highly time-consuming. After recognizing this problem, Patrão et al. introduced an information retrieval system which will be able to screen patients, i.e., find them based on clinical criteria. This system is called Recruit and is based on the Electronic Health Record (EHR) data (2015).

Recruit consists of two parts: the *backend* part, which extracts, transforms and indexes data and the *frontend* part, i.e., a web interface for the user to use the system. The backend combines structured and unstructured data from the original databases and provides an index. Afterwards, the frontend queries the search engine (Apache Solr) which processes the index (Patrão et al., 2015). More specifically, Patrão et al. (2015) developed an ontology-based data warehouse and proceeded with importing this data in an indexing server; this data is considered structured metadata. And they were imported along with unstructured report texts. First, data from several structured databases were integrated into a triple store endpoint (Openlink Virtuoso) (Figure 19). Afterwards, the triple store endpoint was enriched by a set of SPARQL queries (Patrão et al., 2015).

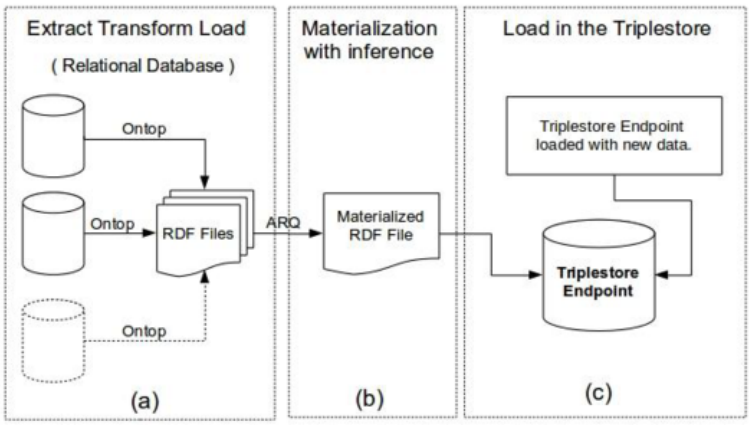

Figure 19. Integration process dataflow (Patrão et al., 2015).

Second, structured (presented as an ontology) and unstructured data were integrated and published into a search engine (Figure 20) (Patrão et al., 2015).

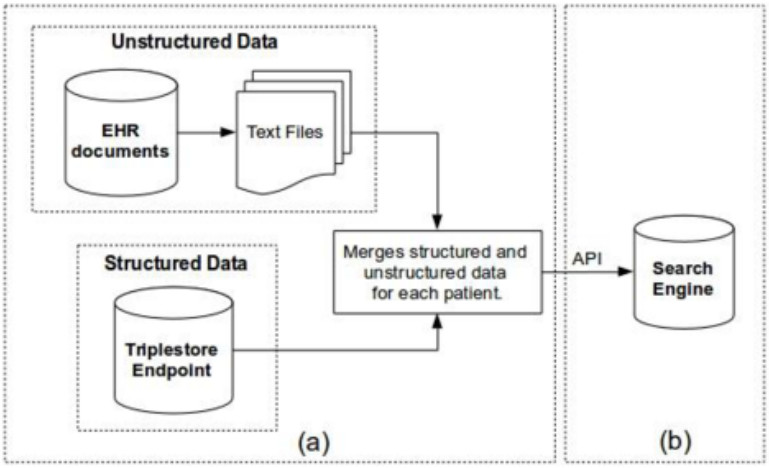

Figure 20. Search engine publishing dataflow (Patrão et al., 2015).

An important advantage of this system is that the system will evaluate the way that users use it and will be strict with the patient data access (Patrão et al., 2015).

- **SemEHR:** “an open source semantic search and analytics tool for EHRs”

Another information system aiming to support recruitment was introduced by Wu et al. (2018). According to them, SemEHR is “a semantic search and analytical system that generates a complete and process- able view of patients from their clinical notes”.

As shown in Figure 21, three subsystems are combined in order to develop SemEHR:

- *The production subsystem,* which extracts unstructured (free-text) clinical notes from heterogeneous EHRs. For this function to be performed, data retrieval, information extraction and semantic indexing ought to be done. For the data retrieval step to be concluded, a harmonization tool for EHRs is used (called CogStack). This tool homogenizes the distributed data which are in heterogenous formats. Each document flows from the data retrieval step into an NLP pipeline and the extracted documents are afterwards analyzed during the semantic indexing step. The analysis results are, then, indexed by an Elasticsearch cluster. The final result is the development of patient-level summaries which are constantly updated whenever a new document is added to the index.
- *The continuous learning subsystem,* which continually addresses study-specific matters. Feedback from users is collected and after it is analyzed, two elements are used in order to optimize the Information Extraction (IE) results: a *rule engine* (creates and applies rules for removing results that are not wanted) and a *machine learning engine* (calculates the value of each concept mention, according to user feedback).
- *The consuming subsystem,* which consists of components that will use IE results and clinical knowledge in order to support tasks, such as trial recruitment. Each consuming task is called “study” and is stored in the SemEHR’s Study Knowledge Graph (KG) along with all the “study’s” parameters (Wu et al., 2018).

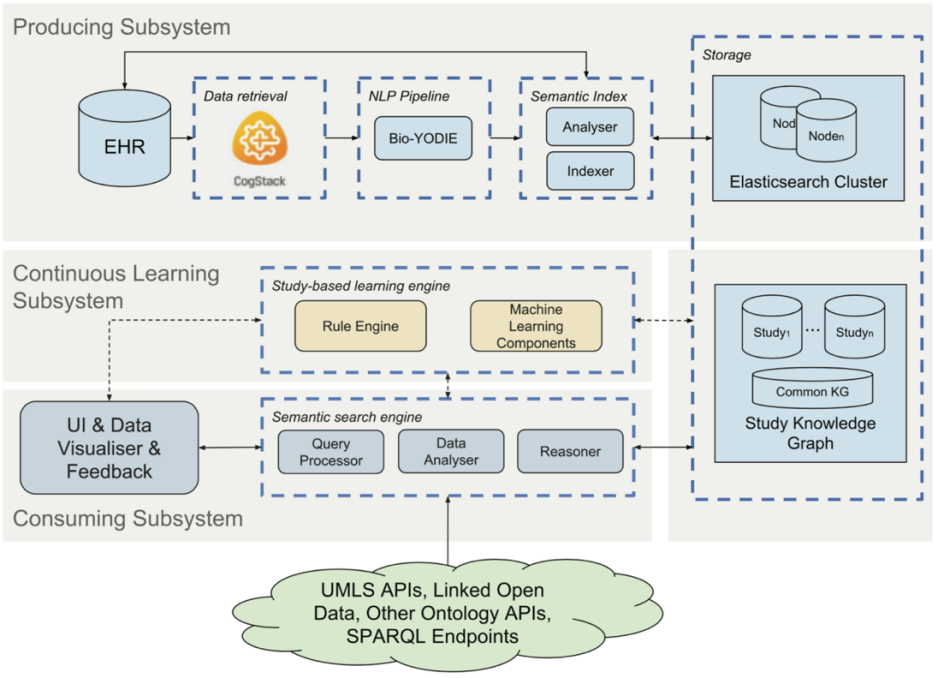

Figure 21. The SemEHR architecture (Wu et al., 2018).

- **Ontology-Based eXtensible (OBX) data model**: a framework for clinical research data in the Immunology Database and Analysis Portal (ImmPort).

Kong et al. (2011) realized the need for the development of a mechanism which will help with the re-use and re-analysis of clinical research data. Therefore, based on the Basic Formal Ontology (BFO) (BFO, 2019) and the Ontology for Biomedical Investigation (OBI) (The OBO Foundry, 2019), Kong et al. built the Ontology-Based eXtensible data model (OBX); this ontology will support the Immunology Database and Analysis Portal (ImmPort) (2011).

The main component of the data model is the *Event Table* which provides information regarding events (planned or not) that actually happened (actual events). As shown in Figure 22, each event can be related with its study design, the time context of its occurrence and one or two objects. Furthermore, a planned event can be described by the *Procedure Specification* subclass and more information regarding its protocol will be presented (Kong et al., 2011).

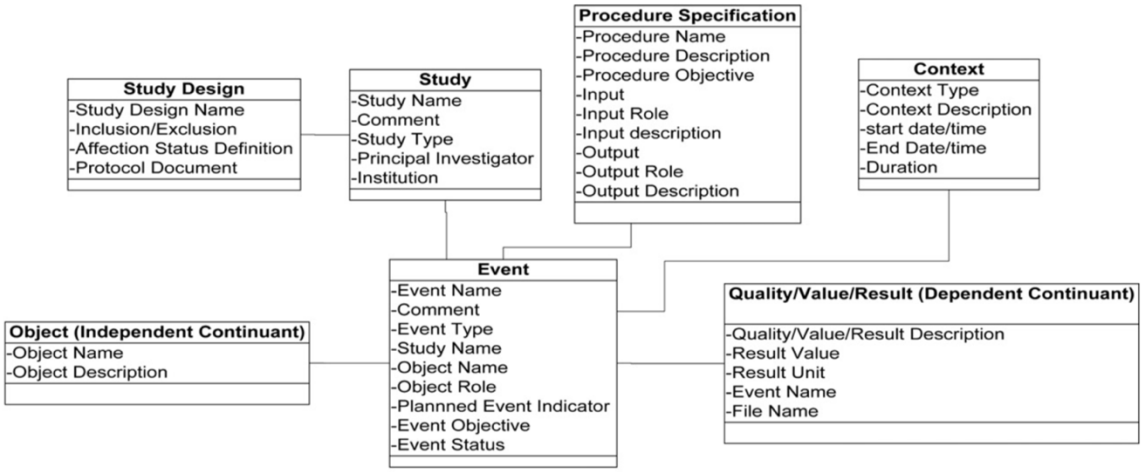

Figure 22. OBX conceptual model representation (Kong et al., 2011).

Afterwards, and based on the OBX core conceptual model, Kong et al. developed a framework which presents some of the specific members of the entities in Figure 22. The entities presented in the framework are of great importance to the ImmPort clinical research database. The class types and subtypes of the entities are shown in Table 3. Each subtype contains certain attributes (e.g., name, type, time, duration etc.) (2011).

Table 3. Class types and subtypes in OBX model (Kong et al., 2011).

Class Types	Subtypes	Inputs	Outputs
Object	Population, Human subject, Biological sample, Animal subject, Compound or Agent, Complex compound, Environment factor, Site, Software, Instrument, Manufacturer.		
Biomaterial Transformation	Subject biomaterial transformation, Substance biomaterial transformation, Substance merging biomaterial transformation, Biosampling process, Surgery process, Device intervention process, Environmental exposure process.	Events with one or more biomaterials	Events with one or more biomaterials
Assay	Subject assessment (Medical history, Family history, Questionnaire), Subject inclusion-exclusion, ECG, Adverse Event, Primary result, Biological sample, Human subject, Lab test or Measurement of analyte.	Events with one or more biomaterials	Data
Data Transformation	Diagnostic process, Study reported premature termination, protocol deviation, Research data analysis (Baseline and Outcome measurement process), Derived result (Diagnosis result).	Data	Data

The outcome of this ontology-based model is the creation of a database schema which will enable the storage and collection of diverse clinical research data originated from studies in several domains and sites. Moreover, OBX provides to researchers the opportunity to integrate clinical data (CRF data) and mechanistic experiment data (type of Assay). Finally, this model offers the

advantage of data sharing and system interoperability from ImmPort database to other organizations using it (Kong et al., 2011).

- **The ObTiMA system:** Ontology-based managing of clinical trials (Design clinical trials and manage the patient data within them).

In 2010, Stenzhorn et al. developed an Ontology-based Trial Management Application (ObTiMA). The data management of this system is based on a master ontology and a semantic mediator, while the system's main components are the Trial Builder and the Patient Data Management System (Fig. 23). Stenzhorn et al. (2010) describe these four components as follows:

- *Master Ontology*

In this ontology concepts regarding research on cancer are presented. According to Stenzhorn et al., although the ontology is focused on the cancer domain, not many limitations can be found over its reusability in other domains; cancer is a field with a vast amount of entities and processes (2010). Therefore, and as the master ontology contains information regarding genetics, administration and legal requirements, it might be able to be used in other research domains, as well (Stenzhorn et al., 2010).

- *Semantic Mediator*

During semantic mediation, information from different data sources is matched, combined and retrieved. This process is characterized as "semantic" because the databases are developed with Master Ontology concepts and relations. Moreover, an ObTiMA's database can be used as a data source to the mediator (Stenzhorn et al., 2010).

- *Trial Builder*

A clinical trial consists of several aspects (as presented in Chapter 1) and the trial builder enables the ObTiMA's user to specify them. For example, via a master protocol the trial's outline can be

described and treatment plans and their time-order can be graphically presented. For collecting the information regarding a treatment, a Case Report Form (CRF) ought to be created at every stage of the treatment plan. The CRFs are developed based on the Master Ontology and present integrated information in the shape of a graphical user interface; the result is the creation of a trial database. The first step is called *Ontology View* and the user creates a CRF item by selecting concepts from the ontology – the master ontology aims to be less detailed than the clinicians' meticulous points of view. Each concept selection will lead to the presentation of its relative concepts. The next step (*Item Editor)* begins after the concept is chosen and presents the concept's attributes, e.g., answer possibilities or data type. The final step is called *Preview Items* and demonstrates the items created in the preceding steps in the exact order that they should be presented on the CRF. If needed, this order can be changed manually (Stenzhorn et al., 2010).

Moreover, another outcome is the development of a CRF Repository. As the CRFs are developed based on the Master ontology and share the same terminologies, existing CRFs can be stored and reused in other clinical studies (Stenzhorn et al., 2010).

- *Patient Data Management System (PDMS)*

This system is automatically developed (and not manually) based on the Trial Builder and more specifically on the master protocol and CRFs created during the Trial Builder. Via PDMS, clinicians can find the help they need in order to follow-up each patient's treatment plans and fill the CRF with information over the treatment status. An advantage of this system is the feedback it provides regarding every new entry of information and therefore, errors are displayed and explained. Furthermore, ObTiMA participates in the data exchange between distributed clinical settings; hence, trial metadata, CRF descriptions and patient data can be imported and exported from the system. Finally, as the system was developed with the goal to integrate data from different clinical trial sites, security regarding legal and ethical requirements must be taken into consideration. In order to guarantee data security, ObTiMA uses two different database servers: one for personal data and another one for data collected during research (Stenzhorn et al., 2010).

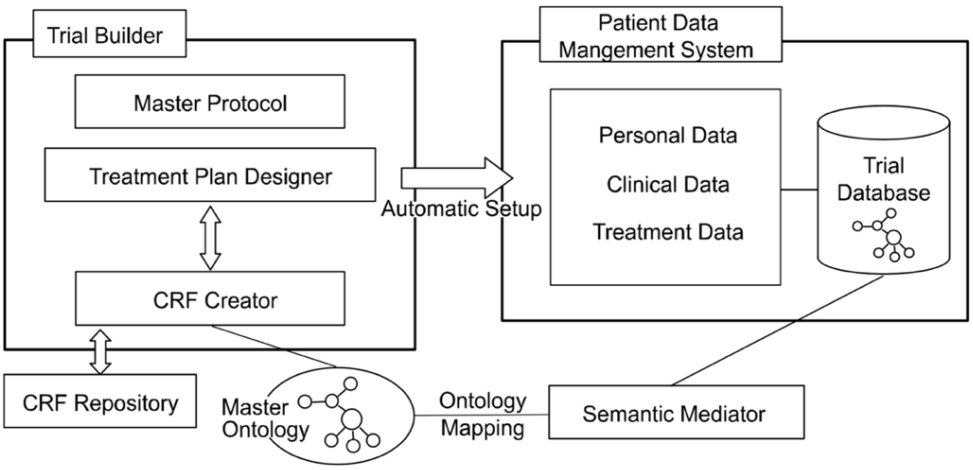

Figure 23. The ObTiMA System Components (Stenzhorn et al., 2010).

- **RCT Schema**: a trial ontology for trial interpretation and application to clinical care.

Randomized Clinical Trial (RCT) knowledge bases are "trial banks" which have systematic reviewing as their target task. For managing systematic reviewing, the information needed is the sum of recommendations regarding a trial report's components: information about trial design, trial execution and trial results. The RCT ontology was developed and evaluated with the competency decomposition method developed by Sim, Olasov and Carini (2004), i.e., each task is hierarchically divided into subtasks and information regarding the completion of each task is delineated in order to determine the ontology components (Sim, Olasov and Carini, 2004).

Table 4. Examples of trial information modeled in RCT Schema (Sim, Olasov and Carini, 2004).

Concept group	Examples
Administration	Trial title, stage, dates, and investigators, study sites, funding, ethics, description of trial committees, and errata
Background	Text description of study background, objectives, and rationale
Design	Design of trial, statistics used, details of randomization and allocation concealment, and subgroup information
Entrance criteria	Inclusion and exclusion criteria
Interventions	Description of interventions and co-interventions, details of treatment masking, and administration
Outcome variables	Definitions of outcome variables, baseline characteristics, side effect variables, details of outcome analysis, assessment, and measurement
Enrollment	Recruitment, screening, enrollment of subjects, and number of subjects in each intervention group
Follow-up and compliance	Follow-up of subjects, compliance, and crossovers
Results	Quantitative study results
Conclusions and publications	Discussion of study's limitations and conclusions, citations to trial publications

As shown in Table 4, RCT Schema provides information over a trial's administration, design, execution and results, as well as, over their subclasses. More specifically:

- *The TRIAL class* can be described by the MAIN-STUDY and SECONDARY-STUDY classes, each of which contain information regarding clinical and scientific background, publication and administrative data, etc.
- *The PROTOCOL class* is another way to describe the study classes above. This class consists of two categories, the INTENDED-PROTOCOL and the EXECUTED-PROTOCOL. The information provided by this class describes the study design, the subject inclusion criteria, recruitment, randomization, follow-ups, adverse events, etc.
- *The SAMPLE-SIZE-CALCULATION class is the one that describes the statistical design of a MAIN-STUDY.*
- *The POPULATION class* (name, size, age and gender) is divided into three subclasses: RECRUITED-POPULATION (its subclasses refer to screened, eligible, enrolled and randomized subjects), EXCLUDED-POPULATION (its subclasses refer to screened but not eligible, eligible but not enrolled, enrolled but randomized and randomized but excluded from intention-to-treat analysis subjects) and ANALYZED-POPULATION.
- *AGE-GENDER-RULE class.*
- *ETHNICITY-LANGUAGE-RULE class.*
- *CLINICAL RULE class* (number of subjects agreeing or not to this rule)*:* its subclasses are RECURSIVE-RULE and BASE-RULE (UMLS terminology describes these rules).
- *The INTERVENTION*-ARM *class* (UMLS terminology is used). This class is described by other information regarding the type of intervention, i.e., device, drug, procedure, other intervention (EXPERIMENTAL-ARM class). Moreover, the COMPARISON-ARM class can be the EXPERIMENTAL-ARM, PLACEBO, USUAL CARE and NO-TREATMENT classes.
- *The OUTCOME class*: PRIMARY-OUTCOMEs, SECONDARY- OUTCOMEs, BASELINE-CHARACTERISTICS, SIDE-EFFECTS and OUTCOME-ASSESSMENT class. For each outcome, information can be described by the STATISTICAL-ANALYSIS-AND-RESULTS class and the REGRESSION-ANALYSIS-AND-RESULTS class (Sim, Olasov and Carini, 2004).

It is crucial to establish a standardized clinical vocabulary in order for the RCT Schema to be efficient and useful. The more terms included in the vocabulary, the more useful the information system will be (Sim, Olasov and Carini, 2004).

Although Sim, Olasov and Carini (2004) developed this ontology in order to improve evidence-based clinical care and practice, it seems like this system might be able to support the first step of a clinical trial: Conceiving the research question. With the developed tasks and subtasks, a systematic review of text-based publications of clinical trials can be less time-consuming, especially if investigators stored their findings into knowledge bases that can be understood by computers (2004).

- **Patient Clinical Data (PCD) Ontology:** Classifying and organizing information.

In 2018, Boshnak et al. built the Patient Clinical Data (PCD) Ontology. According to them, the development of the PCD ontology aims at the representation of clinical data related to Electronic Health Records (EHR). These clinical data are collected during patient visits, present a patient's health status and their accurate representation via PCD will enable researchers and other clinical professionals to have adequate information on a patient's condition. For building the PCD ontology, the first step is the specification of the goal of the ontology (collecting data). The second step is called "Knowledge Acquisition" and its sub-steps are: (a)study clinical terms in an informal way, (b) build the Glossary of Terms (GT) of clinical concepts in a formal way (via unified medical language systems-UMLS), (c)integrate the medical ontologies, create classes, subclasses and their relationships and classify the clinical/medical terms into classes, (d) evaluate the PCD ontology by interviewing physicians and other EHR users. During the third step of the PCD ontology building, "Conceptualization", the stages are: (a) develop the Glossary of Terms (GT) of PCD Ontology by using UMLS (Fig. 24), (b) define and organize classes (cluster synonymous terms and link concepts – use the combination approach: top-down and bottom-up approach), (c) develop the concept dictionary (concept name, class attributes, relations), (d) define table of class attributes (the table should include defined concept, attribute name, value type, cardinality

and for defining the table Fast Healthcare Interoperability Resources – FHIR (HL7.org, 2018)– is used), (e) develop the Conceptual Model of PCD Ontology (Fig. 25) (Boshnak et al., 2018).

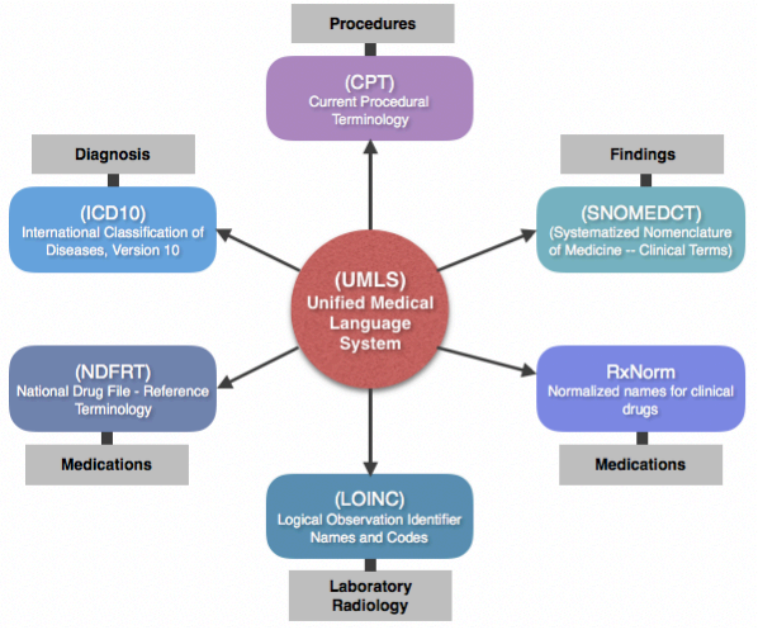

Figure 24. The UMLS source vocabularies used in the PCD ontology (Boshnak et al., 2018).

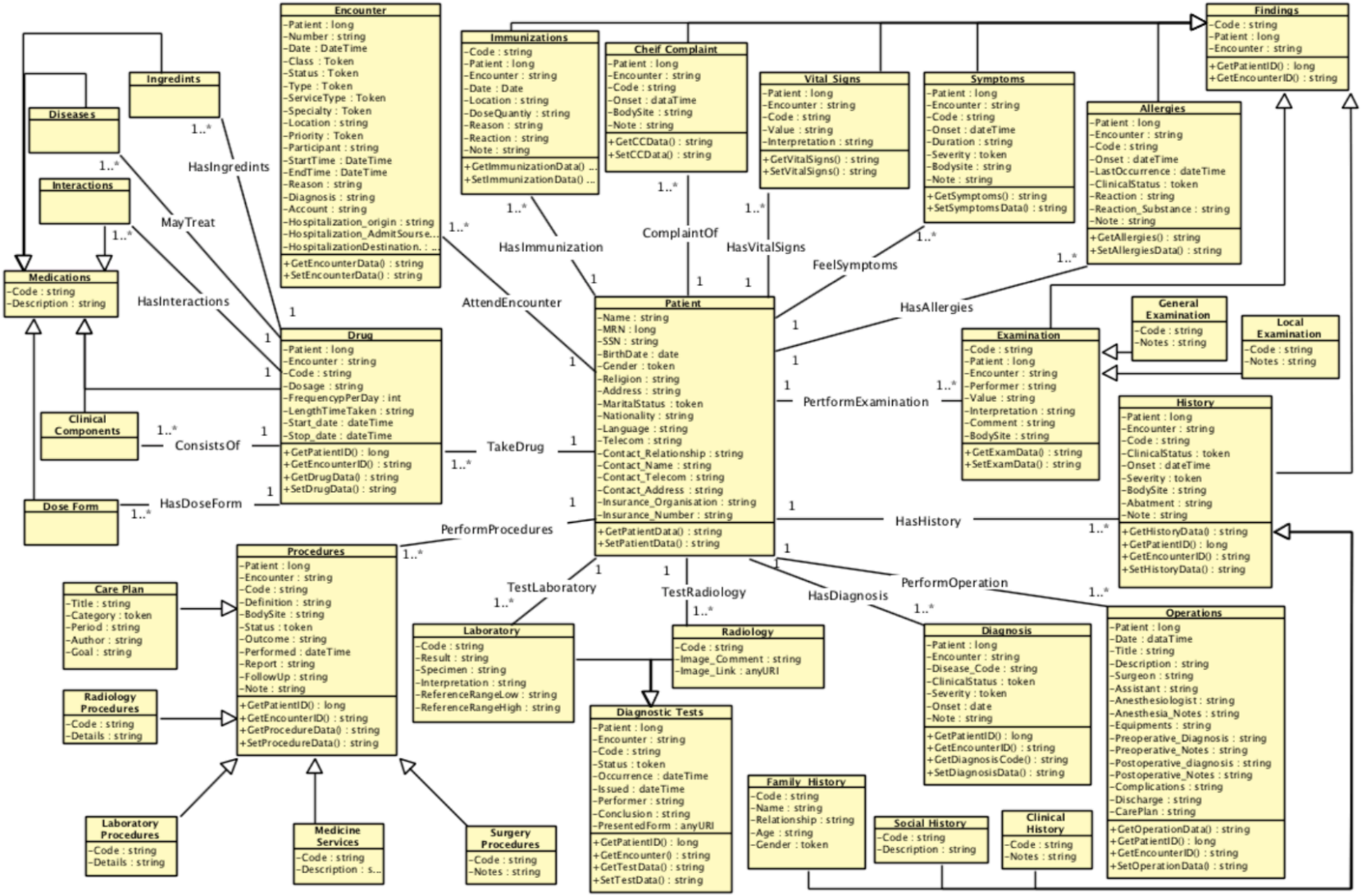

Figure 25. The Conceptual Model of the PCD Ontology (Boshnak et al., 2018).

“Integration” is the fourth step of building the PCD ontology and it includes looking for concepts in the UMLS, importing classes of interest from the system and integrating these classes with the PCD ontology. The fifth step is “Implementation” and the outcome is the coding of the ontology which will lead to a formal ontology development language (Boshnak et al., 2018).

The final outcome of the PCD Ontology is the classification of information. The main classes of this ontology are: Patient, Encounter, Findings, Diagnosis, Procedures, Medications, Operations and Diagnostic tests (Table 5). Moreover, the ontology provides a section called “Properties”, which is divided into object properties (32) and data properties (170). The former properties describe the connection between two objects (e.g., a patient has a disease), while the latter properties provide information about an object and its data (e.g., an address, an emergency contact, etc.) (Boshnak et al., 2018).

Table 5. Patient Clinical Data (PCD) Ontology: Classes and Information provided by each class (Boshnak et al., 2018).

Classes	Description - Information provided
Patient	Provides demographic and clinical data gathered for each patient.
Encounter	Provides information about every contact between patient and provider. Information about patient type, location, specialty, physician, appointment, hospitalization, patient registration.
Findings	Provides information about activities during the encounter between apatient and a physician, e.g., history regarding a patient's complaints, symptoms, vital signs, allergies, immunizations, general and local examination.
Diagnosis	Includes patients' diseases and conditions (the classification is based on the International Classification of Diseases Tenth Revision - ICD10). Its subclasses (19) are the main types of diseases. Information provided by this class: patient's condition, clinical status, the severity of a patient's condition according to the clinician's assessment, the start date of the disease, notes regarding clinical decision.
Procedures	Consists of 5 subclasses. Each subclass is a type of a procedure determined by the clinician according to the examination and diagnosis process.
Medications	Consists of 35 subclasses. For each drug, information regarding treatment plans and instructions is provided.
Operations	Provides information about a patient's performed operations, the attending and responsible medical/clinical stuff, the operation equipment, the preoperative and postoperative diagnosis, etc.
Diagnostic tests	Consists of radiology (17 subclasses) and laboratory (18 subclasses) tests.

The two final steps of building the PCD Ontology are “Evaluation” (the ontology ought to be consistent and the patient clinical data ought to be valid) and “Documentation” (for enabling the use and re-use of the ontology) (Boshnak et al., 2018).

Patient Clinical Data (PCD) Ontology provides a platform in which clinical data can be organized, and therefore, retrieved by healthcare professionals. Although Boshnak et al. proposed this ontology for better clinical decision making and access by researchers and other healthcare professionals, this ontology seems adequate to be used for better tracking of clinical trials participants, especially if this ontology is used by as many clinical trial settings as possible. With the PCD Ontology, clinical data saved in the EHRs (e.g., from study visits and follow-up visits) can be accessible and usable by investigators and therefore, investigators will be able to be fully aware of each participant’s health status and compliance (physically and psychologically). Moreover, the recruitment process might become less time-consuming, because of the common Glossary of Terms and the connection between all the classes mentioned above, as provided by the PCD Ontology (2018).

- **The Semantic Electronic Health Record (SEHR) ontology**

The Semantic Electronic Health Record (SEHR) ontology is based on the PPEPR Methodology (Sahay, Akhtar and Fox, 2008) (Fig. 26) and its development is a step of the Linked2Safety EU project. Analysis of the Linked2Safety project was provided above, but a brief summary is presented at this point. Linked2Safety aims at the development of a platform for patient recruitment across European clinical trials units and a global clinical terminology which will include and integrate several local terminologies, “a consistent, unambiguous and unifying framework of terminology”. Therefore, the SEHR ontology is presented as a solution to distributed clinical sites (Sahay et al., 2011).

Based on the PPEPR Methodology for building an ontology, SEHR ontology proceeds with an alignment of local and global ontologies by using three approaches (independently or in combination) for adapting the local concept. The first approach is called “Top-Down” and

suggests a specialization of the global ontology in order for global concepts to resemble local ones. The second approach, the "Bottom-Up" approach, includes an extension of the local ontology and its generalization in order for local concepts to resemble global ones. Finally, the "Middle-Out" approach (the third approach) proposes the development of an additional specialized class similar to the concepts. In other words, SEHR ontology is a clinical trial ontology which aims at the layering and adaptation of clinical terminologies. During the layering step, global and local ontologies are processed, and the result is the production of layered ontologies and the arrangement of the ontologies into global and local spaces. During the adaptation step, after the local and global ontologies are processed, the outcome is an extended version of layered ontologies which is adapted to the local requirements. Hence, clinical trial data integration is more feasible (Sahay et al., 2011).

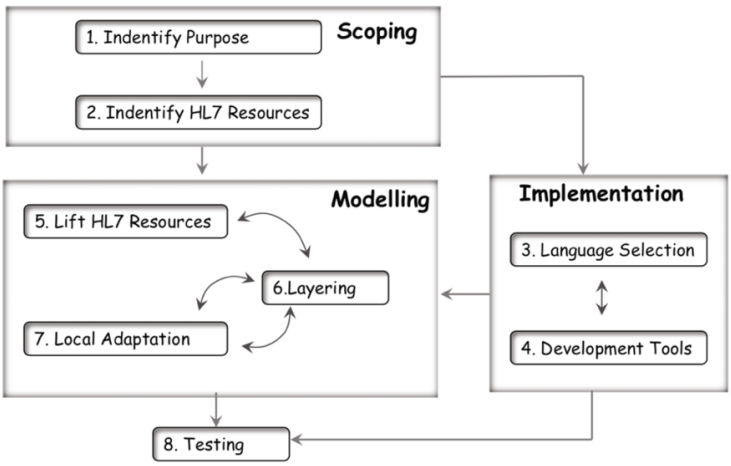

Figure 26. PPEPR Ontology Building Methodology (Sahay et al., 2011).

Evaluation and Categorization of Information Systems and Ontologies in Clinical Research

Information systems have been researched and introduced in the healthcare sector and have offered benefits throughout many processes. It is important, though, that each healthcare information system should be evaluated, and according to Friedman and Wyatt (1997) and Odhiambo-Otieno (2005) this evaluation is necessary for at least five reasons. First, by evaluating information systems, healthcare professionals become more acquainted with the idea of a cost-effective and safe way of managing data ("promotional" reason). Second, scholars will continue their research on more accurate and effective information systems ("scholarly" reason). Third, via the evaluation process, mistakes from previous systems are presented and the more effective methods are pointed out ("pragmatic" reason). Fourth, healthcare professionals can measure the safety in correspondence to the cost and therefore, be ethical ("ethical" reason). Last, evaluation can prove another aspect of the effectiveness of a system which is the accuracy of the information provided to healthcare professionals ("medicolegal" reason). Hence, before proceeding with the establishment and use of an information system, one or more of the above aspects should be considered and the healthcare professionals should decide accordingly.

According to Crepaldi et al. (2018), a complete Health Information System (and therefore, Clinical Research Information Systems and Ontologies) should be evaluated according to five pillars: "the perspective of patients, health professionals, software engineering, security and safety, and managerial issues". Many other researchers would agree with Crepaldi et al. on the points about the evaluation of staff satisfaction (Gugerty, Maranda and Rook, 2006; Hanmer, 1999; Wang and Yang, 2007), the evaluation of security and safety being extremely important (Low and Chen, 2012; Smith et al., 2011) and about the assessment of managerial issues, such as quality, validity, economics, efficiency and effectiveness and flexibility (Hanmer, 1999; Low and Chen, 2012; Odhiambo-Otieno, 2005; Richesson and Andrew, 2019; Smith et al., 2011; Wang and Yang, 2007). In addition to the criteria mentioned above, two crucial criteria, common in the perspectives and suggestions of all the mentioned researchers, is usability (or easy-to-use ability) and functionality

(it includes: security and confidentiality, data consistency and health information standards, appropriateness and accessibility and user acceptability) (Hanmer, 1999; Richesson and Andrews, 2019).

Moreover, according to Park et al. (2018), information systems used in clinical trials should be able to support one or more than one steps and phases of a clinical study's process and one or more than one therapeutic/disease areas. Also, information systems should be characterized according to their access protocols, their data's safety and security, their "controlled vocabulary", their relation or not with existing databases, such as Electronic Health Records (EHRs) and Hospital Information Systems (HIS) and their application or not of a notification system, such as a notification system which will send notifications whenever a patient is eligible for a clinical study's recruitment (Park et al., 2018).

Therefore, according to this literature review over the evaluation and categorization of clinical information systems and clinical research information systems and to some conclusions made after reviewing the clinical research process, information systems and ontologies, the criteria designated in order to evaluate and categorize these information systems (IS) are:

- Usability (easy-to-use system)
- Flexibility, Scalability
- Accessibility (Authorized or open-to-the-public access)
- Data Security/Safety/Privacy
- Data quality (accuracy, consistency, agreement with health information standards)
- Automation
- Standardized vocabulary/terminology
- Ability to support multi-centered trials
- The clinical study's step(s) the IS can support
- The clinical study's phase(s) the IS can support
- The study design type(s) the IS can support (Observational/ Interventional)
- The therapeutic/disease area(s) the IS can support
- The category(/-ies) of clinical research data the IS can provide

- The IS's relation to other databases (e.g., HER or HIS)
- The application of a notification system, e.g., alerts for eligible patients.

Unfortunately, due to restricted information regarding staff and patient satisfaction, evaluation based on these criteria will not be presented (despite their being important attributes for Clinical Information Systems assessment).

Evaluation and categorization of the information systems, ontology-based information systems and ontologies analyzed in this thesis are presented in the tables below (Tables 6-11). The evaluation and categorization were performed based on the criteria mentioned above.

Table 6. Evaluation of the information systems regarding data security/safety/privacy and data quality.

	Data Security/Safety/Privacy	**Data Quality**
Tools in a CIS	Needs evaluation	Needs evaluation
CDW-CIS	Yes (Anonymization)	Yes (it is used for quality management)
COATI	Yes (patients numbers are used)	Yes (GCP Guidelines)
HIS-based patient recruitment	Yes	Yes (supports quality management)
Single Source IS	Yes (access is controlled, but needs imporovement and approvals - data protection issues need to be observed strictly)	Mechanisms for quality control and validation will be provided
CRDW	Yes	No (Restricted)
ISCO	Yes	Yes (quality control tool)
AEDAMS	Yes	Yes (quality checks, improved data quality)
Active Computerized Pharmacovigilance	Unclear	Medium
National Clinical Trials Registry	Yes (patient data is not provided by the registry - only information regarding the study is provided)	Yes (data validation, but data should be monitored and improved at all times
OpenTrials	Yes (secure patient data access via ClinicalStudyDataRequest.com and other databases)	Yes

Table 7. Evaluation of the ontology-based information systems and ontologies regarding data security/safety/privacy and data quality.

	Data Security/Safety/Privacy	**Data Quality**
Epoch	Yes	Yes (tracking of data)
OnWARD	Yes (Role-based acces control mechanism)	Yes (Input Validation)
A semantic interoperability layer	Yes (Filtered data based on the user - in the future)	Adequate
The Linked2Safety project	Yes ("closed-world" room)	Yes (Quality Control)
ODaCCI	Yes (access only to de-identified patient data)	Yes (automatic data validity checks)
Recruit	Yes (Strict patient data access)	Unclear
SemEHR	Unclear	Yes (accuracy)
OBX	No	Unclear
ObTiMA	Yes (security architecture)	Increased, but needs improvement
RCT Schema	Unapplicable (it is an ontology)	Yes (the otology is competent for the tasks responsible for)
PCD Ontology	Unapplicable (it is an ontology)	Unclear
SEHR Ontology	Unapplicable (it is an ontology)	Yes (quality control - SEHR is part of the Linked2Safety project)

Table 8. Evaluation of the information systems.

	Usability	Flexibility/Scalability	Accessibility	Automation	Standardized Vocabulary/Terminology	The IS's relation to other databases (e.g., EHR or HIS or CIS, etc.)	The application of a notification system
Tools in a CIS	Yes (increased usage of the system)	No (only hospitals with the same CIS)	Access to trial registry: resrtricted to selected trial coordinator	Automated alerts / Automated reports editing/ not automated electronic transcripts from the CIS to trial databases	Unclear	Tools are implemented on CIS	Yes (hospital sdmission alerts)
CDW-CIS	Yes	Unclear	Authorized	Unclear	Unclear	Extracts data from EHRs and CISs	No
COATI	Yes (by people familiar to clinical trials)	Yes	Via the World Wide Web	Not complete	Yes	None	No
HIS-based patient recruitment	Yes	No	Authorized	Yes	Unclear	Yes (HIS-based)	Yes (electronic alerts)
Single Source IS	Yes (the HIS is used)	Yes (every hospital is responsible for its HIS)	Physicians and nurses: access to the HIS Study data management team: access to the clinical research database	Automatically and manually used system	No (different terminologies and standards - different HISs)	Based on HIS	Yes (automated notification for potential trial subjects via email)
CRDW	No (uncertainty)	Unclear	Authorized (via a web interface)	Yes	Yes (German mostly)	Yes (EHR & data are extracted from the CIS)	No
ISCO	Yes (user support is available and training-time is low)	No (too specific)	Only to ISCO users	Not complete	Yes (data dictionary was developed), but standardization of coding systems will be needed	Yes (development of the HIS)	No
AEDAMS	Yes (average training-time is 30 minutes)	Yes (scalable to different protocols and to support larger number of events)	Authorized ("users are only granted access after verification of their role on the study by the administration")	Automation (Automated tool)	Yes	None	Yes (notification is done by automatically generated emails)
Active Computerized Pharmacovigilance	Unclear	Unclear	Unclear	Yes (the system can automatically detect new adverse events by using only the existing information in EHR systems)	Yes (MedLEE: Medical Language, Etraction and Encoding system)	"Applied on comprehensive unstructured clinical data from EHRs"	No
National Clinical Trials Registry	Yes (easily operated by patients)	Yes (designed for all types of clinical studies)	Open to the public	Automated	Yes (vocabulary collection)	None	No
OpenTrials	Yes (user-friendly web interfaces)	Yes ("data that can be freely used, modified, and shared by anyone for any purpose" - flexible schemas and data structures)	Open database	Automatically and manually used system	A common dicitionary must be developed - For now: wide categories that can be interpreted in many ways have been introduced	Extracts data from registers	No

Table 9. Evaluation of the ontology-based information systems and ontologies.

	Usability	Flexibility/Scalability	Accessibility	Automation	Standardized Vocabulary/Terminology	The IS's relation to other databases (e.g., EHR or HIS or CIS, etc.)	The application of a notification system
Epoch	No ("cumbersome")	Yes	Authorized (knowledge Base Server - API)	Yes	Yes	None	Yes (ImmunoTrak)
OnWARD	Yes (minimal training)	Yes	Authorized (role-based acces control mechanism)	Semi-automatic (automatic data validation)	Yes (Clinical research ontology)	None	No
A semantic interoperability layer	Unclear	Yes	Filtered data based on the user - in the future	Yes (automated terminology binding)	Yes (the system provides a common, standardized terminology)	Clinical Trial Management Systems (CTMS), Laboratory Information Management Systems (LIMS), Electronic Health Record (EHR), etc.	No
The Linked2Safety project	Unclear	Yes	Researchers and healthcare professionals with access to the platform (Authorized)	Yes	Yes	Analysis and connection of distributed EHRs	No
ODaCCI	Unclear (Not evaluated)	Yes (adaptable due to its general design)	Access only to de-identified patient data	Automatically and manually used system	Yes (Epilepsy and Seizure Ontology - EpSO)	Sometimes data from EHRs are used	No
Recruit	Yes (15 different users/week)	Yes ("its modular design could be applied to other clinical conditions and hospitals")	Strict patient data access	Unclear	Yes	Extracts data from EHRs	No
SemEHR	Unclear	Yes	Unclear	Yes	Yes	Uses data from EHRs	No
OBX	Unclear	Yes (additions can be made easily)	Web-based public resource	Unclear	Yes	Public data repositories (e.g., GenBank, UniProt, the Immune Epitope Database)	No
ObTiMA	Yes (users do not need to know technical details - straightforward - end-user usability was ensured)	Yes (adaptable)	Authorized	Automated IS	Yes (Master ontology)	None	No
RCT Schema	Unclear	Yes	Unclear	Needs to be automated	Yes (UMLS - SNOMED CT in the future))	None	No
PCD Ontology	Yes	Unclear	Accessible to researchers and healthcare facilities users	Unclear	The PCD ontology developes a Glossary of Terms (by using UMLS) and integrates medical ontologies	Data are related to EHRs	No
SEHR Ontology	Medium (most of the users claim to understand the ontology, but additional support should be provided)	Yes (used in many health care sub-domains)	Researchers and healthcare professionals with authorized access to the platform (SEHR is part of the Linked2Safety project)	Unclear	The SEHR ontology provides an integration of local and global ontologies	None	No

Table 10. Categorization of the information systems.

	The clinical study's step(s) the IS can support	The clinical study's phase(s) the IS can support	The study design type(s) the IS can support (Observational/ Interventional)	The therapeutic/disease area(s) the IS can support	The category(/-ies) of clinical research data the IS can provide	Ability to support multi-centered clinical studies
Tools in a CIS	Step 2: Designing a new study & Step 3: Study execution & Step 4: Report Results & Step 5: Interpret the results and apply them to clinical care and policy	Phase I, II, III/IV	Interventional	Variety of disease/therapeutic areas	Data for recruitment and adverse events & Data from stored specimens, images, etc. and from the medical examinations during the trial & Follow-up data	No
CDW-CIS	Step 3: Study execution	Phase I, II, III/IV	Observational and Interventional	Unclear	Follow-up data & Data from the process and outcomes of old trials	Yes
COATI	Step 2: Designing a new study & Step 3: Study execution & Step 4: Report Results	Phase II, III & IV	Observational and Interventional	Any ("Variety of medical conditions")	Data from the participants' medical examinations during the trial & Data for the statistical analysis & Data for adverse events	Yes
HIS-based patient recruitment	Step 3: Study execution	Phase I, II & III	Intervnetional	Acute Myeloid Leukemia (AML), neurology, dermatology	Data for selection criteria, sampling and recruitment	Yes (only few centers)
Single Source IS	Step 3: Study execution	Unapplicable (only for interventional studies)	Observational (for now)	Prostate cancer & Leukemia-only for recruitment	Data for selection criteria, sampling and recruitment	No (monocentric studies)
CRDW	Step 3: Study execution	Phase I, II & III	Interventional	Neurology (e.g., stroke, epilepsy)	Data for selection criteria, sampling and recruitment	Yes (only few centers)
ISCO	Step 3: Study execution	Phase I, II & III	Intervnetional	Oncology	Data for selection criteria, sampling and recruitment &Follow-up data	No
AEDAMS	Step 3: Study execution	Phase I, II, III/IV	Observational and Interventional (variety of study designs)	Variety of disease areas	Data for adverse events	Yes (multi-national trials)
Active Computerized Pharmacovigilance	Step 3: Study execution	Phase I, II, III/IV	Intervnetional	Variety of disease areas	Data for adverse events	Yes
National Clinical Trials Registry	Step 3: Study execution & Step 4: Report Results & Step 5: Interpret the results and apply them to clinical care and policy	Phase I, II & III	Interventional	Variety of disease areas	Data for recruitment & Data found in publications regarding the trial & Data from the process and outcomes of old trials	Yes (information regarding multiple locations is provided)
OpenTrials	Step 1: Conceiving the research question & Step 3: Study execution & Step 4: Report Results & Step 5: Interpret the results and apply them to clinical care and policy	Phase I, II & III	Interventional	Variety of disease areas	Data for recruitment & Data found in publications regarding the trial & Data from the process and outcomes of old trials	Yes ("Where the results on a trial have been reported in multiple different places, a researcher can rapidly review these side by side.")

Table 11. Categorization of the ontology-based information systems and ontologies.

	The clinical study's step(s) it can support	The clinical study's phase(s) it can support	The study design type(s) it can support (Observational/ Interventional)	The therapeutic/disease area(s) it can support	The category(/-ies) of clinical research data it can provide	Ability to support multi-centered clinical studies
Epoch	Step 3: Study execution	Phase I, II & III	Interventional	Immune-mediated disorders (autoimmune diseases, islet, kidney and liver transplantation, allergy, asthma)	Follow-up data (tracking patients and clinical specimens) & Data Data from stored specimens, images, etc.	Yes
OnWARD	Step 2: Designing a new study & Step 3: Study execution & Σtep 4: Report Results	Phase II	Interventional	Heart Biomarker Evaluation in Apnea Treatment	Data for selection criteria, sampling and recruitment	Yes (small to medium scale clinical trials)
A semantic interoperability layer	Step 2: Designing a new study & Step 3: Study execution & Step 4: Report Results	Phase I, II & III	Interventional	Oncology (Breast Cancer)	Data for selection criteria, sampling and recruitment & Data for the statistical analysis	Yes
The Linked2Safety project	Step 3: Study execution	Phase III & Phase IV	Observational & Interventional	Variety of areas, especially chemoinformatics	Data for selection criteria, sampling and recruitment & Data for adverse events in multi-sited trials	Yes
ODaCCI	Step 3: Study execution	Phase I, II, III/IV (for ongoing trials)	Observational and Interventional	Neurological Disorders and Stroke (Epilepsy and Seizures)	Data for recruitment & Data from the participants' medical examinations during the trial & Data for statistical analysis	Yes
Recruit	Step 3: Study execution	Phase I, II & III	Interventional	Oncology	Data for recruitment	Yes (data from several databases)
SemEHR	Step 3: Study execution	Phase I, II & III	Observational and Interventional	Dermatology Disorder, Liver Disease (it has been applied on clinical studies in this areas)	Data for selection criteria, sampling and recruitment	Yes
OBX	Step 1: Conceiving the research question & Step 4: Report Results	Phase I & II	Observational and Interventional	Immunology (Allergy, Infectious Diseases)	Data from stored specimens, images, etc. from previous studies & Data from the process and outcomes of old trials & Data for adverse events	Yes
ObTiMA	Step 2: Designing a new study & Step 3: Study execution	Phase I, II, III/IV	Intervnetional	Oncology (but other domains, as well)	Follow-up data	Yes
RCT Schema	Step 1: Conceiving the research question & Step 5: Interpret the results and apply them to clinical care and policy	Phase I, II & III	Interventional (Randomized Clinical Trials)	Already used for cardiology, radiology, geriatrics and psychiatry	Data found in publications regarding the trial & Data from the process and outcomes of old trials (Data for reporting and analysis)	Unclear
PCD Ontology	Step 3: Study execution	Phase I, II, III/IV	Observational and Interventional	Variety of disease areas	Data from the participants' medical examinations during the trial & Data for recruitment & Follow-up data & Data regarding medicine clinical terms	Yes
SEHR Ontology	Step 3: Study execution (SEHR is part of the Linked2Safety project)	Phase III & Phase IV (SEHR is part of the Linked2Safety project)	Observational & Interventional (SEHR is part of the Linked2Safety project)	Cardiovascular disease, Migraine, Psychiatric Disorder, Breast Cancer, Diabetes, Genetics, Neurology	Data regarding medicine clinical terms and drug terms used globally (standardized "dictionaries")	Yes

Synthesis of the Information Systems (IS) and Ontologies (or Ontology-based IS)

As the goal of this thesis is the development of a guide for choosing the appropriate information systems while conducting a clinical study, a synthesis of the 23 information systems and ontologies and ontology-based information systems presented above ought to be done. This synthesis will be based on the evaluation and categorization tables shown above (Table 6).

The first step of the synthesis would be the categorization of the systems according to the study design type, study step and therapeutic/disease area each system can support, according to the clinical research data it can provide and according to the information system's ability to support multi-centered clinical studies. Due to restricted information regarding the study phase that each information system/ontology supports, the systems will not be categorized based on this criterion (the study phase(s) that each system supports is presented in the table according to the clinical study steps that each system supports and the data it provides). However, it should be mentioned that the systems which can be used for phase IV studies (as presented in Chapter 1) are: Tools in a CIS, CDW-CIS, AEDAMS, Active Computerized Pharmacovigilance, ObTiMA and PCD Ontology, according to the data they provide, and Linked2Safety, COATI and SEHR, according to the literature.

Categorization according to study design type(s) the IS or ontology can support (Tables 10 & 11)

This categorization was accomplished based on the studies that the information systems and ontologies have been used. However, an information system or ontology which contains and provides data for a clinical trial might, also, be able to provide data for an observational study and vice versa, because both categories of clinical studies are in need of almost the same data.

- *Information Systems and ontologies supporting both Observational and Interventional Studies:*

CDW-CIS, Linked2Safety, ODaCCI, a Database System (COATI/PANDA/DART), SemEHR, OBX, AEDAMS, PCD Ontology, SEHR Ontology

- *Information Systems and ontologies supporting only Observational Studies:*

Single Source IS.

- *Information Systems and ontologies supporting only Interventional Studies (clinical trials):*

Tools in a CIS, Epoch, OnWARD, a Semantic Interoperability Layer, HIS-based Recruitment, CRDW, Recruit, ISCO, Active Computerized Pharmacovigilance, ObTiMA, RCT Schema, National Clinical Trials Registry, OpenTrials.

Categorization according to study step(s) the IS or ontology can support (Tables 10 & 11)

This categorization was achieved based on personal opinion which was shaped based on the information provided by each system; in order for an information system/ontology to support a study step, certain data must be included.

- *Step 1. Conceiving the research question*

OBX, RCT Schema, OpenTrials.

- *Step 2. Designing a new study*

Tools in a CIS, OnWARD, a Semantic Interoperability Layer, a Database System (COATI/PANDA/DART), ObTiMA.

- *Step 3. Study execution*

Tools in a CIS, Epoch, OnWARD, a Semantic Interoperability Layer, CDW-CIS, Linked2Safety, ODaCCI, a Database System (COATI/PANDA/DART), HIS-based Recruitment, Single Source IS, CRDW, Recruit, SemEHR, ISCO, AEDAMS, Active Computerized Pharmacovigilance, ObTiMA, OpenTrials, PCS Ontology, SEHR Ontology.

- *Step 4. Report Results*

Tools in a CIS, OnWARD, a Semantic Interoperability Layer, a Database System (COATI/PANDA/DART), OBX, National Clinical Trials Registry, PCD Ontology.

- Step 5. *Interpret the results and apply them to clinical care and policy*

Tools in CIS, RCT Schema, National Clinical Trials Registry, OpenTrials.

Categorization according to disease/therapeutic area the IS or ontology can support (Tables 10 & 11)

This table's element was completed based on the disease/therapeutic areas that the information systems and ontologies have been used with. In some cases, it was mentioned in the literature that a variety of disease areas can be supported, while in some other cases no information regarding the disease area was provided (unclear).

- *Variety of disease areas:* Tools in a CIS, Linked2Safety, a Database System (COATI/PANDA/DART), AEDAMS, Active Computerized Pharmacovigilance, National Clinical Trials Registry, OpenTrials, PCD Ontology.
- *Immunology:* Epoch (Immune-mediated disorders), OBX (allergy, infectious disease).
- *Heart Biomarker Evaluation in Apnea Treatment:* OnWARD.
- *Oncology:* a Semantic Interoperability Layer (Breast Cancer), HIS-based Recuitment (AML), Single Source IS (prostate cancer, leukemia), Recruit, ISCO, ObTiMA, SEHR Ontology (breast cancer).
- *Neurology:* ODaCCI (epilepsy, seizures), HIS-based Recruitment, CRDW (stroke, epilepsy), SEHR Ontology (migraine).
- *Dermatology:* HIS-based Recruitment, SemEHR.
- *Liver Diseases:* Epoch, SemEHR.
- *Cardiology:* RCT Schema, SEHR Ontology (cardiovascular disease).
- *Radiology:* RCT Schema.
- *Geriatrics:* RCT Schema.
- *Psychiatry:* RCT Schema, SEHR Ontology.
- *Diabetes:* SEHR Ontology.
- *Genetics:* SEHR Ontology.
- *Unclear:* CDW-CIS

Categorization according to the category(-ies) of clinical research data the IS or ontology can support (Table 10 & 11)

- *Data for selection criteria, sampling and recruitment*

Tools in a CIS, OnWARD, a Semantic Interoperability Layer, Linked2Safety, ODaCCI, HIS-based Recruitment, Single Source IS, CRDW, Recruit, SemEHR, ISCO, National Clinical Trials Registry, OpenTrials, PCD Ontology.

- *Data from stored specimens, images, etc. from previous studies*

Tools in a CIS, Epoch, OBX.

- *Data for adverse events*

Tools in a CIS, Linked2Safety, a Database System (COATI/PANDA/DART), OBX, AEDAMS, Active Computerized Pharmacovigilance.

- *Data regarding medical and clinical terms and drug terms used globally (standardized "dictionaries")*

PCD Ontology, SEHR Ontology.

- *Data found in publications regarding the trial*

RCT Schema, National Clinical Trials Registry, OpenTrials.

- *Data from the process and outcomes of old trials*

CDW-CIS, OBX, RCT Schema, National Clinical Trials Registry, OpenTrials.

- *Follow-up data*

Tools in a CIS, Epoch, CDW-CIS, ISCO, ObTiMA, PCD Ontology.

- *Data from the participants' medical examinations during the trial*

Tools in a CIS, OdaCCI, a Database System (COATI/PANDA/DART), PCD Ontology.

- *Data for the statistical analysis*

A Semantic Interoperability Layer, ODaCCI, a Database System (COATI/PANDA/DART).

Categorization according to the information system's and ontology's ability to support multi-centered clinical studies (Table 10 & 11)

- *Information Systems supporting multi-centered clinical studies*

Epoch, OnWARD (small to medium scale), a Semantic Interoperability Layer, CDW-CIS, Linked2Safety, ODaCCI, a Database System (COATI/PANDA/DART), HIS-based Recruitment (only few centers), CRDW (only few centers), Recruit, SemEHR, OBX, AEDAMS, Active Computerized Pharmacovigilance, ObTiMA, National Clinical Trials Registry, OpenTrials, PCD-Ontology, SEHR Ontology

- *Information Systems unable to support multi-centered clinical studies*

Tools in a CIS, Single Source IS, ISCO

- *Unclear:* RCT Schema

The evaluation of the information systems should be initiated by not selecting the systems and ontologies that do not achieve data quality and data security/safety/privacy requirements (Tables 6 & 7). This choice is based on the ICH-Good Clinical Practice (GCP) guidelines for clinical research, which emphasize the necessity of quality control which "should be applied to each stage of data handling to ensure that all data are reliable and have been processed correctly" (International Conference of Harmonization, 1996). Moreover, according to the ICH-GCP Guidelines, only authorized and secure data access should be allowed, approval for data changes must be given, data ought to be characterized by accuracy, and records that can identify subjects must remain secure for privacy matters (1996). More specifically, and as the guidelines for Good Clinical Practice claim, "Compliance with this standard provides public assurance that the rights, safety and well-being of trial subjects are protected, consistent with the principles that have their origin in the Declaration of Helsinki, and that the clinical trial data are credible" (ICH Expert Working Group, 1996). In some cases, these criteria were not mentioned in the literature (unclear) and these systems will not be included in this evaluation. Therefore, the information systems and ontology-based information systems with both data quality and data security/safety/privacy assurance, as shown in Tables 6 and 7, are: Epoch, OnWARD, a Semantic Interoperability Layer (adequate data quality), CDW-CIS, Linked2Safety, ODaCCI, a Database

System (COATI/PANDA/DART), HIS-based Recruitment, Single Source IS, ISCO, AEDAMS, ObTiMA (increased data quality, but needs improvement), National Clinical Trials Registry and OpenTrials. The next step is to evaluate the above information systems and the three ontologies (RCT Schema, PCD Ontology and SEHR Ontology) based on their usability, flexibility/scalability, automation and terminology/vocabulary used (standardized or not) (Tables 8 & 9).

Usability: From the information systems and ontologies with data quality and data security/safety/privacy assurance (presented above), only Epoch might lack in usability as it was characterized as "cumbersome". The level of usability of the Semantic Interoperability Layer, the Linked2Safety and the ODaCCI is unclear.

Flexibility/Scalability: The flexible information systems and ontology-based information systems are Epoch, OnWARD, Semantic Interoperability Layer, Linked2Safety, ODaCCI, a Database System (COATI/PANDA/DART), Single Source IS, AEDAMS, ObTiMA, National Clinical Trials Registry and OpenTrials. The HIS-based Recruitment and the ISCO are not flexible. The flexibility of the CDW-CIS is unclear.

Automation: The automated information systems and ontology-based information systems are: Epoch, Semantic Interoperability Layer, Linked2Safety, HIS-based Recuritment, AEDAMS, ObTiMA and National Clinical Trials Registry. The systems which are used both automatically and manually are: ODaCCI, Single Source IS and OpenTrials. The OnWARD is characterized as semi-automatic. The automation of the Database System (COATI/PANDA/DART) and the ISCO is not complete and the level of automation of the CDW-CIS is unclear.

Standardized Vocabulary/Terminology: The information systems and ontologies with standardized vocabularies are: Epoch, OnWARD, Semantic Interoperability Layer, Linked-Safety, ODaCCI, a Database System (COATI/PANDA/DART), ISCO, AEDAMS, ObTiMA and National Clinical Trials Registry. OpenTrials does not have a standardized vocabulary yet. For the CDW-CIS and the Single Source IS, the existence of a standardized vocabulary/terminology is unclear.

The information systems and ontologies with unclear information regarding usability, flexibility/scalability and standard vocabulary/terminology were excluded from the guidelines. According to the categorization and evaluation of the clinical research information systems and ontologies above, the guidelines for conducting a clinical study are:

For clinical trials (or interventional studies):

Step 1. Conceiving the research question: OpenTrials will provide data found in publications regarding a trial and data from the process and outcomes of old trials and researchers will be able to proceed with a systematic review and comparisons of older data. The National Clinical Trials Registry will provide data from the process and outcomes of old trials, as well.

Step 2. Designing a new study: OnWARD will help with searching for previously conducted clinical studies of interest and offer data for establishing selection criteria, sampling and recruitment, while ObTiMA will provide information for designing the study via its Trial Builder component. Moreover, the National Clinical Trials Registry will provide data for recruitment. The Database System (COATI/PANDA/DART) will offer data for adverse events.

Step 3. Study execution: Epoch will provide follow-up data and data for the specimens selected during the trial (therefore, Epoch will provide to future studies data from specimens from previous studies). OnWARD will provide data for selection criteria, sampling and recruitment. ODaCCI will provide data from the participants' medical examinations during the trial and for selection criteria, sampling and recruitment. ISCO will provide data for selection criteria, sampling and recruitment and follow-up data. ObTiMA will provide follow-up data. OpenTrials will provide data for selection criteria, sampling and recruitment and follow-up data. AEDAMS and the Database System (COATI/PANDA/DART) will provide data for adverse events. Database System (COATI/PANDA/DART) will, also, provide data for statistical analysis and data from medical examinations during the trial. The National Clinical Trials Registry will provide data for recruitment.

Step 4. Report Results: OnWARD, the National Clinical Trials Registry will enable researchers to develop reports regarding a clinical study and the Database System (COATI/PANDA/DART) will provide data for the adverse events that ought to be reported.

Step 5. Interpret the results and apply them to clinical care and policy: OpenTrials is a database which can provide effectively and correctly available results of clinical trials as it can provide data found in publications regarding a trial and data from the process and outcomes of old trials. Therefore, researchers will be able to proceed with a systematic review and comparisons of older data and maybe, a new research question will be conceived (the cycle will be complete).

For observational studies only ODaCCI, Single Source IS (which cannot support multi-centered studies), Database System (COATI/PANDA/DART) and AEDAMS can be used. Hence, no guidelines can be shaped for a complete observational study.

Limitations of the information systems and ontologies suggested for completing a clinical study are presented in Table 12.

Table 12. Limitations of the suggested information systems and ontologies.

	Limitations
Epoch	no usability
OnWARD	semi-automated
ODaCCI	incomplete automation
ISCO	no flexibility, incomplete automation
ObTiMA	–
National Clinical Trials Registry	–
OpenTrials	semi-automated, a commond dictionary ought to be developed
AEDAMS	–
Database System (COATI/PANDA/DART)	usable only for people familiar with clinical trials, incomplete automation

To sum up, researchers should take advice from the guidelines presented above regarding the information systems they should use for completing an interventional clinical trial. Moreover, clinical investigators should, also, choose which information system to use according to the disease/therapeutic area on which their research is based, such as oncology, neurology, etc. Finally, more information regarding the information systems/ ontologies accessibility, relation to other databases (EHRs, HIS, etc.) and application of a notification system is presented in Table 6 for the clinical study management team to select the most appropriate information system/ontology for their clinical study.

Conclusion, Limitations and Suggestions for Future Research

Guidelines for choosing the appropriate information systems for conducting an interventional clinical trial were developed. After categorizing and evaluating 23 information systems and ontologies, nine information systems were considered to be adequate for being used in a clinical research process (the data quality and data security/safety/privacy requirements were met, and "unclear" comments did not characterize them). These information systems are: Epoch, OnWARD, ODaCCI (Ontology-guided Data Curation for Multisite Clinical Research Data Integration**)**, ISCO (Information System for Clinical Organizations), ObTiMa (Ontology-based Trial Management Application), National Clinical Trials Registry, OpenTrials, AEDAMS (Adverse Event Data Management System) and Database System (COATI/PANDA/DART).

In the beginning of this research, the same guidelines for observational studies were to be developed, as well. Unfortunately, due to lack of information in the papers used for the synthesis, not enough information systems supporting observational studies could be included in the final evaluation. However, the study design type (Observational or Interventional) of each system was reported based on the clinical study that the system was implemented on and this does not mean that the system cannot be used in the other design type, as well; this comment can be considered as a limitation of this thesis. Another limitation of this thesis is that the evaluation did not include patient-reporting and patient satisfaction (whenever applicable) assessment due to unavailable information.

Insufficient information is a limitation in this thesis, as some papers did not provide adequate information regarding the evaluation criteria used in this thesis. Therefore, the information systems and ontologies with "unclear" comments in their evaluation were not included in the synthesis. Moreover, another limitation was that some papers did not provide information regarding the study phase(s), study step(s) and study design type(s) the information system or ontology can support. Hence, some information systems and ontologies were categorized based on the data they provide and not on the authors' suggestions (because they were not available).

These limitations could be resolved in future work by contacting the authors and creators of these information systems and requesting more information regarding the unclear parts of the systems' categorization and evaluation. Due to time restriction, communication with the authors was not feasible for this thesis.

Another suggestion for future work could be the conduction of a systematic review of ontologies used in clinical research. Information systems are continually being developed and standardized terminologies and vocabularies supporting them are an imperative part of their function. Moreover, as many information systems are combined for the completion of a clinical study, a specific ontology ought to be able to support all these systems and the disease/therapeutic area that the clinical study researches. Therefore, this need of a common ontology seems to be the reason for the high number of clinical study ontologies and ontology-based clinical research information systems found in the literature and analyzed in this thesis.

References

Alonso-Calvo, R., Perez-Rey, D., Paraiso-Medina, S., Claerhout, B., Hennebert, P. and Bucur, A. (2015). Enabling semantic interoperability in multi-centric clinical trials on breast cancer. *Computer methods and programs in biomedicine*, 118(3), pp. 322-329. Available at: http://dx.doi.org/10.1016/j.cmpb.2015.01.003.

Antoniades, A., Georgousopoulos, C., Forgo, N., Aristodimou, A., Tozzi, F., Hasapis, P., Perakis, K., Bouras, T., Alexandrou, D., Kamateri, E., Panopoulou, E., Tarabanis, K. and Pattichis, C. (2012). Linked2Safety: A secure linked data medical information space for semantically-interconnecting EHRs advancing patients' safety in medical research. In: *2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE)*, pp. 517-522. Doi: 10.1109/BIBE.2012.6399767.

Bailey, H. and Urquhart, C. (2003). Information needs in a clinical trials unit. *Health Informatics Journal*, *9*(2), pp. 111-126. Doi: 10.1177/1460458203009002006.

Basic-formal-ontology.org (2019). *BFO Basic Formal Ontology.* [online] Available at: http://basic-formal-ontology.org [Accessed 1 Aug. 2019].

Bellary, S., Krishnankutty, B. and Latha, M. S. (2014). Basics of case report form designing in clinical research. *Perspectives in Clinical Research,* 5(4), pp. 159. Doi: 10.4103/2229-3485.140555.

Boshnak, H., AbdelGaber, S., Abdo, A. and Yehia, E. (2018). Ontology-Based Knowledge Modelling for Clinical Data Representation in Electronic Health Records. *International Journal of Computer Science and Information Security,* 16(10), pp.68-86.

Chan, A. W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., Hróbjartsson, A., Mann, H., Dickersin, K., Berlin, J. A., Doré, C. J., Parulekar, W. R., Summerskill, W. S. M., Groves, T., Schulz, K. F., Sox, H. C., Rockhold, F. W., Rennie, D. and Moher, D. (2013). SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Annals of Internal Medicine.* 158(3), pp.200-207. Doi: 10.7326/0003-4819-158-3-201302050-00583.

ClinicalTrials.gov, (2018). *ClinicalTrials.gov Results Data Elements Definitions for Interventional and Observational Studies.* [online] Available at: https://prsinfo.clinicaltrials.gov/results_definitions.html [Accessed 12 Sep. 2019].

ClinicalTrials.gov, (2019). *Learn about clinical studies.* [online] Available at: https://clinicaltrials.gov/ct2/about-studies/learn#WhatIs. [Accessed 25 Jul. 2019].

ClinicalTrials.gov, (2019). *How to register your study.* [online] Available at: https://clinicaltrials.gov/ct2/manage-recs/how-register#RequiredRegistration. [Accessed 27 Jul. 2019].

Crepaldi, N. Y., de Lima, I. B., Vicentine, F. B., Rodrigues, L. M. L., Sanches, T. L. M., Ruffino-Netto, A., Alves, D. and Rijo, R. P. C. L. (2018). Towards a Clinical Trial Protocol to Evaluate Health Information Systems: Evaluation of a Computerized System for Monitoring Tuberculosis from a Patient Perspective in Brazil. *Journal of Medical Systems,* 42(6), pp.113. Available at: https://doi.org/10.1007/s10916-018-0968-8 .

Cui, L., Huang, Y., Tao, S., Lhatoo, S. D. and Zhang, G. Q. (2017). ODaCCI: Ontology-guided Data Curation for Multisite Clinical Research Data Integration in the NINDS Center for SUDEP Research. In: *AMIA Annual Symposium Proceedings*, pp. 441-450. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333343/pdf/2497467.pdf.

De Moor, G., Sundgren, M., Kalra, D., Schmidt, A., Dugas, M., Claerhout, B., ... and Kush, R. (2015). Using electronic health records for clinical research: the case of the EHR4CR project. *Journal of biomedical informatics*, 53, pp.162-173. Available at: http://dx.doi.org/10.1016/j.jbi.2014.10.006.

Dugas, M., Breil, B., Thiemann, V., Lechtenbörger, J., & Vossen, G. (2009). Single source information systems to connect patient care and clinical research. *Studies in health technology and informatics,* 150, pp.61-65. Doi: *10.3233/978-1-60750-044-5-61*

Dugas, M., Lange, M., Berdel, W. E. and Müller-Tidow, C. (2008). Workflow to improve patient recruitment for clinical trials within hospital information systems–a case-study. *Trials*, *9*(2). Doi: 10.1186/1745-6215-9-2.

Elwood, J. M. (1980). Interpreting clinical trial results: seven steps to understanding. *Canadian Medical Association Journal,* 123(5), pp. 343-345. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1704805/.

European Medicines Agency, (2006). *ICH Topic E5(R2) Guideline for Good Clinical Practice.* London: European Medicines Agency. Available at: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e6-r1-guideline-good-clinical-practice_en.pdf. [Accessed 29 Jul. 2019].

FDA.gov, (2018). *Informed Consent for Clinical Trials.* [online] Available at: https://www.fda.gov/patients/clinical-trials-what-patients-need-know/informed-consent-clinical-trials#What_is_Informed_Consent . [Accessed 27 Jul. 2019].

FDA.gov, (2018). *What Are the Different Types of Clinical Research?.* [online] Available at: https://www.fda.gov/patients/clinical-trials-what-patients-need-know/what-are-different-types-clinical-research. [Accessed 25 Jul. 2019].

FDA.gov (2019). *Institutional Review Boards Frequently Asked Questions. [*online] Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/institutional-review-boards-frequently-asked-questions. [Accessed 27 Jul. 2019].

Friedman, C. P. and Wyatt, J. C. (1997). *Evaluation Methods in Medical Informatics.* 1st ed. New York: Springer-Verlag.

Friedman, L. M., Furberg, C. D., Demets, D.L., Reboussin, D. M. and Granger, C. B. (2015). *Fundamentals of Clinical Trials.* New York:Springer, pp. 5-9.

Geibel, P., Trautwein, M., Erdur, H., Zimmermann, L., Jegzentis, K., Bengner, M., Nolte, C. H. and Tolxdorff, T. (2015). Ontology-based information extraction: identifying eligible patients for clinical trials in neurology. *Journal on Data Semantics*, *4*(2), 133-147. Doi: 10.1007/s13740-014-0037-5.

Geissbuhler, A. (1998). Clinical Information Systems - What is the bottom line? *Journal of the American Medical Informatics Association,* 5(6), pp.585-586. Doi:10.1136/jamia.1998.0050585.

Goldacre, B. and Gray, J. (2016). OpenTrials: towards a collaborative open database of all available information on all clinical trials. *Trials*, *17*(1), pp.164. Doi: 10.1186/s13063-016-1290-8.

Gugerty, B., Maranda, M. and Rook, D. (2006). The Clinical Information System Implementation Evaluation Scale. *Studies in health technologies and informatics,* 122, pp.621. Available at: https://www.researchgate.net/profile/Michael_Maranda/publication/6695972_The_clinical_information_system_implementation_evaluation_scale/links/546e1c5d0cf2bc99c21520c0.pdf.

Hanmer, L. (1999). Criteria for the evaluation of district health information systems. *International Journal of Medical Informatics*, 56(1-3), pp.161-168. Available at: https://doi.org/10.1016/S1386-5056(99)00046-5.

HL7.org, (2018). *HL7 FHIR Release 4.* [online] Available at: http://hl7.org/fhir/.

Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D. G. and Newman, T. B. (2013). *Designing Clinical Research*. 4th ed. Philadelphia: Wolters Kluwer.

ICH Expert Working Group. (1994). ICH Harmonised Tripartite Guideline for Clinical Safety Data Management: definitions and standards for expedited reporting E2A. In: *International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use*. Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf.

ICH Expert Working Group. (1996). ICH Harmonised Tripartite Guideline-Guideline for Good Clinical Practice E6(R1). In: *International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use*. Available at:

http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf.

Ichgcp.net, (2019). *5.18 Monitoring.* [online] Available at: https://ichgcp.net/518-monitoring/ [Accessed 29 Jul. 2019].

Kong, Y. M., Dahlke, C., Xiang, Q., Qian, Y., Karp, D. and Scheuermann, R. H. (2011). Toward an ontology-based framework for clinical research databases. *Journal of biomedical informatics*, *44*(1), pp. 48-58. Doi: 10.1016/j.jbi.2010.05.001.

Krishnankutty, B., Bellary, S., Kumar, N. B. R. and Moodahadu, L. S. (2012). Data management in clinical research: an overview. *Indian J. Pharmacol.* 44(2), pp. 168-172. Doi: 10.4103/0253-7613.93842.

Leroux, H., McBride, S. and Gibson, S. (2011). On selecting a clinical trial management system for large scale, multi-centre, multi-modal clinical research study. *Student Health Technol Inform.* 188, pp.89-95. Available at: https://www.ncbi.nlm.nih.gov/pubmed/?term=21893916.

Llewellyn-Bennett, R., Bowman, L. and Bulbulia, R. (2016). Post-trial follow-up methodology in large randomized controlled trials: a systematic review protocol. *Systematic reviews*. 5(1), pp.214. Doi: 10.1186/s13643-016-0393-3.

Low, C., and Chen, Y. H. (2012). Criteria for the evaluation of a cloud-based hospital information system outsourcing provider. *Journal of medical systems*, *36*(6), pp. 3543-3553. Doi: 10.1007/s10916-012-9829-z.

Mann, C. J. (2003). Observational research methods. Research design II: cohort, cross sectional, and case-control studies. *Emergency medicine journal*, *20*(1), pp.54-60. Doi: http://dx.doi.org/10.1136/emj.20.1.54.

McCray, A. T. and Ide, N. C. (2000). Design and implementation of a national clinical trials registry. *Journal of the American Medical Informatics Association*, *7*(3), 313-323. Available at: https://doi.org/10.1136/jamia.2000.0070313.

NHS.uk (2019). *NHS data model and dictionary.* [online] Available at: https://www.datadictionary.nhs.uk [Accessed 30 Jul. 2019].

Obofoundry.org (2019). *Ontology for Biomedical Investigations.* [online] Available at: http://www.obofoundry.org/ontology/obi.html [Accessed 1 Aug. 2019].

Odhiambo-Otieno, G. W. (2005). Evaluation criteria for district health management information systems: lessons from the Ministry of Health, Kenya. *International Journal of Medical Informatics*, *74*(1), pp.31-38. Doi: 10.1016/j.ijmedinf.2004.09.003.

Oliveira, A. G. and Salgado, N. C. (1999). A database system for integrated clinical trial management, control, statistical analysis and ICH-compliant reporting. *Proceedings. AMIA Symposium*, pp.72–76. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232832/pdf/procamiasymp00004-0109.pdf.

Oliveira, A. G. and Salgado, N. C. (2006). Design aspects of a distributed clinical trials information system. *Clinical Trials*, *3*(4), pp.385-396. Doi: 10.1177/1740774506069156.

Park, Y. R., Yoon, Y. J., Koo, H. Y., Yoo, S., Choi, C. C., Beck, S. and Kim, T. W. (2018). Utilization of a clinical trial management system for the whole clinical trial process as an integrated database: a system development. *Journal of Medical Internet Research,* 20(4). Doi: 10.2196/jmir.9312: 10.2196/jmir.9312.

Patrão, D. F., Oleynik, M., Massicano, F. and Sasso, A. M. (2015). Recruit-An Ontology Based Information Retrieval System for Clinical Trials Recruitment. In *MedInfo*, pp. 534-538. Doi: 10.3233/978-1-61499-564-7-534.

Richesson, R. L. and. Andrews, J. E. (2019). Clinical Research Informatics. Cham: Springer Nature Switzerland AG, pp. 173.

Richesson, R. L., Malloy, J. F., Paulus, K., Cuthbertson, D. and Krischer, J. P. (2008). An automated standardized system for managing adverse events in clinical research networks. *Drug safety*, *31*(10), pp. 807-822. Doi: https://doi.org/10.2165/00002018-200831100-00001.

Riva, J. J., Malik, K. M., Burnie, S. J., Endicott, A. R., and Busse, J. W. (2012). What is your research question? An introduction to the PICOT format for clinicians. *The Journal of the Canadian Chiropractic Association*, *56*(3), p.167-171. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3430448/.

Sahay, R., Akhtar, W. and Fox., R. (2008). PPEPR: plug and play electronic patient records. In: *Symposium on Applied Computing (SAC).* [online] Fortaleza: ACM, pp.2298-2304. Doi: 10.1145/1363686.1364232

Sahay, R., Btalaperas, D., Kamateri, E., Hasapis, P., Beyan, O. D., Strippoli, M. P. F., … and Bouras, T. (2013). An ontology for clinical data integration. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics.* Manchester, pp.3244-3250. Doi: 10.1109/SMC.2013.553.

Sahay, R., Fox, R., Zimmermann, A., Polleres, A. and Hauswrith, M. (2011). A methodological approach for ontologising and aligning health level seven (HL7) applications. In: *International Conference on Availability, Reliability, and Security* for Business, Enterprise and Health Information Systems. Vienna: Springer, pp. 102-117.

Salgado, N. C. and Oliveira, A. G. (2000). Towards a common framework for clinical trials information systems. *Proceedings. AMIA Symposium*, pp.754–758. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2244087/pdf/procamiasymp00003-0789.pdf.

Schreiweis, B., Trinczek, B., Köpcke, F., Leusch, T., Majeed, R. W., Wenk, J., Bergh, B., Ohmann, C., *Röhrig, R., Dugas, M. and* Prokosch, H. U. (2014). Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. *International journal of medical informatics*, *83*(11), pp. 860-868. Available at: https://doi.org/10.1016/j.ijmedinf.2014.08.005.

Shankar, R. D., Martins, S. B., O'Connor, M. J., Parrish, D. B. and Das, A. K. (2006). Epoch: an ontological framework to support clinical trials management. In: *Proceedings of the international workshop on Healthcare information and knowledge management*, Virginia:ACM, pp. 25-32. Doi:10.1145/1183568.1183574.

Shankar, R. D., Martins, S. B., O'Connor, M., Parrish, D. B., & Das, A. K. (2007). An ontology-based architecture for integration of clinical trials management applications. In: *Annual Symposium Proceedings*, Chicago: AMIA, pp. 661-665. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2655871/.

Sim, I., Olasov, B. and Carini, S. (2004). An ontology of randomized controlled trials for evidence-based practice: content specification and evaluation using the competency decomposition method. *Journal of Biomedical Informatics*, *37*(2), pp. 108-119. Doi: 10.1016/j.jbi.2004.03.001.

Sim, I., Tu, S. W., Carini, S., Lehmann, H. P., Pollock, B. H., Peleg, M. and Wittkowski, K. M. (2014). The Ontology of Clinical Research (OCRe): an informatics foundation for the science of clinical research. *Journal of biomedical informatics*, *52*, pp. 78-91. Available at: https://doi.org/10.1016/j.jbi.2013.11.002.

Smith, J. B., Lacey, S. R., Williams, A. R., Teasley, S. L., Olney, A., Hunt, C., Cox K. C. and Kemper, C. (2011). Developing and testing a clinical information system evaluation tool: prioritizing modifications through end-user input. *Journal of Nursing Administration*, *41*(6), pp.252-258. Doi: 10.1097/NNA.0b013e31821c4634.

Smith, B. and Scheuermann, R. H. (2011). Ontologies for clinical and translational research: Introduction. *J Biomed Inform.*, 44(1), pp. 3-7. Doi: 10.1016/j.jbi.2011.01.002.

Stenzhorn, H., Weiler, G., Brochhausen, M., Schera, F., Kritsotakis, V., Tsiknakis, M., Kiefer, S. and Graf, N. M. (2010). The ObTiMA system-ontology-based managing of clinical trials. *Studies in health technology and informatics*, 160(Pt2), pp.1090-1094. Doi: 10.3233/978-1-60750-588-4-1090.

Thabane, L., Thomas, T., Ye, C. and Paul, J. (2008). Posing the research question: not so simple. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, 56(1), pp.71-79. Doi: https://doi.org/10.1007/s12630-008-9007-4.

Thiese, M. S. (2014). Observational and interventional study design types: an overview. *Biochemia medica*. *24*(2), pp. 199-210. Doi: https://doi.org/10.11613/BM.2014.022.

Tran, V., Johnson, N., Redline, S. and Zhang, G. (2011). OnWARD: Ontology-driven web-based framework for multi-center clinical studies. *Journal of Biomedical Informatics,* 44, pp.S48-S53. Doi: 10.1016/j.jbi.2011.08.019.

Vetter, T. R. and Mascha, E. J. (2017). Defining the primary outcomes and justifying secondary outcomes of a study: usually, the fewer, the better. *Anesthesia & Analgesia,* 125(2), pp.678-681. Doi: 10.1213/ANE.0000000000002224.

Wales.nhs.uk (2017). *NHS Wales data dictionary.* [online] Available at: http://www.datadictionary.wales.nhs.uk [Accessed 30 Jul. 2019].

Wang, X., Hripcsak, G., Markatou, M. and Friedman, C. (2009). Active computerized pharmacovigilance using natural language processing, statistics, and electronic health records: a feasibility study. *Journal of the American Medical Informatics Association*, *16*(3), pp.328-337. Available at: https://doi.org/10.1197/jamia.M3028.

Wang, J. J. and Yang, D. L. (2007). Using a hybrid multi-criteria decision and aid method for information systems outsourcing. Computers &Operations Research, 34(12), pp. 3691-3700. Doi: 10.1016/j.cor.2006.01.017.

Weisskopf, M., Bucklar, G. and Blaser, J. (2014). Tools in a clinical information system supporting clinical trials at a Swiss University Hospital. *Clinical Trials*, 11(6), pp. 673-680. Doi: 10.1177/1740774514546702.

World Health Organization (1984). *Adverse Reaction Terminology.* [online] Available at: https://apps.who.int/iris/handle/10665/61950 [Accessed 30 Jul. 2019].

World Health Organization. (2002). *The importance of pharmacovigilance*. Available at: https://apps.who.int/iris/bitstream/handle/10665/42493/a75646.pdf.

World Health Organization (2015). *WHO Statement on Public Disclosure of Clinical Trial Results.* [online] Available at: https://www.who.int/ictrp/results/reporting/en/ [Accessed 28 Jul. 2019].

World Health Organization (2018). *International Classification of Diseases (ICD-11)*. [online] Available at: https://www.who.int/classifications/icd/en/ [Accessed 30 Jul. 2019].

World Medical Association. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, *79*(4),pp.373-374. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2566407/.

Wu, H., Toti, G., Morley, K. I., Ibrahim, Z. M., Folarin, A., Jacjson, R., Kartoglu, I., Agrawal, A., Stringer, C., Gale, D., Gorrell, G., Roberts, A., Broadbent, M., Stewart, R. and Dobson, R. J. (2018). SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. *Journal of the American Medical Informatics Association,* 25(5), pp.530-537. Doi: 10.1093/jamia/ocx160.

Zapletal, E., Rodon, N., Grabar, N. and Degoulet, P. (2010). Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. *MedInfo*, pp. 193-197. Doi: 10.3233/978-1-60750-588-4-193.