Clinical Research Informatics

26

Philip R.O. Payne, Peter J. Embi, and James J. Cimino

After reading this chapter, you should know the answers to these questions:

- What is clinical research and what factors influence the design of clinical studies?
- What are the types of information needs inherent to clinical research and how can those information needs be stratified by research project phase or activity?
- What types of information systems can be used to address or satisfy the information needs of clinical research teams?
- How can multi-purpose platforms, such as electronic health record (EHR) systems (see Chap. 12), be leveraged to enable clinical research programs?
- What is the role of a clinical trial management system (CTMS) for supporting and enabling clinical research, and what types of functionality are common to such systems?
- What is the role of standards in supporting interoperability across and between actors and entities involved in clinical research activities?
- What are current and future CRI research "grand challenges" and how will they optimize or otherwise alter the conduct of clinical research?
- How does clinical research informatics relate to the field of Biomedical Informatics and the broader clinical and translational science continuum?

P.R.O. Payne, PhD, FACMI (✉)
Department of Biomedical Informatics,
The Ohio State University Wexner Medical Center,
3190 Graves Hall, 333 West 10th Avenue, Columbus,
OH, 43210, USA
e-mail: philip.payne@osumc.edu

P.J. Embi, MD, MS
Departments of Biomedical Informatics and Internal Medicine, The Ohio State University Wexner Medical Center, 3190 Graves Hall, 333 West 10th Avenue, Columbus, OH, 43210, USA
e-mail: peter.embi@osumc.edu

J.J. Cimino, MD
Laboratory for Informatics Development, NIH Clinical Center, 10 Center Drive, Room 6-2551, Bethesda, MA, 20892, USA
e-mail: ciminoj@mail.nih.gov

26.1 Introduction

The conduct of clinical research is fundamental to the generation of evidence that can in turn facilitate improvements in human health. However, the design, execution, and analysis of clinical research is an inherently complex information- and resource-intensive endeavor, involving a broad variety of stakeholders, workflows, processes, data types, and computational resources. At the intersection point between biomedical informatics and clinical research, a robust and growing sub-discipline of informatics has emerged, which for the remainder of this chapter we will refer to as **clinical research informatics** (**CRI**) (Embi and Payne 2009; Payne et al. 2005). Numerous reports have shown that innovations and best practices generated by

E.H. Shortliffe, J.J. Cimino (eds.), *Biomedical Informatics*,
DOI 10.1007/978-1-4471-4474-8_26, © Springer-Verlag London 2014

the CRI community have contributed to improvements in the quality, efficiency, and expediency of clinical research (Chung et al. 2006; Payne et al. 2005; Sung et al. 2003). Such benefits can be situated in a full spectrum of contexts that extends from the activities of individual clinical investigators to the operations of multi-center research consortia that involve geographically and temporally distributed participants.

Given the recognition of CRI as a distinct and increasingly important sub-discipline of biomedical informatics, it is imperative that a common basis for defining and understanding CRI science and practice be established. Such a foundation must by necessity include explicit linkages to the major challenges and opportunities associated with the planning, conduct, and evaluation of clinical research programs. To provide a common frame of reference for the remainder of this chapter, we will use the National Institutes of Health (NIH) definition of clinical research :[1]

> ***Clinical Research*** *involves*, "*the range of studies and trials in human subjects that fall into the three sub-categories*: (1) ***Patient-oriented research***: *Research conducted with human subjects* (*or on material of human origin such as tissues*, *specimens and cognitive phenomena*) *for which an investigator* (*or colleague*) *directly interacts with human subjects*. *Patient-oriented research includes*: (*a*) *mechanisms of human disease*; (*b*) *therapeutic interventions*; (*c*) *clinical trial*; *and* (*d*) *development of new technologies*. (2) ***Epidemiologic and behavioral studies***. (3) ***Outcomes research and health services research***."

A lack of sufficient information technology (IT) and applied biomedical informatics tools, expertise and methods, as well as a reliance on workflows largely defined by historical precedent rather than optimal operational strategies, account for significant impediments to the rapid, effective, and resource-efficient conduct of clinical research projects (Payne et al. 2005). Compounding these challenges is the rapid pace of advancement in biomedical science and the resulting need for advances in diagnostics and therapeutics that can be validated and disseminated quickly and cost effectively (Butte 2008a, b; Embi and Payne 2009; Payne et al. 2005, 2009). The confluence of these factors has led to a number of major challenges and opportunities related to current and future CRI research and practice. For example, the importance of making clinical phenotype data available for the secondary use in support of clinical research has become a competitive requirement for research enterprises of all sizes (Chung et al. 2006; Embi and Payne 2009). Similarly, the increasing complexity of clinical research programs and the difficulty of recruiting sufficiently large patient cohorts, when combined with the regulatory overhead of conducting studies in large academic institutions, has led to an increase in the conduct of clinical studies in community practice settings. Such community-based research paradigms introduce new levels of complexity to the technical and policy aspects of data capture, management, and sharing plans (Embi and Payne 2009). This rapid evolution and the realities of an increasingly expansive clinical research landscape has led investigators and other decision makers in the health care and life sciences communities to call for increased investments in and delivery of innovative solutions to such information needs (Ash et al. 2008; Chung et al. 2006; Embi and Payne 2009; Payne et al. 2005; Sung et al. 2003). At the highest level, clinical research is a domain faced with significant information management challenges. At the same time, clinical research is an area of scientific endeavor that is at the forefront of attention for the governmental, academic, and private sectors, all of whom have significant scientific and financial interests in the conduct and outcomes of such efforts. These challenges and opportunities, when viewed collectively, have called, and continue to call, for

[1] NIH. (2011). Glossary & Acronym List Retrieved June 20, 2011, from http://grants.nih.gov/grants/glossary.htm (Accessed December 12, 2012)

the development and validation of innovative biomedical informatics methods and tools specifically designed to address clinical research information needs. It is this overall context that has motivated an increasing focus on the both basic and applied Clinical Research Informatics (CRI), which can be defined broadly as follows (Embi and Payne 2009):

Clinical Research Informatics (CRI) is the sub-domain of biomedical informatics concerned with the development, evaluation and application of informatics theory, methods and systems to improve the design and conduct of clinical research and to disseminate the knowledge gained.

Examples of focus areas in which CRI researchers and practitioners apply biomedical informatics theories and methods can include the following:

- Evaluation and modeling of clinical research workflow
- Social and behavioral studies involving clinical research professionals and participants
- Designing optimal human-computer interaction models for clinical research applications
- Improving information capture and data flow in clinical research
- Leveraging data collected in EHRs
- Optimizing site selection, investigator and patient recruitment
- Improving reporting to regulatory agencies
- Enhancing clinical and research data mining, integration, and analysis
- Phenomic characterization of patients for cohort discovery and analytical purposes
- Integrating research findings into individual and population level health care
- Defining and promoting ethical standards in CRI practice
- Educating researchers, informaticians, and organizational leaders about CRI
- Driving public policy around clinical and translational research informatics

Building upon the preceding definitions and overarching challenges and opportunities relevant to CRI, in the remainder of this chapter we will provide an overview of the types of activities commonly undertaken as part of a variety of representative clinical research use cases, introduce the role of major classes and types of information system that enable or facilitate such activities, and conclude with a set analyses regarding the future directions of the field. The overall objective of this chapter is to provide the reader with the ability to evaluate critically the current and anticipated roles of biomedical informatics knowledge and practice as applied to clinical research.

26.2 A Primer on Clinical Research

In the following section, we will briefly introduce the characteristics of the modern clinical research environment (Sect. 26.2.1), including the design and execution of an exemplary class of clinical studies that were introduced in Chap. 11 and are known as randomized controlled trials (Sect. 26.2.2). This primer on clinical research will serve as the context for the remainder of the chapter, in which we will introduce major information needs and their relationships to a variety of basic and applied biomedical informatics practice areas and IT applications.

26.2.1 The Modern Clinical Research Environment

Clinical research comes in many forms and may include a variety of specific activities. All forms,

however, share a common set of requirements related to the comprehensive management of study data – specifically, collection of data on human research subjects – and analysis of those data. As clinical research designs traverse the spectrum from passive or observational studies to interventional trials, the acuity of activities and associated data-management needs increases commensurately. For example, as introduced in Chap. 11, in a retrospective study subjects are selected based on the presence or absence of a particular condition and retrospective or pre-existing data are obtained from historical records (such as EHRs), whereas in natural history studies, subjects are recruited and followed in prospective manner, with additional collection of data performed solely for the purposes of research, rather than the normal process of patient care.

Further along the spectrum are clinical trials, in which research subjects participate in some additional activity, or *intervention*, that is intended either to induce a change in the subject or to prevent the occurrence of some change that would otherwise be expected. The intervention might be as simple as administering a substance already found in the human body (such as a vitamin) to measuring a change in that substance (such as a the amount of the vitamin found in the blood or urine). More complex studies involve interventions that have an impact on human disease, such as the administration of a preventive vaccine, the administration a curative drug, or a surgical procedure to remove, insert, repair or replace a structure or device in the subject's body. As with passive studies, data collection is critical to the proper performance of research and may become intense, with the collection of clinical information occurring more frequently and involving data describing the intervention materials (such as the purity of a drug or the performance of a device) in addition to data related to the human subject and their response to the intervention under study.

Although not an intrinsic requirement of clinical research, the inclusion of comparison groups is generally considered an important part of good scientific method. In some cases, **historical controls** can be used for comparison with a group of subjects under study. For example, if a disease is known to have a particular fatality rate, subjects could be given a potentially life-saving treatment and their fatality rate can be measured and compared to past experience. In **quasi-experiments**, comparison subject groups can also be selected for based on some known characteristic that distinguishes the two groups, such as gender or race, or their willingness to undergo a particular intervention.

A more rigorous method of establishing comparison groups is through randomization (Chap. 11), in which prospective subjects are assigned to different groups (often referred to as **study arms**) and undergo different interventions. Typically, randomization might take into account observable characteristics (such as gender and race) to create balanced groups, especially where the characteristics are known to have some influence on the effect of the intended intervention. Randomization also serves to distribute subjects based on unobserved characteristics, for example, unknown genetic traits, in order to reduce differences in the groups that might bias the results of the study. In a randomized controlled trial (RCT), one subject group will often receive a **control intervention** (for example, the usual treatment for a condition or even no treatment) while one or more other groups receive an **experimental intervention**.

Although intended to reduce bias, the randomization process itself must be carefully executed such that it does not introduce new sources of bias. For example, randomization can include **blinding**, in which the subject, the investigator, or both (as in **double-blinded** studies), are kept unaware of group assignment until after all assessments have been made. This might include the use of a **placebo** for a group receiving no treatment, in order to avoid the possibility that subjective improvement in a prior condition or the occurrence of random events (such as normally occurring illnesses) or are not ascribed to the intervention. This also may prevent subjects from deciding not to participate after randomization in a way that might unbalance the study groups (for example, if subjects prefer not to

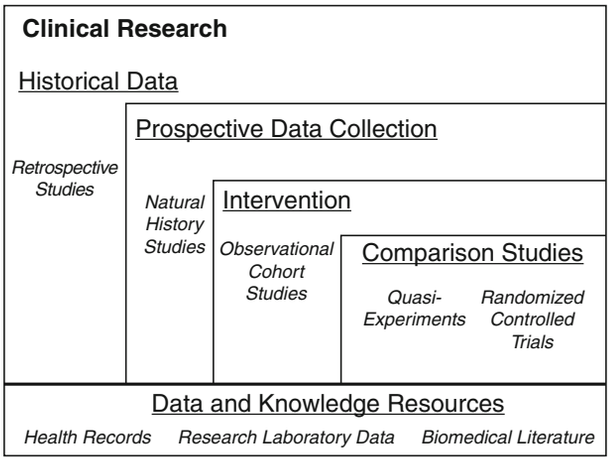

Fig. 26.1 Overview of clinical study designs and associated information and data management needs. Underlying such design patterns are a common thread of systematic data management, leveraging resources such as health records, research-specific laboratory data, as well as broader knowledge collections such as the published biomedical literature

participate if they know they are not getting the experimental intervention) or even bias the assignments (for example, people less prone to take care of themselves might drop out if they find they are assigned to an intervention that requires a great deal of effort on their part).

The gold standard of clinical studies (introduced in Chap. 2) is generally considered to be the double-blinded, randomized, placebo-controlled trial (Cimino et al. 2000). However, such studies may not always be practical. For example, the use of a placebo when an effective therapy is known may be unethical, the blinding of a surgical repair may not be practical, or the condition under study may be so rare that only historical controls are available.

While different study designs have unique and differentiated information needs, they uniformly involve some form of systematic data management, as noted previously. Such data management activities usually include initial data collection, aggregation, analysis, and results dissemination, to name a few of many such tasks. As shown in Fig. 26.1, different study methods introduce new issues as successively more complex interventions and study design patterns are employed. For the remainder of this chapter, we will focus our discussion on RCTs as our prototypical study design, since they tend to involve most if not all of the informatics issues and information needs encountered in other study designs. Further information on the design characteristics, data management needs, and associated best practices related to various types of clinical trials can be found in a number of excellent textbooks on the subject (Gallin and Ognibene 2012), and further discussion is beyond the scope of this chapter.

26.2.2 Phased Randomized Controlled Trials

Most clinical studies begin with the identification of a set of driving or motivating hypotheses. The research questions that serve to define such hypotheses might be raised through an analysis of gaps in knowledge as found in the published biomedical literature or be informed by the results of a previous study. It is important to note that clinical research endeavors exist on a spectrum of scientific activity that is commonly referred to as **clinical and translational research**. A particular type of translational research, often referred to as T1-type translation (see Chap. 25), is a process by which basic science discoveries are used to design novel therapies. Such discoveries are then evaluated during clinical research studies, first pre-clinical and subsequent clinical trial phases (Payne et al. 2005). A second type of translational research, often referred to as T2 translation, involves methods such as those borrowed from **implementation science** and clinical informatics, and focus on translating the findings of such clinical research studies into common practice. A common colloquialism for this process of translating a novel basic science discovery through clinical research and into clinical practice is "bench to bedside" science.

Individual and distinct RCTs are often conducted for different purposes, most often motivated by the need to fill fundamental knowledge gaps about a particular intervention under study. By combining such knowledge gaps with the underlying biomedical mechanisms of physiology

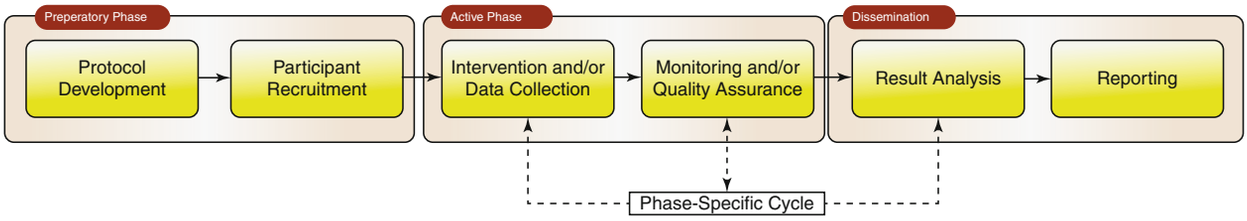

Fig. 26.2 Overview of the clinical research process for Phase I-III trials, divided into three major phases (preparatory, active, dissemination)

and disease, a motivating hypothesis or collections of hypotheses are established as to why a given intervention might lead to a given result or finding. Such hypotheses result in a natural sequence of research questions that can be asked relative to a novel intervention. Usually, an individual research study is designed to address one specific research question and hypothesis. In the case of the development and evaluation of a new therapeutic intervention, like a new drug, an individual research study is designed to address each **phase** in a line of research inquiry that will determine the efficacy and effectiveness of such a therapy (Spilker 1991). In most cases, this adheres to the following model:

- **Phase I**: Investigators evaluate the novel therapy in a small group of participants in order to assess overall safety. This safety assessment includes dosing levels in the case of non-interventional therapeutic trials, and potential side effects or adverse effects of the therapy. Often, Phase I trials of non-interventional therapies involve the use of normal volunteers who do not have the disease state targeted by the novel therapy.
- **Phase II**: Investigators evaluate the novel therapy in a larger group of participants in order to assess the efficacy of the treatment in the targeted disease state. During this phase, assessment of overall safety is continued.
- **Phase III**: Investigators evaluate the novel therapy in an even larger group of participants and compare its performance to a reference standard which is usually the current standard of care for the targeted disease state. This phase typically employs an RCT design, and often a multi-center RCT given the numbers of variation of subjects that must be recruited to adequately test the hypothesis. In general, this is the final study phase to be performed before seeking regulatory approval for the novel therapy and broader use in standard-of-care environments.
- **Phase IV**: Investigators study the performance and safety of the novel therapy after it has been approved and marketed. This type of study is performed in order to detect long-term outcomes and effects of the therapy. It is often called "post-market surveillance" and is, in fact, not an RCT at all, but a less formal, observational study.

The phase of an RCT has implications for the kinds of questions being asked and the kinds of processes carried out to answer them. From an informatics perspective, however, the tasks are usually very similar. At a high level, the conduct of a Phase I, II or III clinical trial can be thought of in an operational sense as consisting of three major stages: preparatory, active, and dissemination (Fig. 26.2).

During these three stages, a specific temporal series of processes is executed. First, during the **preparatory phase**, a protocol document is generated as part of the project development process. The protocol document usually contains background information, scientific goals, aims, hypotheses and research questions to be addressed by the trial. In addition, the protocol describes policies, procedures, and data collection or analysis requirements. A critical aspect of the protocol document is the definition of a protocol schema, which defines at a highly granular level the temporal sequence of tasks and events required to both deliver the intervention under study and to ensure that data are collected and managed in a systematic manner commensurate

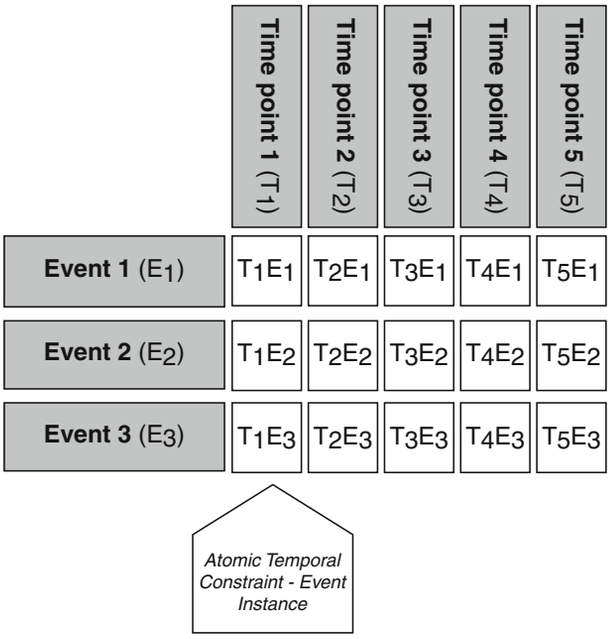

Fig. 26.3 Generic layout of a clinical trial protocol schema, composed of atomic temporal constraints. Event instances are shown as Time Point (T) – Event (E), using the notation: T_xE_y, where x is the Time Point descriptor, and y is the Event descriptor. In some instances, a transposed version of this grid is used

with the study hypotheses and aims. Such protocol schemata are often represented as a temporal grid (Fig. 26.3).

Once a protocol is deemed ready for execution, the feasibility of the study design (e.g., addressing questions such as "are there enough participants available in the targeted population to satisfy the study design defined in the protocol document?") is assessed either quantitatively (e.g., using historical data) and/or heuristically. Throughout the preparatory phase, a concurrent process of seeking regulatory approval from local and national bodies (e.g., local Institutional Review Boards ((Bernstam et al.), the Food and Drug Administration (FDA), etc.) occurs. Once a protocol plan is complete, deemed feasible, and regulatory approval has been received, potential participants are recruited and screened to determine if they meet the inclusion and exclusion criteria for the study (e.g., specific demographic and/or clinical parameters required for subjects to be eligible for the study). Once a potential participant has been deemed eligible for the study, they are provided with an informed consent document, which must be signed prior to proceeding with the enrollment process. **Enrollment** in the context of clinical trials means officially registering as a study subject, and is normally associated with the assignment of a study-specific identifier. Once a person agrees to become a participant, they are enrolled, and in the case of studies with multiple study groups or arms, randomized into one of those arms.

The preceding activities lead to the initiation of the next step in the research process, which we refer to as the **active phase**. During the active phase, the participant receives the therapeutic intervention indicated by their study arm and is actively monitored to enable the collection of study-specific data. This therapeutic intervention and active monitoring process is often iterative, involving multiple cycles of interventions and active monitoring. Follow-up activities begin once a participant has completed the interventional stage of a study. During this stage, subjects are contacted on a specified temporal basis in order to collect additional data of interest, such as long-term treatment effects, disease status or survival status (Spilker 1991).

Finally, during the **dissemination phase**, the results of the study are evaluated and formalized in publications or other knowledge dissemination media, for translation into the next phase of an RCT or into clinical practice. In some cases, such as adaptive study designs, this dissemination phase feeds back into the planning and active phases to allow for rapid revisions to a study design and iterative participant enrollment and data collection in support of such revised hypotheses and designs.

The quality of data produced by a clinical trial is assessed using multi-dimensional metrics that take into account the design, execution, analysis and dissemination of the study results. The quality of a clinical trial is also judged with respect to the significance or relevance of the reported study results within a clinical context (Juni et al. 2001). One key metric used to assess clinical trial quality is validity, which can be defined both internally and externally. **Internal validity** is defined as the minimization of potential biases during the design and execution of the trial, while **external validity** is the ability to generalize study results into clinical care (Juni et al. 2001).

It is important to note in a discussion of the role of biomedical informatics relative to clinical research that a large number of both basic and applied informatics practice areas concerning this domain focus upon platforms, interventions, and methods intended to reduce or mitigate such sources of bias, thus enhancing the validity and generalizability of study results.

26.3 Information Needs and Systems in the Clinical Research Environment

As can be inferred by the preceding section and its introduction to the definitional aspects of clinical research, such activities regularly involve a variety of data, information, and knowledge sources, as well a complicated set of complementary and overlapping workflows. At the highest level, these characteristics of the clinical research environment can be related to a number of critical information needs, as summarized in Table 26.1. This representation of the information needs inherent to clinical research is presented using the specific context of a prototypical RCT, but the basic types of needs and example solutions provided can be extended to apply to the broader spectrum of research designs and patterns introduced in Sect. 26.2.

Building upon this broad definition of the information needs inherent to clinical research, in the following sub-sections we: (1) review the types of information systems that can support the phases that comprise a clinical study (Sect. 26.3.1); (2) explore the functional components that make up a clinical trials management system (Sect. 26.3.2); and (3) discuss the role of standards in enabling interoperability between such information systems (Sect. 26.3.3).

26.3.1 Information Systems Supporting Clinical Research Programs

It is helpful to conceptualize the conduct of clinical trials as a multiple-stage sequential model, as was introduced in Sect. 26.2.2 and is expanded upon in this section (Payne et al. 2005) (Fig. 26.4). At each stage in such a model, a combination of research-specific and general technologies can be employed to support or address related information needs.

There are numerous examples of general-purpose and clinical systems that are able to support the conduct of clinical research:

- **Bibliographic databases** and **information retrieval tools** such as PubMed and OVID (see Chap. 21) can be used to assist in conducting the background research necessary for the preparation of protocol documents (Briggs 2002; Ebbert et al. 2003; Eveillard 2000; Eysenbach et al. 2001; Eysenbach and Wyatt 2002).
- **Electronic health records** (EHRs, see Chap. 12) can be used to collect clinical data on research participants in a structured form that can reduce redundant data entry (Bates et al. 2003; Clark et al. 2001; Marks et al. 2001; McDonald 1997; McDonald et al. 1999; Padkin et al. 2001).
- **Data warehouses and associated data or text mining tools** can be used in multiple capacities, including: (1) determining if participant cohorts who meet the study inclusion or exclusion criteria can be practically recruited given historical trends, and (2) identifying specific participants and related data within existing databases (Butler 2001; Evans 2002; Marks and Power 2002).
- **Clinical decision-support systems** (CDSS, see Chap. 22) can be used to alert providers at the point-of-care that an individual may be eligible for a clinical trial (Bates et al. 1998; Butte et al. 2000; Embi et al. 2005; Marks and Power 2002).

In addition to the preceding general technologies, a number of research-specific technologies have been developed:

- **Feasibility analysis applications and data simulation and visualization tools** can streamline the pre-clinical research process (e.g., disease models) and assist in the analysis of complex data sets in order to assess the feasibility of a given study design (Holford et al. 2000; Kim et al. 2002).

Table 26.1 Overview of definitional information needs in the clinical research environment

Information needs	Major sub-components	Description
Support for research planning and conduct	Collaborative document and knowledge management Data sources and tools for feasibility analyses Regulatory approval workflows	*Study teams often involve geographically and temporally distributed participants, who need to engage in iterative protocol development and approval processes. Such activities by necessity incorporate document versioning, annotation, and associated metadata management tasks. Once a protocol has been developed, access to data sets for the purposes of assessing the feasibility of a given study design is critical, and often involves the use of de-identified data sets drawn from a data warehouse or research registry. Finally, the submission, tracking, and documentation of regulatory approvals often necessitate the coordination and management of complex, document-oriented workflows and record keeping tasks.*
Facilitation of data management, access, and integration	Secondary-use of EHR-derived data for research purposes Research project specific data capture, management, and reporting Distributed data management (spanning traditional organizational boundaries) Syntactic and semantic interoperability	*The ability to use primary clinical data from EHR or equivalent platforms to support secondary use in a research program has the potential to reduce redundancy and potential errors while increasing data quality. However, using such data in a secondary capacity also requires that appropriately structured data be captured and codified in clinical systems, and then be made available to research teams and research data management systems in a timely and resource efficient manner. In addition to such secondary use of clinical data, most clinical studies require the regular capture and management of study-specific data elements, a task that is usually accomplished via the use of Electronic Data Capture (EDC) or Clinical Trial Management Systems (CTMS). Finally, given the propensity to conduct studies that span traditional organizational boundaries in order to realize economies of scale and/or access sufficiently large patient populations, it is often necessary to query, integrate, and manage distributed data sets, and ensure their syntactic and semantic interoperability. Such a need is usually addressed through the use of Service Oriented Architectures, Cloud Computing, Data Warehousing, and Metadata Management technologies.*
Workforce training and support	Dissemination of study, methodological, and technical training materials Support for team collaboration and knowledge sharing	*A central need when conducting clinical studies is the ability to ensure that individuals involved in the execution of a protocol share common methods, data management practices, and workflows (thus reducing potential sources of study bias). Ensuring such shared knowledge and practices, particularly in distributed or multi-site settings, requires the use of distance education and team-science tools and platforms to enable knowledge sharing and distance learning paradigms.*
Management information capture and reporting	Support for research billing Operational instrumentation and reporting Regulatory monitoring Data quality assurance	*The business and management aspects of the conduct of clinical studies is complex, often requiring the disambiguation of standard-of-care and research specific charges as part of billing operations, as well as the tracking of key performance and data quality metrics that may be required to satisfy contractual commitments to the entities funding such studies. Furthermore, the monitoring of study data for critical or sentinel events that should or must be reported for regulatory purposes is both necessary and of extreme importance. All of the aforementioned activities require the application of a variety of management information system, business intelligence, and reporting tools, leveraging a broad variety of enterprise, administrative, and study-specific data sources.*

(Continued)

Table 26.1 (continued)

Information needs	Major sub-components	Description
Participant recruitment tools and methods	Cohort discovery Eligibility determination and alerting Participant registration, consent, and enrollment execution and tracking	*The identification of participant cohorts that satisfy key study design criteria, such as inclusion and exclusion criteria, is frequently a major barrier to the timely and efficient execution of clinical studies. A variety of information needs, related to the identification and engagement of such cohorts, to point-of-care alerting regarding potential study eligibility, to the management of registration, consent, and enrollment records is inherent to this information need. Such requirements are usually satisfied through a multi-modal approach, leveraging both clinical and research-specific information systems.*
Data standards	Standards for interoperability between research systems Standards for interoperability between research, enterprise (e.g. EHR), and administrative systems	*As has been noted relative to several of the preceding information needs, there is a frequent and reoccurring requirement for both syntactic and semantic interoperability between research-specific information systems, as well as between research-specific and clinical or administrative systems. Such a need necessitates the design, selection, and application of a variety of data standards, as well as the ability to map and harmonize between shared information models to support interactions between systems using a variety of standards.*
Workflow support	Integration of tools for combined standard-of-care and research visits Data, information, and knowledge transfer between stakeholders, project phases, activities, and associated information systems or tools.	*Much as was the case related to data standards, a closely aligned information need exists relative to the ability to support complex workflows between information systems and actors involved in the conduct of clinical research. Such workflow support requires both computational and application-level workflow orchestration, as well as the ability to define and apply reusable data analytic "pipelines."*
Data, information and knowledge dissemination	Knowledge management for clinical evidence generated during trials Guidelines and CDSS delivery mechanisms Publication mechanisms Data registries	*The ultimate objective of clinical research is to generate and apply new evidence in support of improvements in clinical care and human health. In order to do so, it is necessary to disseminate the findings generated during such studies in a variety of formats, including reusable/actionable knowledge resources, clinical guidelines, decision support rules, and/or publications and reports. In addition, increasing emphasis is being placed on the transparency and reproducibility of study designs, which is largely being accomplished through the creation of public registries via which study data sets can be shared and made available to the broader biomedical community.*

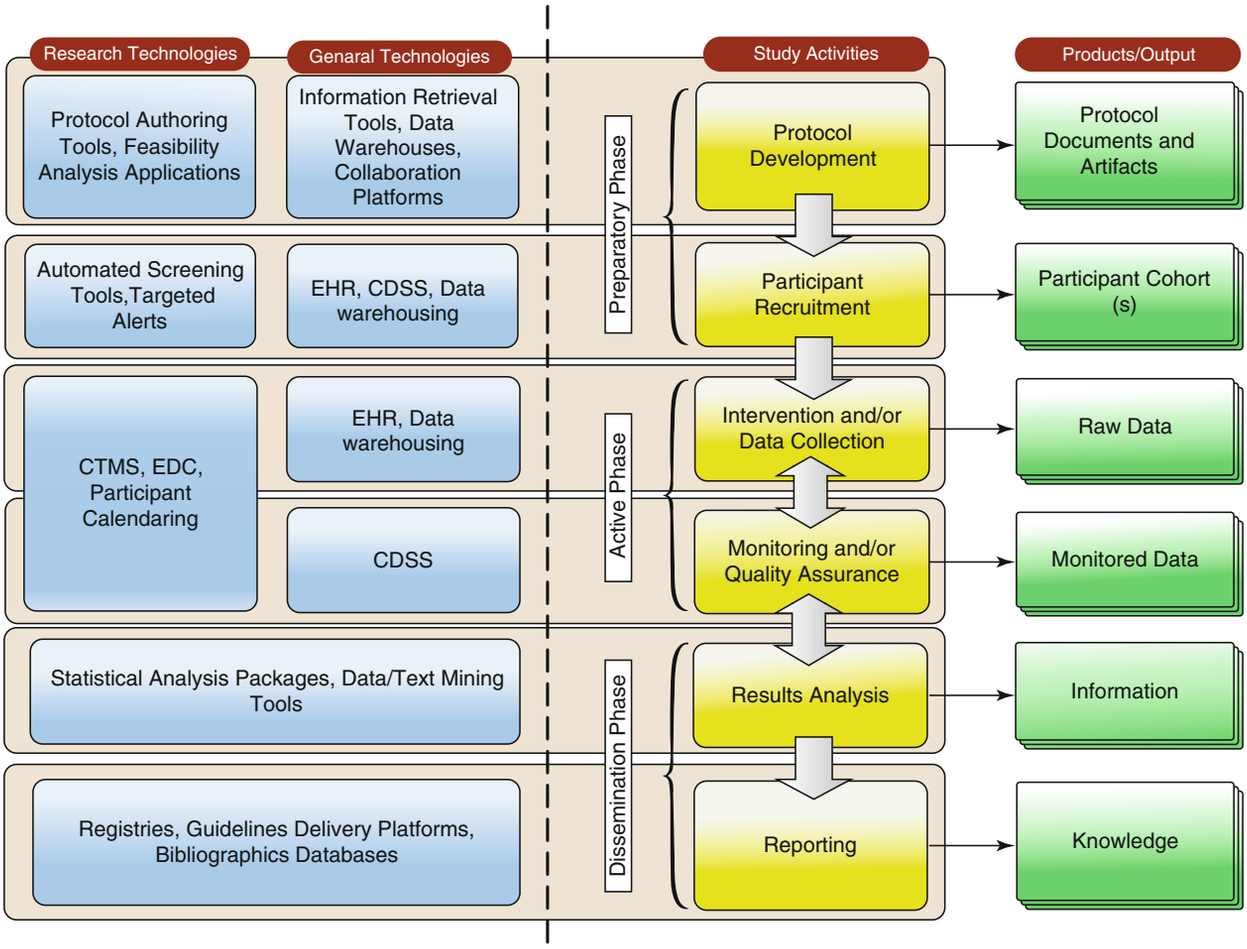

Fig. 26.4 Overview of study activities, and related research-specific and general information technologies, as well as targeted products or outputs associated with the sequential clinical research workflow paradigm

- **Protocol authoring tools** can allow geographically distributed authors to collaborate on complex protocol documents (Fazi et al. 2000, 2002; Goodman 2000; Rubin et al. 2000; Tai and Seldrup 2000).
- **Automated screening tools and targeted alerts** can assist in the identification and registration of research participants (Butte et al. 2000; Lutz and Henkind 2000; Marks and Power 2002; Pressler et al. 2012).
- **Electronic data capture (EDC) and Clinical Trial Management Systems (CTMS)** can be used to collect research-specific data in a structured form, and reduce the need for redundant and potentially error-prone paper-based data collection techniques (Harris et al. 2009; Kuchenbecker et al. 2001; Marks et al. 2001; Merzweiler et al. 2001; Wubbelt et al. 2000).
- **Research-specific decision support systems such as participant calendaring tools** provide protocol-specific guidelines and alerts to researchers, for example tracking the status of participants to ensure protocol compliance (Marks et al. 2001; Tai and Seldrup 2000).

26.3.2 Clinical Research Management Systems

One of the most widely used technology platforms in the clinical research domain is the **clinical research management system** (**CRMS**). Such platforms were historically referred to as clinical *trials* management systems (CTMS), but the term CRMS is gaining popularity as such systems are increasingly used to manage the conduct of studies including but not limited to trials.

CRMS platforms are usually architected as composite systems that incorporates a number of task and role-specific modules intended to address core research-related information needs (Chung et al. 2006; Payne et al. 2005, 2009). Exemplary instances of such modules include the following:

- **Protocol Management** components that support document management functionality to enable the submission, version control, and dissemination of protocol related artifacts and associated metadata annotations.
- **Participant Screening and Registration** tools that allow for the application of electronic eligibility "check lists" to individual patients or cohorts in order to assess study eligibility, and when appropriate, record the registration and associated "baseline" data that are required per the study protocol.
- **Participant Calendaring** functionality allows for the instantiation of general protocol schemas (e.g., a definition of a protocols temporal series of tasks, events, and associated data collection tasks) in a participant specific manner, accounting for complex reasoning tasks including the dynamic recalculation of temporal intervals between evens based on actual completion dates/times, as well as the "windowing" of events in which a given task or event is allowed to fall within a range of dates rather a specific, atomic temporal specification.
- **Electronic Data Capture** (**EDC**) components allow for the definition, instantiation, and use of electronic case report forms (e.g., forms that define study and task/event specific data elements to be collected in support of a given trial or research program). Such **electronic case report forms** (**eCRFs**) are the basic instrument by which the majority of study-specific data are collected, and are usually populated via a combination of: (1) manual data entry (including abstraction from source documentation such as medical records); (2) the importation of secondary use data from clinical systems; or (3) a hybrid of the two preceding approaches.
- **Monitoring tools** enable the application of logical rules and conditions (e.g., range-checking, enforcement of data completion, etc.) using a rules engine or equivalent technology, in order to ensure the completeness and quality of research related data. Such tools may also be used to monitor patient compliance with study schemas, as reflected in the previously described patient calendar functionality.
- **Query and Reporting Tools** support the planned and ad-hoc extraction and aggregation of data sets from multiple eCRFs or equivalent data capture instruments as used with the CTMS. These types of tools are commonly used by biostatisticians and other quantitative scientists to perform interim and final analyses of study results, outcomes, and to enable higher-order safety analyses. In addition, such tools may be employed to comply with a broad variety of data submission and reporting standard set by both public- and private-sector entities, as described in Sect. 3.3.
- **Security and Auditing** functionality enables site, role, and study-specific access controls and end-user authentication/authorization relative to all of the preceding functionality, as well as the ability to track and report upon end-user interaction with and modifications to data contained in the CRMS. Such functionality is critical to enabling compliance with a broad variety of regulatory and privacy/confidentiality frameworks that apply to the use of protected health information (PHI) for research purposes.

In most CRMS platforms, the aforementioned functional modules share one or more common research databases or in the case of service-oriented architectures (SOA), common data services (See Chap. 5 for more details on SOA technologies). In more advanced platforms, these common data structures are populated with research-specific and/or clinical data from enterprise systems and sources (such as electronic health records, personal health records, and data warehousing platforms) via either a SOA paradigm (e.g., data service publication and consumption) or an **extract**, **transform**, **and load** (**ETL**) **approach** (See Chaps. 2 and 6 for further

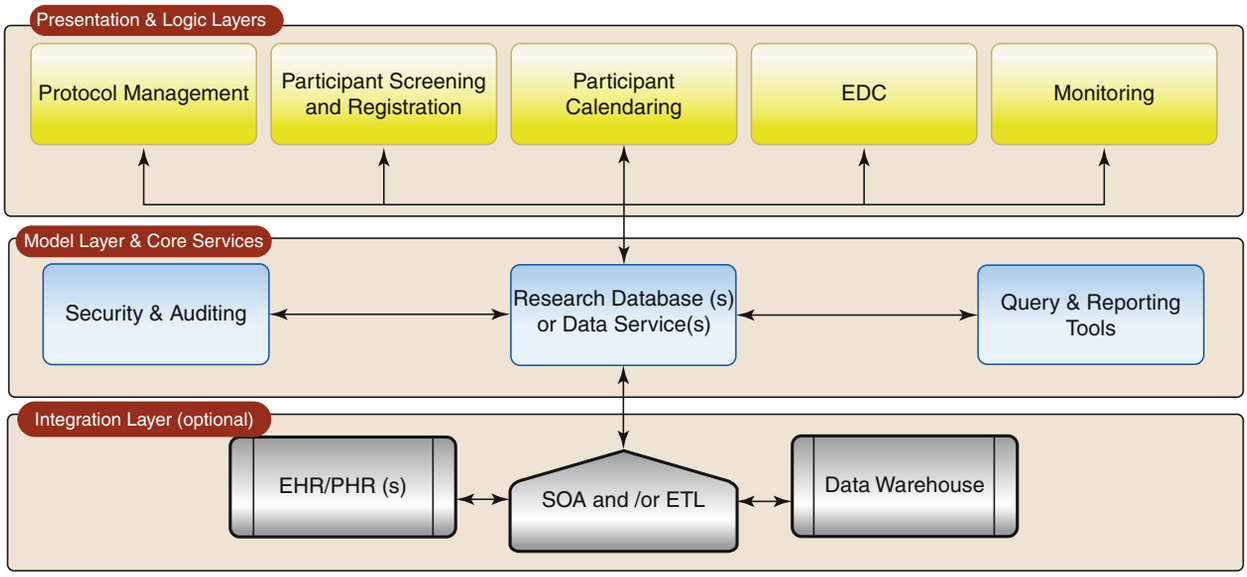

Fig. 26.5 Overview of the prototypical architecture of a clinical trial management system, divided into: (1) presentation and logic layers; (2) model layer and core services; and (3) an optional integration layer

details concerning architectural and methodological approaches to secondary use of clinical data). This overall architecture is illustrated in Fig. 26.5.

26.3.3 Data Standards in Clinical Research

The use of standards to represent clinical research information provides the same challenges and benefits found in other informatics application areas (see Chap. 7). Data may be captured with standard terminologies or translated into standards to support data reporting and sharing which, in turn, require agreed-upon standard frameworks to support such exchanges. Standards are even being developed for the representation of clinical trial protocols themselves. Figure 26.6 depicts how the various kinds of standards fit into the overall schema of clinical research, ranging from data models that define how data are to be represented, through standards for terminologies to actually represent the data and structures for exchanging them, out to standards for reporting and sharing. The standards described here are some of the current and most prevalent ones, but they continue to evolve and new standards relevant to the CRI domain are constantly emerging.

26.3.3.1 Emerging Standards and Domain Modeling in Clinical Research

Formats for data sharing typically include a data model for the information to be shared, leaving to individual contributors the later task of mapping local data into the exchange model. An alternative approach is the model-driven architecture, in which an underlying data model is created for the express purpose of representing all aspects of an information design, including data representation. Previously, the models used for clinical research management systems have been those required to support system functionality. New efforts are underway to create standards for modeling the actual research protocols, to enable a logical representation that includes the semantic aspects of the protocol (for example, the relationships between specific interventions and observations intended to measure their effects). While use of such models may make the research process somewhat more complicated, the mapping to standards used for exchanging data becomes greatly simplified.

For example, the National Cancer Institute (NCI) sponsors the *Cancer Biomedical Informatics Grid* (caBIG) program (Buetow 2009) which, among its many activities, has established a Clinical Trials Management

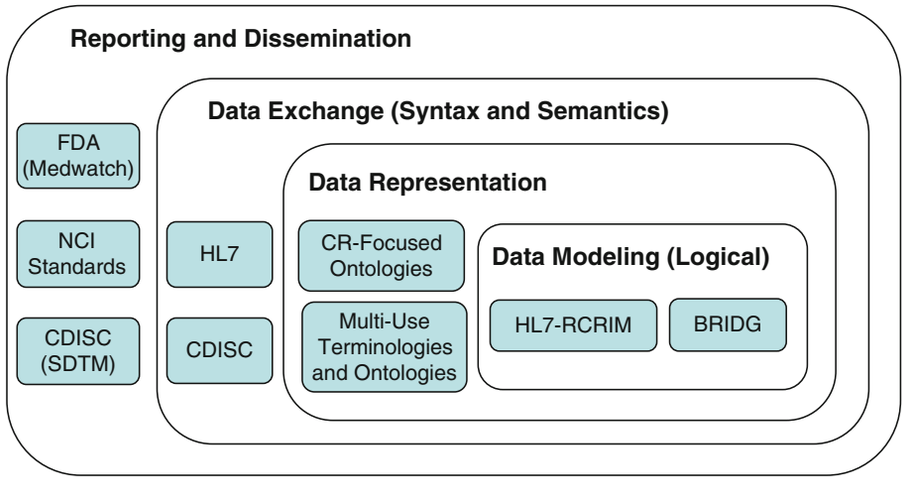

Fig. 26.6 Relationships of among various CRI standards. Data modeling, at the core, determines how terms from terminologies and ontologies will be recorded in clinical research databases. Exchange standards determine how data will map from the model to the messages used for interchanging the data. The use of messages is determined by the requirements of regulatory agencies and collaborating research groups. See text and Chap. 7 for explanation of acronyms

Systems Workspace (CTMS WS) that is developing standards to enable the design and execution of computable clinical trials. These efforts include the development of domain-specific workflow models and use cases to inform the design of CTMS's in a manner consistent with the "real-world" needs of clinical trials investigators, staff and sponsors. For example, a set of Life Science Business Architecture Models (Boyd et al. 2011) have been created to describe the vocabulary, goals and processes that are common in the business of life science research, including the actors, activities and data involved, using use cases described with the **Unified Modeling Language** (**UML**).

Similarly, Health-Level 7 (HL7; see Chap. 7) is an open standards development organization that develops consensus standards for all manner of clinical and administrative data, and is also working on clinical research-specific standards, such as the **Regulated Clinical Research Information Management** (**RCRIM**) model (Ohmann and Kuchinke 2009) in order to define messages, document structures, terminology and semantics related to the collection, storage, distribution, integration and analysis of research information. The main focus of the work is on data related to studies involving US Food and Drug Administration (FDA) regulated products (drugs and devices).

A key component of the previously described CTMS WS effort is the development of a data model known as the Biomedical Research Integrated Domain Group (BRIDG) Model.[2] BRIDG is designed to harmonize models from the HL7 RCRIM, the *Clinical Data Interchange Standards Consortium* (CDISC; see discussion of sharing and reuse later in this section) (Kuchinke et al. 2009), and models being developed by the CTMS WS itself. The modeling components of this project have focused primarily on logical abstractions of classes and data types, rather than domain-specific concepts, and are being put to practical use in a number of caBIG programs and resultant IT applications (Ohmann and Kuchinke 2009).

26.3.3.2 Using Standard Controlled Terminologies for Clinical Research Data

As described previously, the design of clinical protocols includes rigorous attention to the types of data to be collected and the format of those data. This often involves the use of controlled terminologies to capture categorical data. The terminology may be as small as "yes/no" or a

[2] hfttp://www.cdisc.org/bridg (Accessed December 12, 2012)

ten-point pain scale for capturing subjects' symptoms, or it may be as vast as a list of all possible drugs or diseases in a subject's medical history. In many cases, researchers will simply compose sets of terms that meet their immediate needs and then require all investigators participating in the study to apply them consistently.

Because the terms used in clinical research are often identical to those used in clinical care, standard multi-use terminologies (such as those described in Chap. 7) are often appropriate for use in capturing clinical research data. However, there are some aspects of clinical research that are not well represented in mainstream terminologies; and in these cases, terminologies and their richer forms, ontologies, that are more focused on clinical research, are required. In particular, clinical research data and workflow models require controlled terminologies and ontologies that define domain-specific concepts and standard **common data elements** (**CDEs**). Collections of standard terms for CDEs can be found in the NCI's Cancer Data Standards Repository (caDSR) and Enterprise Vocabulary Service (EVS).[3] In a similar manner, the Ontology for Biomedical Investigations (OBI) (Kong et al. 2011) is being developed by a consortium of representatives from across the spectrum of biomedical research, and includes terms to represent the design of protocols and data collection methods, as well as the types of data obtained and the analyses performed on them.

There are several reasons for considering the use of *standard* controlled terminologies in the capture of clinical research data. One reason is to take advantage of clinical data that are already being collected on research subjects for other purposes. A common example is the use of data on morbidity and mortality that are collected using one of the various versions and derivatives of the *International Classification of Diseases* (ICD; see Chap. 7). In the US, for example, patient diagnoses are reported for billing purposes using the *Clinical Modifications* of the ninth edition of ICD (ICD-9-CM). While such coded information is readily available, researchers repeatedly find that ICD-9-CM codes assigned to patient records have an undesired level of reliability or granularity, especially when compared to with the actual content of the records (Iezzoni 1990). Thus the convenience of using such standard codes may be outweighed by the imprecision, which can adversely affect study design and analytical results.

A second reason for adopting a standard controlled terminology is simply to avoid "reinventing the wheel." As is described in Chap. 7, a great deal of effort has been expended in the creation of domain-specific terminologies that are comprehensive, unambiguous, and maintained over time. Designating such terminologies for use in a protocol design can relieve researchers of having to worry about the quality of the terminology. For example, a researcher is unlikely to encounter novel concepts when recording subjects' demographic data, such as gender, marital status, religion and race. Specifying, for example, that an ISO standards should be used for these data elements greatly simplifies the protocol-design process.

A third reason for choosing standard terminologies relates to the ability to compare data collected in one study with those collected in others. For example, the use of a standard scale for recording a subject's pain will allow comparison of results from a study of one treatment with those from a second study of another treatment. The selection of an appropriate standard for a particular purpose is not straightforward (for example, by 2012 the NIH Pain Consortium was listing six different scales[4]). The choice may be

[3] http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr (Accessed December 12, 2012)

[4] http://painconsortium.nih.gov/pain_scales/NumericRatingScale.pdf (Accessed December 12, 2012) http://painconsortium.nih.gov/pain_scales/COMFORT_Scale.pdf (Accessed December 12, 2012) http://painconsortium.nih.gov/pain_scales/FLACCScale.pdf (Accessed December 12, 2012)
http://painconsortium.nih.gov/pain_scales/CRIESPainScale.pdf (Accessed December 12, 2012)
http://painconsortium.nih.gov/pain_scales/ChecklistofNonverbal.pdf (Accessed December 12, 2012)
http://painconsortium.nih.gov/pain_scales/Wong-Baker_Faces.pdf (Accessed December 12, 2012)

determined simply based on the emerging popularity of one terminology over another in a wide community of those investigating similar problems.

A fourth use of standard terminologies relates to reporting requirements. Government agencies sometimes require the reporting of clinical research data and, when they do, often require certain data to be reported using a particular standard. For example, the FDA requires the use of the *Medical Dictionary for Regulatory Activities* (MedRA) for reporting all adverse events occurring in drug trials (Brown et al. 1999), while the Cancer Therapy Evaluation Program (CTEP) at the National Cancer Institute (NCI) requires the use of *Common Terminology Criteria for Adverse Events* (CTCAE) (Trotti et al. 2003). In an analogous manner, at the international level, the World Health Organization requires the use of the Adverse Reactions Terminology (WHO-ART).[5] Faced with such reporting requirements, researchers sometimes choose to record data in these terminologies as they are being captured. In those cases where the clinical questions being answered require more detailed data, however, researchers must resort to recording data with some other standard (such as SNOMED; see Chaps. 7 and 25), or a controlled terminology of their own creation, and then translating them to the terminology or terminologies required for reporting purposes. See Fig. 26.7 for a comparison of how similar clinical concepts are represented in various standard terminologies.

26.3.3.3 Sharing or Reusing Multimodal Data to Support Clinical Research

Standards also exist for organizing clinical research data to enable sharing, reuse, and aggregation. CDISC, introduced previously and in Chap. 7, is a standards group motivated by the needs of the pharmaceutical and bio-technology industry entities that sponsor or otherwise support many clinical studies. CDISC is creating a standard for submitting regulatory information to the FDA, while in a similar manner, HL7 has created a standard *Clinical Document Architecture* (CDA; see Chap. 7).

The caBIG project provides a variety tools to allow researchers, clinicians and patients to share, integrate and analyze data (Buetow and Niederhuber 2009). These tools are distributed as open-source software that is made freely available to a consortium of individuals and organizations who contribute to the common goal of advancing translational research (see Chap. 25). Although the work centers around cancer research, few aspects of the models or tools are specific to that domain, and researchers from other specialties are finding caBIG resources to be valuable for their own research.

Informatics for Integrating Biology and the Bedside (i2b2) is a project being developed under a National Center for Biomedical Computing grant from NIH. Originating from the research and development activities of Partners HealthCare System and Harvard University, i2b2 is developing an information system framework to allow clinical researchers to use existing clinical data for discovery research (Murphy et al. 2006). The i2b2 platform includes a workflow framework and a data repository, as well as tools for terminology management and natural language processing. Of note, many of the over 60 institutions receiving Centers for Translational Science Awards (CTSA) from the National Center for Advanced Translational Science (NCATS) are adopting i2b2 technologies to support research and collaboration.

For those seeking to share data, and to avail themselves of data shared by others, the National Center for Biotechnology Information (NCBI) at the NIH's National Library of Medicine is creating a public repository of individual-level data, including exposure history, signs, symptoms, diagnostic test results, and genetic data. Called the *Database of Genome and Phenome* (dbGAP), this project provides stable data sets that allow multiple researchers to reference the same samples in their publications of secondary analyses of the data (Mailman et al. 2007). Additional data from

[5] http://www.umc-products.com (Accessed 12/12/2012)

ICD-9-CM:
Symptoms
787 Symptoms involving digestive system
787.0 Nausea and vomiting
787.01 Nausea with vomiting
787.02 Nausea alone
787.03 Vomiting alone

CTCAE:
Gastrointestinal Disorders
Nausea
Grade 1: Loss of appetite without alteration in eating habits
Grade 2: Oral intake decreased without significant weight loss, dehydration or malnutrition
Grade 3: Inadequate oral caloric or fluid intake; tube feeding, TPN, or hospitalization indicated
Grade 4: Life-threatening consequences (version 3.0 only)
Grade 5: Death (version 3.0 only)

MedDRA (partial)**:**
10017947 - Gastrointestinal disorders
10018012 - Gastrointestinal signs and symptoms
10028813 - Nausea
10028815 - Nausea alone
10066962 - Procedural nausea
10036285 - Postoperative nausea
10028818 - Nausea postoperative

SNOMED-CT (partial)**:**
404684003 - Clinical finding
118234003 - Finding by site
386617003 - Digestive system finding
386618008 - Gastrointestinal tract finding
422587007 - Nausea
51885006 - Morning sickness
37031009 - Motion sickness
33902006 - Air sickness
21162009 - Outerspace sickness
17783003 - Car sickness
18530007 - Sea sickness
249502005 - Train sickness
16932000 - Nausea and vomiting
64581007 - Postoperative nausea
1488000 - Postoperative nausea and vomiting

Fig. 26.7 Examples of terms used to represent research subjects' report of nausea, taken from *ICD-9-CM*, *CTCAE*, *MedDRA*, and *SNOMED*. See text for explanation of acronyms

clinical trials, currently limited to summary results, are also being made available by the NLM through the *ClinicalTrials.gov* resource, which is a repository of descriptive metadata related to historical and actively recruiting clinical trials (Zarin et al. 2011).

26.3.3.4 Clinical Research Reporting Requirements

Requirements for reporting research data, particularly those related to outcomes and adverse events, are generally accompanied by specifications for the format of the data being reported. For example, the FDA's Center for Drug Evaluation and Research (CDER) accepts reports using the HL7 Individual Case Safety Report, while the NCI's CTEP allows submission of adverse event information to its Adverse Event Expedited Reporting System (AdEERS[6]) either manually, using a Web-based application, or electronically via a web-services API. As mentioned earlier in this section, these agencies require that data be coded with standard terminologies, such as MedDRA and CTCAE, respectively.

Several reporting requirements have emerged for the purpose of making clinical trial results publicly available, both to support reuse of the data by researchers and as information sources for patients and their families. In 2000, the US National Library of Medicine launched ClinicalTrials.gov to provide a mechanism for researchers to voluntarily register their trials so that those interested in participating as research subjects can identify, via the World Wide Web, studies relevant to their condition. ClinicalTrials.gov currently includes information from over 100,000 trials from over 170 countries. In 2004, the European Union initiated as similar effort, called the European Union Drug Regulating Authorities Clinical Trials (EudraCT). ClinicalTrials.gov and EudraCT also support the reporting of the clinical trials results. While the submissions are nominally voluntary, federal agencies often mandate the reporting as a requirement for obtaining research funds or to obtain approval for regulated drugs and devices. In the US, for example, the Food and Drug Administration Amendments Act of 2007 (FDAAA) strongly reinforced these requirements. In addition, the over 900 peer-reviewed biomedical journals that participate in the International Consortium of Medical Journal Editors (ICJME) now require public, prospective registration in ClincialTrials.gov or similar databases of clinical trials of all interventions (including devices) in order for resultant manuscripts to be considered for publication.

Each repository has defined its own mechanisms for transmitting protocol data. ClinicalTrials.gov, for example, allows investigators to enter their data through an interactive Web site or to upload data in a defined XML (**eXtensible Markup Language**) format (see Chap. 5). Figure 26.8 shows an example of outcomes and adverse event data in this format. Clinical research data management systems that can export their study in this format can save the research much manual effort and assure accurate data entry (Zarin 2011).

26.4 Future Directions for CRI

As the preceding sections illustrate, significant progress has been made to advance the state of the CRI domain, and such advances have already begun to enable significant improvements in the quality and efficiency of clinical research (Chung et al. 2006; Murphy et al. 2012; Payne et al. 2005, 2010). These advances can be viewed as having been achieved at the individual investigator level (e.g., improvements in protocol development, study design, participant recruitment, etc.), through approaches and resources developed and implemented at the institutional level (e.g., development of methods and resources in data warehousing that enable storage and retrieval of clinical data for research, development of novel clinical trials management systems, etc.), and through mechanisms that have enabled and facilitated the endeavors multi-center research consortia to drive team science (e.g., innovations that enable data management and interchange for multi-center studies, etc.).

Many of these advances have been motivated by national and international funding and policy efforts that span the research and clinical-care enterprises. Among these are research funding

[6] http://ctep.cancer.gov/protocolDevelopment/electronic_applications/adeers.htm (Accessed December 12, 2012)

```
<clinical_results>
  <study_identifiers>
    <nct_id>NCT00687609</nct_id>
    <brief_title>Pilot Evaluation of Atomoxetine on Attention Deficit…. </brief_title>
    <results_first_received_date>July 22, 2010</results_first_received_date>
  </study_identifiers>
  <participant_flow>
    <group group_id="P1"><title>Atomoxetine</title>
      <description>0.5 milligrams per kilogram (mg/kg) daily …</description>
    </group>
    <period><title>Overall Study</title>
      <milestone><title>STARTED</title></milestone>
  </participant_flow>
  <baseline>
      <measure><title>Age</title><units>Years</units>
        <param>Mean</param><dispersion>Standard Deviation</dispersion>
        <measurement group_id="B1" value="16.3" spread="0.95"/>
      </measure>
  </baseline>
  <outcome>
    <type>Primary</type>
    <title>Attention Deficit Hyperactivity Disorder (ADHD) Rating Scale…</title>
    <description>Measures the 18 symptoms contained in the….</description>
    <time_frame>12 weeks</time_frame>
    <group group_id="O1"></group>
    <measure><title>Number of Participants</title><units>participants</units></measure>
  </outcome>
  <reported_events>
    <other_events>
      <event><sub_title>Nausea</sub_title>
        <counts group_id="E1" events="8" subjects_affected="5" subjects_at_risk="7"/>
      </event>
    </other_events>
  </reported_events>
</clinical_results>
```

Fig. 26.8 Example of XML representation of a clinical trial, including outcomes data, from Clinical Trials.gov. For clarity, some sections of the document have been omitted

efforts like those we have mentioned (e.g., the NCI's caBIG initiative programs (Buetow 2009; Oster et al. 2008; Saltz et al. 2008), and the CTSA initiative). In addition, investments being made to accelerate the adoption and "meaningful-use" of health IT for clinical practice (see Chaps. 12 and 27) are laying the groundwork for grand opportunities to accelerate the research and discovery enterprise. Examples of this include initiatives by the US Office of the National Coordinator for Health IT (ONC) and the US Centers for Medicare and Medicaid services (CMS), that anticipate the health IT infrastructure in the US will ultimately be able to support information management and exchange in a manner that will enable the reuse of data and information from clinical care for improvements in public health and research. As described by the ONC-based leaders of this initiative, this should greatly enable the creation of the so-called **learning health system** (Friedman et al. 2010).

As such efforts progress, and the demand for more evidence-based health care increases, the methods, theories and tools of CRI will be essential complements to those of clinical informatics in order to realize the potential of increasingly interconnected systems and ever-growing databases that can enable discovery and advance human health. While most of the efforts to date have appropriately focused on the development

of technological solutions to the issues of data capture, storage and retrieval, future advances will increasingly require effort not just to advance the development and management of technologies and platforms, but also to enhance the foundational science of CRI in an increasingly electronic world (Payne et al. 2010). By facilitating an understanding of the information-dense aspects of clinical research, CRI methods and resources will increasingly drive hypothesis generation as well as facilitate the conduct of research programs to generate new and meaningful knowledge. CRI approaches and theories will enable the meaningful use of EHRs and other biomedical information systems that extend beyond the early definitions of meaningful use limited to the systematic capture of key data elements solely for clinical care. Ultimately, systems will drive not only adherence to current guidelines and increasingly the translation of scientific discoveries into practice via evidence-based-medicine, but also will feed back data from routine practice to generate evidence by informing hypotheses and driving future research based on real-world clinical experiences, thereby completing the translational cycle.

Even as progress continues toward such a goal, the landscape of research continues to change, thereby motivating ongoing developments in the CRI domain. For example, research efforts are expanding beyond the traditional environments of single academic medical centers to multi-center, community-based and global locations of research. While there are a variety of reasons for this, cost-effectiveness and efficiency are often cited among them. Given the information intensive nature of research and these fundamental changes to the nature and location of research activities, new CRI solutions and methods will be needed to enable efficient and effective research across geographical and institutional boundaries. To address this, new funding for research into such CRI solutions and methods is emerging from agencies including NIH, the Agency for Healthcare Quality and Research, and the Patient Centered Outcomes Research Institute (Lauer and Collins 2010; Slutsky and Clancy 2010).

Even with all of the progress in CRI over the past several years, as a 2009 study of self-identified CRI professionals documented, there exists a range of fundamental challenges and opportunities facing the domain. These were sorted into 13 distinct categories that spanned multiple stakeholder groups (Fig. 26.9). In addition to helping to define the current state of the domain, the challenges and opportunities identified offer a view of the work that will face CRI professionals in the coming years (Embi and Payne 2009).

One key element that will need to be addressed in order to achieve the advances envisioned for CRI is the growth of a dedicated workforce of experts focused in the CRI domain. Currently, most CRI professionals come to the field from many different disciplines and professional communities, including computer science, information technology, clinical research and various health care domains. Recent initiatives by consortia like the CTSA institutions as well as those of professional associations like AMIA have begun to provide professional communities and venues, such as the AMIA Summits on Translational Science[7], for scientific information sharing among those working in the CRI discipline. In addition, while there is as yet no dedicated scientific journal focused on the CRI domain, CRI-focused publications are increasingly found in major peer-reviewed journals, including a number of recent special issues published in both the Journal of the American Medical Informatics Association (JAMIA) and Journal of Biomedical Informatics (JBI) highlighting distinguished papers from the AMIA Joint Summits on Translational Science (Sarkar and Payne 2011).

Despite such progress, there remains a need to address the shortage of professionals dedicated to advancing the CRI domain. While formal programs specifically for training professionals in CRI are limited, National Library of Medicine supported fellowship programs focused on CRI

[7] http://www.amia.org/meetings-and-events (Accessed 12/7/2012)

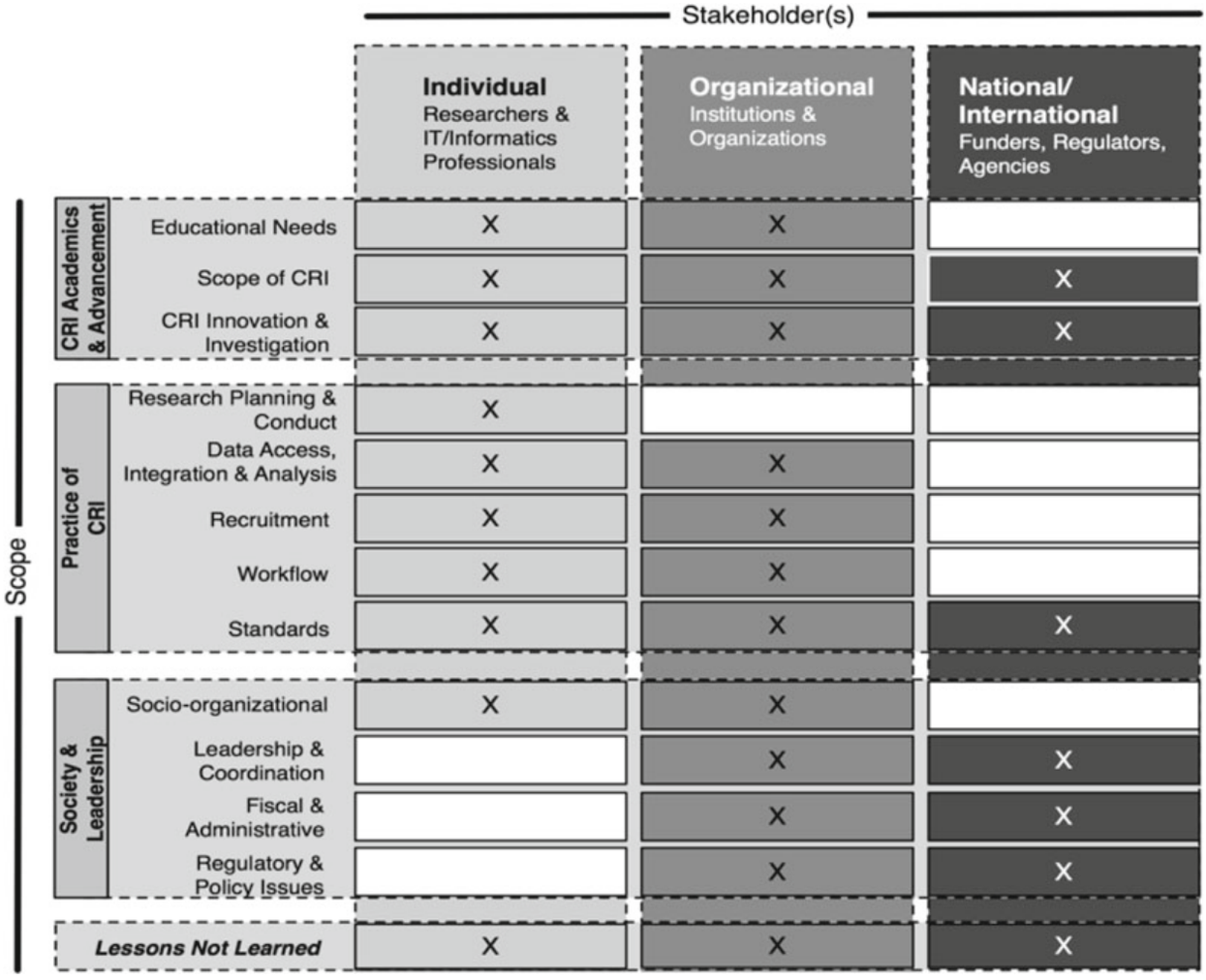

Fig. 26.9 Major challenges and opportunities facing CRI: This figure provides an overview of identified challenges and opportunities facing CRI, organized into higher-level groupings by scope, and applied across the groups of stakeholders to which they apply (Reproduced with permission, Embi and Payne 2009)

are emerging[8] and are expected to grow in the coming years. As with many biomedical informatics sub-disciplines, CRI curricula can be expected to be interdisciplinary, requiring the study of topics ranging from research methods and biostatistics, to regulatory and ethical issues in CRI, to the fundamental informatics and IT topics essential to data management in biomedical science. In addition, given the expectation that clinical information systems and environments will increasingly be sources of data and subjects for research, there is also a need to train not only technicians conversant in both clinical research and biomedical informatics to work in the CRI space, but also to educate clinical informaticians, clinical research investigators and staff, and institutional leaders concerning the theory and practice of CRI. Programs like AMIA's 10×10 initiative[9] and tutorials at professional meetings offer examples of what can be expected to grow. For example, The Ohio State University currently offers a distance education program focusing on CRI via the aforementioned 10×10 initiative[10].

As CRI continues to mature as a discipline, the current efforts focused on the relative "low

[8] http://www.nlm.nih.gov/ep/GrantTrainInstitute.html (Accessed 12/20/2012)

[9] http://www.amia.org/education/10x10-courses (Accessed 12/7/2012)

[10] http://medicine.osu.edu/bmi/education/distance/10x10/pages/index.aspx (Accessed 12/18/2012)

hanging fruit" of overcoming the significant day-to-day IT challenges that plague our traditionally low-tech research enterprise can be expected to give way to fundamental and systematic advances. In this way, CRI progress can be expected to mirror that seen years ago in the now-relatively more mature clinical informatics domain. Future years can be expected to see CRI not only instrument, facilitate and improve current clinical research processes, but it will generate advances that can be expected to change fundamentally the pace, direction, and effectiveness of the clinical research enterprise and discovery. Through CRI biomedical advances, discovery, health care quality improvement, and the systematic generation of evidence will become as routine and expected as advances in clinical informatics have already become in fostering the systematic application of evidence into health care practice.

26.5 Conclusion

This chapter has sought to introduce the following major themes: (1) design characteristics that serve to define contemporary clinical studies; (2) foundational information needs inherent to clinical research programs and the types of information systems can be used to address or satisfy such requirements; (3) the role of multi-purpose platforms, such as Electronic Health Record (EHR) systems, that can be leveraged to enable clinical research programs; (4) the role of standards in supporting interoperability across and between actors and entities involved in clinical research activities; and (5) future directions for the CRI domain and how such endeavors they may alter or optimize the conduct of clinical research. As we have explained, the clinical research environment is faced with significant workflow and information management challenges, and it is therefore increasingly garnering attention from the governmental, academic, and private-sectors. This progress explains CRI's emergence as a distinct and highly valued sub-discipline of biomedical informatics. Part of the evolution of CRI can be attributed to the extraordinary increase in the scope and pace of clinical and translational science research and development that has been catalyzed by a variety of funding and policy initiatives that seek to re-engineer the way in which governmental, public, and private entities advance basic science discoveries into practical therapies. CRI has accordingly become a dynamic and relevant sub-domain of biomedical informatics knowledge and practice, providing a broad spectrum of research and development opportunities in context of both basic and applied informatics science.

Discussion Questions

1. How do the foundational information needs of clinical research differ depending on the type and phase of study being undertaken? Do study phases have an impact on the primacy of such information needs?
2. What is the role of biomedical informatics with regard to decreasing bias in RCTs and thus enhancing the internal validity, external validity, and generalizability of study results?
3. How can clinical or general purpose information systems and research-specific tools be employed synergistically to address clinical research-specific information needs, such as participant recruitment or the population of study-specific data capture instruments?
4. How do the core functional components of common clinical trial management systems (CTMS) overlap with or otherwise replicate the functionality of electronic health record (EHR) systems? To what extent does this similarity or difference inform the need for syntactic and/or semantic interoperability among such systems?
5. In what situations is the use of clinical research-specific terminologies or ontologies appropriate? In such situations, what challenges exists relative to

the selection, use, and maintenance of appropriate standards?

6. What is the role of data standards in enabling the dissemination and reuse of study-generated data sets? How can the use of such standards enable the cross-linkage or integrative analysis of data sets derived from multiple but independent studies?
7. Compare and contrast the future directions of CRI with those of other BMI sub-disciplines and focus areas described in this book. To what extent are they similar and different, and what are the implications of such findings relative to the role of common informatics theories and methods and their applicability to the clinical research domain?

Suggested Readings

Ash, J. S., Anderson, N. R., & Tarczy-Hornoch, P. (2008). People and organizational issues in research systems implementation. *Journal of the American Medical Informatics Association, 15*(3), 283–289. This paper summarizes critical people-centric issues, including cultural norms, policies, and perceptions, that influence the adoption and optimal use of informatics approaches intended to facilitate clinical research.

Chung, T. K., Kukafka, R., & Johnson, S. B. (2006). Reengineering clinical research with informatics. *Journal of Neurocytology, 54*(6), 327–333. This paper describes an exploration of socio-technical factors to be addressed in order to realize the benefits of integrating Biomedical Informatics theories and methods with the design, conduct, and dissemination of clinical studies.

Embi, P. J., & Payne, P. R. (2009). Clinical research informatics: challenges, opportunities and definition for an emerging domain. *Journal of the American Medical Informatics Association, 16*(3), 316–327. This manuscript summarizes a multi-modal study that serves to both define the field of clinical research informatics, as well as critical opportunities and challenges associated with the field.

Gallin, J. I., & Ognibene, F. (2012). *Principles and practice of clinical research* (3rd ed.). London: Academic Press. This book provides a broad overview of various aspects of clinical research with particular attention to data capture and quality.

Kush, R. D., Helton, E., Rockhold, F. W., & Hardison, C. D. (2008). Electronic health records, medical research, and the tower of Babel. *The New England Journal of Medicine, 358*(16), 1738–1740. This report provides a broad overview of critical standards commonly encountered in the clinical research domain.

Payne, P. R., Johnson, S. B., Starren, J. B., Tilson, H. H., & Dowdy, D. (2005). Breaking the translational barriers: the value of integrating biomedical informatics and translational research. *Journal of Neurocytology, 53*(4), 192–200. This report provides a targeted review of literature and supporting survey data highlighting informatics-specific issues related to the delivery and support of technologies and services capable of supporting and facilitating clinical research in academic health centers.

Richesson, R. L., & Krischer, J. (2007). Data standards in clinical research: gaps, overlaps, challenges and future directions. *Journal of the American Medical Informatics Association, 14*(6), 687–696. This paper provides an in-depth look at the current roles of the various terminology and data modeling standards that are relevant to clinical research informatics.

Sung, N. S., Crowley, W. F., Jr., Genel, M., Salber, P., Sandy, L., Sherwood, L. M., et al. (2003). Central challenges facing the national clinical research enterprise. *JAMA : The Journal of the American Medical Association, 289*(10), 1278–1287. This paper provides summary of critical cultural, policy, methodological, and technical challenges that serve as impediments to the efficient and timely conduct of clinical research in the United States. This report is the result, in part, of seminal work in this area conducted by the IOM.