Clinical Trial Registries, Results Databases, and Research Data Repositories

21

Karmela Krleža-Jerić

Abstract

Trial registration, results disclosure, and sharing of analyzable data are considered powerful tools for achieving higher levels of transparency and accountability for clinical trials. The emphasis on disseminating knowledge and growing demands for transparency in clinical research are contributing to a major paradigm shift in health research. In this new paradigm, knowledge will be generated from the *culmination* of all existing knowledge – not just from bits and parts of previous knowledge, as is largely the case now. The full transparency of clinical research is a powerful strategy to diminish publication bias, increase accountability, avoid unnecessary duplication of research (and thus avoid research waste), efficiently advance research, provide more reliable evidence for diagnostic and therapeutic interventions, and regain public trust. Transparency of clinical trials, at a minimum, means sharing information about the design, conduct, results, and analyzable data. Not only must the information itself be explicitly documented, but an access location or medium for distribution must be provided. In the case of clinical trials, the public disclosure of data is realized by posting cleaned and anonymized data in well-defined, freely accessible clinical trial registries and results databases. Making cleaned, anonymized individual participant data sets analyzable is still a challenge.

Basic electronic tools that enable sharing clinical trial information include registries hosting protocol data, results databases hosting aggregate data, and research data repositories hosting reusable/analyzable data sets and other research-related information. These tools are at different levels of development

K. Krleža-Jerić, MD, MSc, DSc (✉)
IMPACT Observatory, Mediterranean Institute for Life Sciences- MedILS,
Split & Cochrane Croatia, University of Split School of Medicine, Split, Croatia

Electronic Health Information Laboratory -EHIL, CHEO, Ottawa, ON, Canada
e-mail: karmela@krleza.com

453

and plagued with heterogeneity as international standards exist only for trial registration. The lack of standards related to publishing data in repositories makes it difficult for researchers to decide where to publish and search for data from completed studies.

Background

The movement toward open science and open data (i.e., making raw data from research available for analysis) is slowly beginning to penetrate clinical trials [1]. For clinical trials, any discussion of raw data refers specifically to the cleaned and anonymized individual participant data (IPD). However, consumers of these data ultimately need analyzable data sets, which include IPD, metadata, and adjacent (or supporting) documents.

The clinical trial enterprise is international, and therefore the development of clinical trial registries, results databases, and research data repositories should be at an international level and with open access. Such international standards should be flexible to allow elaboration of required fields and addition of more fields as needed.

There are three broad types of clinical trial data that can be shared publicly or openly: protocol, results and findings, and raw data sets [2]. More precisely, these include:

(a) The registration of selected protocol elements in trial registries which might be complemented by publication of full protocols in journals.
(b) The public disclosure of summary results (aggregate data) in databases, usually developed by clinical trial registries; these are usually beyond publications in peer-reviewed journals.
(c) The public availability of analyzable data sets; these data sets are based on cleaned, anonymized individual participant data (IPD) and adjacent trial documentation.

There are several modes or mechanisms of finding and accessing IPD-based analyzable data sets for secondary analysis (often called pooled or meta-analysis of IPDs). These include (a) direct researcher-to-researcher contact (reviewer contacting initial data producers), (b) initiatives and projects that play intermediary role, and (c) publicly accessible repositories.

(a) *Direct researcher-to-researcher contact*: The reviewer gets the data directly from the original data creator by contacting him or her. The reviewer identifies studies mainly by following the literature and/or by visiting trial registries.

(b) *Intermediary contact in which the researcher requests data from special* initiatives or projects including Clinicalstydatarequest, [3] Yoda [4], the Project DataSphere [5] and recently launched Vivli [6, 7]: The reviewer applies for data to an independent panel, a sort of peer-reviewed panel that is formed by a group of data pharmaceutical industry providers or producers (generally the pharmaceutical industry at present). The panel is usually independent international panel. Increasingly, government agencies are also moving to this direction, such as the European Medicine Agency (EMA) [8].

(c) *Open-access, publicly accessible research data repositories* (in further text repositories). They might be either domain repositories that specialize in hosting clinical trial data or general repositories that host clinical trial data in addition to hosting raw data from several or all research areas. There are currently several such open-access general research data repositories in public domain that host CT data.

In this chapter, we focus on registries, databases, and repositories.

Rationale

Trial registration, results disclosure, and making analyzable IPD-based data publicly available all share the same underlying rationale. All three are based on the principles of making the most out of clinical research, diminishing research waste, and enhancing knowledge creation. Trial registration, results disclosure, and data sharing are considered powerful tools for achieving higher levels of transparency and accountability of clinical trials [9]. Increasing emphasis on knowledge sharing and growing demands for transparency in clinical research are contributing to a major paradigm shift in health research that is well underway. In this new paradigm, knowledge will be generated from the *culmination* of all existing knowledge – not just from bits and parts of previous knowledge, as is largely the case now [10].

A stepwise process of opening clinical trial data began with the registration of protocol elements, but it was clear from the very beginning that without results disclosure, the registration would be an empty promise. Later on, it became well understood that transparency would be not be achieved without results *and* data disclosure. Actually, one could argue that results disclosure includes publication in a journal, posting summary results in open-access Internet-based database or registry, and publishing analyzable data sets in research data repository.

We are firmly in the era of evidence-informed decision-making in health for both individuals and populations at all levels – local, regional, national, and global. This decision-making is multifaceted, from the individual patient via physician to health administrators and policy-makers [10]. Registration of protocol items, publication
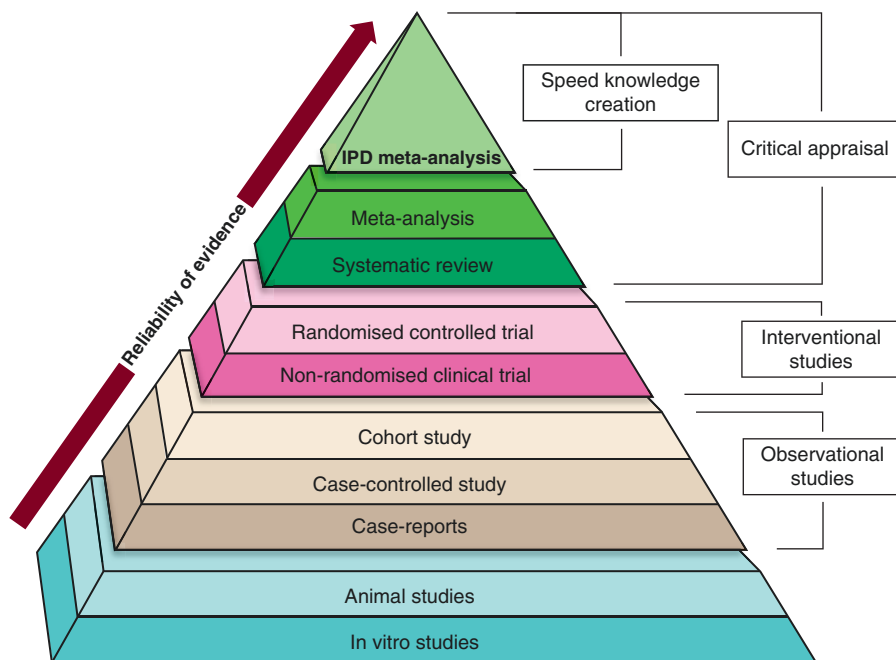
Fig. 21.1 Evidence pyramid – reliability of evidence that can be used for decision-making in health

of the complete protocol, and public disclosure of trial findings in peer-reviewed journals – complemented with public (Internet-based) disclosure of results including aggregate data and IPD-based analyzable data sets – represent a *totality* of evidence and knowledge for a given topic area and are integral to supporting efforts toward evidence-informed decision-making.

Evidence is needed to support many personal and policy decisions in health and in research. Randomized clinical trials, systematic reviews, and increasingly IPD-based meta-analyses are considered gold standards for evidence creation, illustrated by their positions at the top of the pyramid of evidence (Fig. 21.1). Actually, there has been quite an evolution from the acceptance of a systematic review (i.e., reana-lyzing the *aggregate or summarized data*, usually obtained from publications) as a gold standard to the growing notion that the gold standard should require the meta-analysis using *the raw data*. This position of clinical trials on the evidence pyramid implies that the reliability of results generated by clinical trials is indeed very important. As the evidence gained from clinical trials might be directly implemented into clinical decision-making, it follows that the quality of these results should be continually scrutinized. Unfortunately, the reliability of trial-based evidence is questionable due to publication and outcome reporting bias of trials, as well as the lack of data sharing – which means that others cannot replicate or verify results. Consequently, incomplete evidence can lead to biased clinical decisions, with often harmful consequences, and can damage public trust in research and medical inter-ventions. Following medical deontology, doctors' prescription habits are supposed to be judiciary, which requires complete and total knowledge of the benefits and

potential harms of prescribed medications. This is difficult at best and impossible if the information about the given diagnostic tools, medications, or devices is not available or is incomplete and thus biased [9, 10].

The full transparency of clinical research is a powerful strategy to diminish publication bias, increase accountability, avoid unnecessary duplication of research, avoid waste, advance research more efficiently [2], provide more reliable evidence for diagnostic and therapeutic prescriptions, speed knowledge creation, and regain public trust [10]. Transparency of clinical trials, at a minimum, means sharing information about design, conduct, and results. The information itself must be explicitly documented, but then an access location or medium for distribution must be provided. Until recently, the public disclosure of clinical trial data was realized by posting them in well-defined, freely accessible clinical trial registries and results databases. Since the first version of this chapter in 2012 [11], a lot has changed. Open-access research data repositories have been developed, and the analyzable data sets (i.e., IPDs and adjacent documentation needed to make data analyzable) can be made publicly available by publishing them in such repositories.

Considering that trials take place internationally and that the knowledge gained by them may be used by anyone anywhere in the world, their quality is also constantly and internationally scrutinized. Thus, the related standards should be internationally defined and relevant. While there are standards for trial registration and registries, the standards for results disclosure and, most importantly, standards for preparing clinical trial data for public sharing (including the definition of the requirements for repositories that host them) have yet to be developed.

Trial Registration

Development of Trial Registration

Although the need for trial registration (i.e., publishing protocol information) has been discussed for several decades, only at the beginning of this millennium did trial registration garner widespread attention from many stakeholders representing varied perspectives. The practical development of trial registration began around 2000 with two critical boosts in 2004 and in 2006. The 2004 New York State Attorney General vs. Glaxo case [12, 13] inspired the International Council of Medical Journal Editors (ICMJE) [14] and Ottawa statement [15] as well as the recommendations of the Mexico Ministerial Summit organized by the World Health Organization (WHO) [16]. These led to the development of international standards for trial registration by the WHO, which were launched in 2006 and changed the landscape of trial registration worldwide [17]. As we learned by the IMPACT Observatory scoping review [18], a number of circumstances had coincided by the year 2000 (earlier than initially thought) which enabled the development of data sharing, beginning with trial registration. These include:

- Internet-enabled storage and retrieval of large data sets
- The definition of data, metadata, and evidence-based (now increasingly called evidence-informed) medicine

- The use of evidence gained by systematic reviews and initial IPD-based meta-analysis in decision-making
- The appreciation of the impact of trial registration on knowledge creation, sharing, and Knowledge translation-KT
- The existence and experience of two major registries: the International Standard Randomized Clinical Trials Number (ISRCTN) http://www.isrctn.com/, based in the UK, and ClinicalTrials.gov, based in the USA
- Growing awareness of the need to enhance transparency
- The willingness of the international research community to embark on this undertaking
- The awareness of the harmful consequences of decision-making in the context of partial evidence
- The powerful arguments from oncology, pediatrics, rare diseases, AIDS, pregnancy, perinatal medicine, and media reporting trail-related scandals
- The need to stop wasting precious resources in unnecessary duplication of research

The initial international trial registration standards that were launched by WHO in 2006 provided essential contribution toward achieving the evidence-informed decision-making. These standards clearly identify existing registries and trials that need to be registered, define the minimum data set, designate the timing of registration, assign unique numbers to trials, and set international standards to facilitate the development of new national or regional registries as well as the comparability of data across registries. It is important to note that as of 2018, there are no international standards for results disclosure or public sharing of analyzable data. However, these are likely to be developed in the near future and will create numerous opportunities for informatics and information technology (IT) experts to leverage and apply to new applications. Additionally, further evolution of trial registration and its standards has been taking place, again leading to new applications and resources that will undoubtedly impact the development of new research and our subsequent understanding of health, disease, and effective therapies.

The goal of research transparency includes having protocol documents electronically available. For example, the protocol documents should be posted on the registry website, and all trial-related data from them ideally can be cross-referenced to results and findings. However, in reality, a trial protocol can be very complex and lengthy, which can make finding the needed information difficult. To overcome this, an international group defined the set of Standard Protocol Items for RandomIzed Trials (SPIRIT), developed SPIRIT guidelines, and made them publicly available [19–21].

SPIRIT is expected to increase the clarity of clinical research protocols and ensure that the collection of necessary items is indeed specified in the protocol, thus contributing to the overall quality of the protocol and presumably the study and results it generates. The use of SPIRIT guidelines in development of protocols might also facilitate public disclosure, especially in combination with the growing use of electronic data management [22]. It is important to note that even if full protocols are publicly

available, the existing minimum data set of the WHO international standards will still be important as the summary of a protocol. Trial registration standards will have to be revisited frequently as methodology evolves, demands for transparency increase, and with ongoing evaluation and analysis. Trial registries will most certainly expand to include results or cross-references to results databases.

Trial Registries

A *clinical trial registry* is an open-access, Internet-based repository of defined protocol information. Many different kinds of clinical trial registries exist in the public and private domains, such as international-, country-, and region-specific registries, as well as corporate (sponsor-driven) registries. The presence of multiple registries might be seen as a natural consequence of increased pressure and interest and as a positive development; however, a proliferation of registries could potentially lead to information overload and confusion for patients, clinicians, policy-makers, and research sponsors. For example, an inexperienced user may not know which clinical trial registries to trust. It might be expected that this situation will gradually correct itself as the evidence and best practice accumulate. Certainly, the proliferation of trial registries underscores the critical need for international standards that would define required features of registries as well as the content and supporting information that they must provide. Fortunately, such standards exist.

Standards, Policies, and Principles

Because clinical trials are conducted throughout the world, trial registration standards have to be defined on the international level. WHO developed international standards for trial registration, which were endorsed by the ICMJE, most medical journal editors, the Ottawa group, some public funders, organizations, and countries. It is important to note that individual countries often implement international standards by adopting and extending them with additional fields to host more information in their particular registries.

WHO international standards have helped shape many, if not all, trial registries and have been contributing to the quality and the completeness of data for registered trials. Also, it is expected that they will play a major role in further evolution of trial registration. They are sometimes referred to as WHO/ICMJE standards (or even cited only as ICMJE requirements, because the journal editors endorsed the WHO international standards in their instructions to authors and in related FAQs). These international standards define the scope (i.e., *all* clinical trials need to be registered), the registries that meet the well-defined criteria, the timing (i.e., prospective nature of the registration prior to the recruitment of the first trial participant), the content (a minimum data set that needs to be provided to the registry, initially referred to as a 20-item minimum data set), and the assignment of the unique identifier (ID). These international standards also define the criteria that the registry has to meet, which

includes level (nationwide or regional), ownership and governance (public or private nonprofit), trial acceptance, open access, and structure. In particular, structurally, the registry must have at least enough fields to host minimum data set that initially contained the following 20 items:

1. Unique trial number and the name of registry
2. Trial registration date
3. Secondary ID
4. Funding source(s)
5. Primary sponsors
6. Secondary sponsors
7. Responsible contact person
8. Research contact person
9. Public title
10. Scientific title
11. Countries of recruitment
12. Health condition or problem studied
13. Interventions (name, dose, duration of the intervention studied, and comparator)
14. Inclusion/exclusion criteria
15. Study type (randomized or not, how many arms, who is blinded)
16. Anticipated start date (and later on the actual start date)
17. Target sample size
18. Recruitment status (not yet recruiting, recruiting, temporarily stopped recruiting, or closed for recruitment)
19. Primary outcome(s) (name, prespecified time point of measurement)
20. Key secondary outcomes

Since 2012 few additional items were added to the list, each with precise definition and description, thus forming the version 1.3.1 of the WHO data set [23]. These new items are:

21. Ethics review
22. Completion date
23. Summary results
24. IPD sharing statement

In order to foster the implementation of standards, to facilitate creation of new registries, to identify the best practice, and to help develop trial registration policies, WHO formed a freely accessible search portal in 2007, followed in 2008 by the formation of a network of registries and of the Working Group on Best Practice for Clinical Trial Registries. The WHO International Clinical Trials Registry Platform (ICTRP) is a unique global portal to the trials in registries that meet criteria as data providers (i.e., WHO primary registries and ClinicalTrials.gov), but the platform does not provide access to the full extent of registries' data. Instead, the predefined 24-item data set provided by the registries is displayed (in English). The unique

identifier displayed is meant to be used in any communication about a trial, including in the ethics committees/boards' communications, consent forms, reports, publications, amendments, and press releases. This enables users and computer applications to collect trial data from many sources, allowing users to view the full picture of a given trial, from start to finish.

WHO ICRTP is also supporting a development of policies and regulations and posts them on its website. Many organizations are developing policies on clinical trial registration. While some countries recommend the trial registration (Canada, Australia) or make it a compulsory prerequisite in drug marketing authorization process (approving new drug for the market) such as the USA and the EU, so far only few countries have also developed regulations making trial registration compulsory. Some of these countries (e.g., India) also have registries, while Argentina, Israel, and Switzerland have regulations but do not yet have a registry.

Characteristics and Design Features of Trial Registries

The distinction between patient and trial registries might be confusing as they both capture certain disease-related information and often use Internet-based depositories. However, these two types of registries are quite different. Patient registries (Chap. 13) contain records and data on *individuals*, whereas trial registries focus on the descriptive aspects of a *research study* at various stages of its implementation and often provide a link to *study results*. While trial registries can be accessed via the WHO ICTRP global search portal, at present there is no single global search portal that can be used to identify or access patient registries.

Clinical trial registries contain predefined information about ongoing and completed clinical trials, regardless of the disease or condition addressed. Patient registries contain the disease-specific information of individual patients. In a clinical trial registry, each entry represents one trial and contains selected information from protocol documents of the trial. Clinical trials are prospective interventional studies, and they may recruit either healthy volunteers or patients with various diseases. Each trial may include any number from a few to thousands of participants. In a patient registry, each entry is an individual patient with the same disease or a condition of the same group, often chronic diseases (e.g., cancer, psychosis, and rare disease patient registries).

The most important difference between trial and patient registries is the purpose. The main goal of trial registries is to provide various stakeholders with information about ongoing and completed trials, in order to enhance transparency and accountability as well as to reduce the publication bias, increase the quality of published results, prevent harmful health consequences, and most importantly, provide knowledge that will ultimately enhance patient care. Patient registries, on the other hand, are developed in order to answer epidemiological questions such as incidence and prevalence and better understand the natural course of disease including morbidity or mortality.

Some trial registries also aim to inform potential trial participants about open or upcoming trials in order to enhance recruitment. Besides being tools for

transparency, registries can also function as learning tools, and one could argue that registries might help improve the quality of the protocol and, as a result, the quality of the trials as they are completed. For example, while entering data in predefined fields, the researcher might realize that he or she is lacking some information (i.e., elements he or she forgot to define and include in the protocol) and will address the missing element(s) by editing and enhancing the protocol.

The first version of the protocol is the initial protocol that has been approved by the local ethics committee and submitted to the trial registry. Updates for trial registries are expected and consist of providing information about the protocol in various stages of the trial: prior to recruitment, during the implementation (recruitment, interventions, follow-up), and upon completion. During trial implementation, changes of protocol, called *amendments*, often take place for various reasons. Amendments to a protocol are instantiated as new protocol versions, which are dated and numbered sequentially as version 2, 3, 4, etc. Annual updates of registry data enable posting of such amendments after approval by the ethics committees. The ability to manage multiple versions of protocol documents is an important feature for a trial registry. The basic rule for the registry is to preserve all of the descriptive data of a protocol that is ever received. Once registered, trials are never removed from the registry, but rather a *status* field indicates the stage of a trial (e.g., prior to recruitment, recruiting, do not recruit any more, completed). Earlier versions of protocol-related data are kept, are not overwritten, and should still be easily accessible by trial registry users.

WHO endorses trial registries that meet international standards and calls these *primary registries*. Registries that do not meet all the criteria of international standards are considered *partner registries*, and they provide data to the WHO search portal via one or more primary registries. The need for international access and utilization of registries implies the need for a common language. While some of these registries initially collect data in the language of the country or region, they provide data to the WHO portal in English because the WHO ICTRP currently accepts and displays protocol data in English only.

It is important to note that registries that adhere to international standards tend to add additional data fields to meet their registry-specific, often country-specific, needs. Regardless of these additional fields, the essential 24 items should always be included and well-defined. Although they are bound by the international standards, the presentation of a registry's website (i.e., the web-based access and query interface) is not the same across primary registries. Some registries collect and display protocol descriptive data beyond the basic predefined 24-item fields. Those registries that collect more data typically have more extensive and detailed data for each trial record and are potentially more useful for consumers. Some registries have free-text entry fields with instructions about which data need to be provided in the fields targeted to those registering their trials, while other registries employ self-explanatory and structured fields, such as drop-down lists [24].

The WHO formed the Working Group on Best Practice for Clinical Trial Registries in 2008 in order to identify best practices, improve systems for entering new trial protocol records, and support the development of new registries [25]. The working group includes primary and some partner registries. Since the first edition
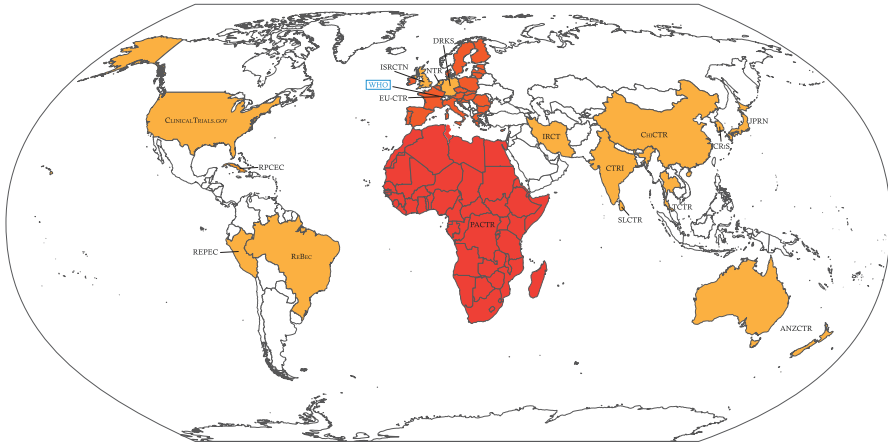
Fig. 21.2 Network of registries providing data to WHO search portal and the WHO portal – ICRTP. This map provides the worldwide distribution of registries that directly provided data to WHO as of July 2018. *ANZCTR* Australian New Zealand Clinical Trials Registry, *ReBec* Brazilian Clinical Trial Registry, *ChiCTR* Chinese Clinical Trial Registry, *CRiS* Clinical Research Information Service, Republic of Korea, ClinialTrials.gov (USA), *CTRI* Clinical Trials Registry, India, *EU-CTR EU* Clinical Trials Register, *RPCEC* Cuban Public Registry of Clinical Trials, *DRKS* German Clinical Trials Register, *IRCT* Iranian Registry of Clinical Trials, ISRCTN.org (UK), *JPRN* Japan Primary Registries Network, *NTR* The Netherlands National Trial Register, *PACTR* Pan African Clinical Trial Registry, *REPEC* Peruvian Clinical Trial Registry, *SLCTR* Sri Lanka Clinical Trials Registry, *TCTR* Thai Clinical Trials Registry, WHO Search Portal, Geneva. Note: The source of information: WHO ICRTP [17]. Since 2012 three registries, EU-CRT, TCTR, and REPEC joined the WHO primary registry network that directly provide data to WHO

of this book in 2012, 3 additional primary registries were developed, and as of June 2018, there were 17 registries that directly provide data to the WHO portal, specifically 16 WHO primary registries and the ClinicalTrials.gov registry which is not a part of primary registry network but provides data to the search portal. As can be seen from the geographic distribution shown in Fig. 21.2, the network includes at least one registry per continent.

Clinical trial registries can cross-reference a registered trial to its website if one exists; many large trials establish their own websites. Also, registries provide links and cross-references to publications in peer-reviewed journals, and some also cross-reference to trial results databases and research data repositories. It is expected that the number of these links will increase as results databases and repositories continue to be developed.

Timing

A responsible registrant, usually a specially delegated individual from the trial team or sponsoring organization, provides protocol-related data to the trial registry. Because all research protocols must be reviewed and approved by the ethics

committee or board of the local institution in order to conduct the study, the descriptive protocol data set is usually submitted to the trial registry after institutional ethics approval. Otherwise, registration in the trial registry is considered conditional until the ethics approval is obtained.

Although international standards require registration prior to recruitment of trial participants, this is still not fully implemented [24, 26]. Such prospective registration is important as it not only guarantees that all trials are registered but also that the initial protocol is made publicly available. For various reasons, the protocol might be changed early on, and/or a trial might be stopped within the first few weeks. Information about early protocol changes or stopped trials is lost unless trials are prospectively registered. Full data sharing is essential for the advancement of science and helps to avoid repeating such trials. Registries record the date of initial registration and date all subsequent updates. Additionally, the assignment and subsequent use of a unique ID for each trial upon registration enables any stakeholder to easily find what interests them.

Some countries hesitate to simply "import" the international standards or policies out of fear that these might change and put the country (regulator, or funding agency) in an odd position. One can debate the justification of such positions, but they are a reality. Implicit application of international standards occurs more often, with or without referencing them. Such is the case with the Declaration of Helsinki (DoH) [27], which obliges physicians via their national medical associations and is thus implicitly implemented. The DoH gradually addressed clinical trial registration and results disclosure, and the latest, 2013, Declaration explicitly calls for the registration and results disclosure of trials [27–29].

Quality of Registries

The quality of various trial registries can be judged by the extent to which they meet the predefined goal of achieving high transparency of trials. Considering that meeting international standards is a prerequisite to qualify as a WHO primary registry, the quality and utility of trial registries mainly depend upon the quality and accuracy of data and the timing of reporting [17]. To realize research transparency, clinical trials need to be registered prior to the recruitment of trial participants; this principle has not yet been fully achieved [26, 30, 31].

Registries constantly work on ensuring and improving the quality of data. The aim is to have correct data that are meaningful and precise. Accuracy of data requires regular updates in case of any changes and keeping track of previous versions. Registries impose some logical structure onto submitted data, but the quality is largely in the hands of data providers (i.e., principal investigators or sponsors). Many researchers and some registries perform analysis and evaluation of registry data [24, 31, 32]. IT experts might contribute by developing new, system-based solutions for quality control of entered trial data. Quality of data is a particularly sensitive issue as trial registries are based upon self-reporting by researchers, their teams, or sponsors. Following international standards and national requirements are

prerequisites for attaining an acceptable level of data quality. (Note that the practical and theoretical aspects of data quality are described in Chap. 11.)

The numerous and ongoing analyses and evaluations of implementation of standards and the quality of registries will enable revisions and updates, thereby improving trial registries at large. Furthermore, trial registries should reflect the reality of clinical trials methodology, which is constantly developing. Understandably, this presents a continuing challenge to those involved with the IT aspects of the data collection.

Registries that meet international standards might accept trials from any number of countries with data in the country's native language; therefore, it is essential to ensure the high quality of the translation of terms from any other language to English. Criteria that define quality also include transfer-related issues such as coding and the use of standard terms, such as those developed by the Clinical Data Interchange Standards Consortium (CDISC) [33]. For this reason, definitions of English terms used across registries created in different countries also require standardization, and there have been efforts to this end, notably those on the standard data interchange format developed by CDISC. Standardization of terms is an important issue, and solutions must balance the resources required for researchers and trial registry administrators to implement standard coding against the potential benefits for information retrieval, interoperability, and knowledge discovery. The ability of protocol data to be managed and exchanged electronically, including difficulties with computerized representation due to various coding standards for several elements such as eligibility criteria, is described in Chap. 10.

One of concerns for trial registries is the issue of duplicate registration. Duplicate registration of trials, especially of multicenter and multi-country trials, has been observed from the very beginning and was discussed by the WHO Scientific Advisory Group (SAG) while developing the standards. The concern is that duplicate registration in WHO primary registries/registries acknowledged by the ICMJE might lead to counting one trial as two, or even as several trials, and might skew conclusions of systematic reviews. Therefore, these registries perform intra-registry deduplication process, while the WHO search portal established mechanisms of overall deduplication called *bridging*. In that process, most registries have created a field for an identification number (ID) that a particular trial was given by another registry. They usually also have the field for the ID from the source, which is assigned by the funder and/or sponsor. Parallel registration in a hospital, sponsor-based, or WHO partner registry does not count as duplicate registration; only the registration in more than one primary registry of the WHO/registries recognized by the ICMJE qualifies as duplication. This is because those other registries have to provide their data to one primary registry or ClinicalTrials.gov to meet criteria of international standards and then data are provided to the WHO search portal.

It is important to note that clinical trials are sometimes justifiably registered in more than one primary registry. For example, international trials might be registered in more than one primary registry if regulators in different jurisdictions require registration in specific registries. In these cases, researchers need to cross-reference IDs assigned from one registry to another. For this reason, the creation of a field in the registry to host the ID(s) received by other registries is important. Also, it is

important that researchers provide the same trial title and the same version of protocol information in case of duplicate registration. The latter is particularly important in case of delayed registration in one of the registries and/or of initial data entry from a protocol that was already amended. Primary registries usually date the e-data entry, but it would be very useful to also number and date the protocol versions.

In 2009, as a part of implementing international standards, WHO established the universal trial number (UTN) [17], and registries developed a field to host it. This number is also meant to help control duplicate registrations. While designing a registry, it is thus necessary to anticipate the field to host the UTN. Likewise, nonprimary registries as well as eventual trial websites should create fields for UTN and IDs assigned by primary registries.

Evolution and Spin-Off

Mandates for registries determine their scope, substance, and consequent design. Although relatively new, trial registries are experiencing constant and rapid evolution, and the learning curve is steep for registrants, registry staff, registry users, and of course, IT professionals. The major impetus for the progress of trial registries followed the development of the WHO international standards in 2006 that expanded their scope from randomized controlled trials (RCTs) to all trials, regardless of the scope and type, and from a few items that indicated the existence of a trial to a summary of the protocol. At the same time, registries expanded fields and started to accept trials from other countries. Initially, registration included only RCTs that aimed at developing new drugs and collected only basic information. Of course, there is still significant potential for improvement. For example, many trials are still registered retrospectively or with a delay, but this is expected to get better with time [30, 34, 35].

Further evolution of the international trial registration standards is expected to respond to the evolution of trial methodology. For example, phases 0, I, and II might need different fields, while some fields designed for RCTs no longer apply. This has to be kept in mind while designing a registry.

Some registries, such as ClincalTrials.gov, primarily originated from a mandate to enable potential trial participants to find a particular RCT and to enroll in it. Overall the main purpose of registries has shifted from a recruitment tool to a transparency tool while still focusing on benefits to trial participants. While registries still facilitate patients and clinicians searching by various criteria for ongoing studies, they are also becoming a source of data on various completed trials.

The trigger for trial registration was the lack of transparency and the subsequent and disastrous health consequences shown by the New York State Attorney General vs. Glaxo trial [12, 13]. This case mobilized stakeholders and elicited consequent action from various interest groups, i.e., journals, research communities, consumer advocates, regulators, etc. Nowadays, trial registries aim to inform research and clinical decisions as well as to control publication bias in response to scientific and ethical requirements of research. As a result of the international dialogue among various stakeholders, most registries now aim to meet the needs of all involved in order to elevate research to another level.

Apparently, the compliance with international standards is weak and selective when registration is voluntary, but it is gradually becoming compulsory in many jurisdictions. Still, even when regulated, compulsory registration does not necessarily meet all the requirements of the WHO international standards. For example, in the USA, registration in ClinicalTrials.gov is required by law [36]. Investigators must comply or risk a penalty; however, the law does not require registration of all trials, and it allows a delay of 21 days for registration of trials that are covered by the Food and Drug Administration Amendments Act (FDAAA) of 2007.

The experience gained so far is expected to inspire the registration of other types of studies or the development of other research-type registries. Such "spin-off" is already taking place and includes registration of observational studies in trial registries. Another example of a spin-off is the international initiative to develop a registry of systematic reviews of clinical trials and corresponding standards. The registry PROSPERO, international prospective register of systematic reviews [37], was launched in February 2011. It is expected that such registries will function based on similar principles as trial registries. For example, PROSPERO is prospectively registering a systematic review (i.e., its design and conduct, protocol, or equivalent) and is displaying a link to eventual publication of the completed review. All the information is provided by the researcher and publicly displayed on PROSPERO's website. The registration and the usage are free of charge and freely accessible. Individual studies are the unit (record) of entry in such registries, and a mechanism for cross-referencing of study entries across various registries will be established. For example, systematic review registries might establish a cross-reference to trial registries. Such spin-off would require development of standards and creation of specific fields. Registries might provide fields to capture results or link to various levels of reporting trial results and findings, such as links to publications, capturing aggregate results data in results fields, and linking to a database with microlevel data and registry of systematic reviews.

In addition to the WHO international trial registration standards, some countries develop their own specific standards, which may meet and expand or somewhat differ from the existing standards. For example, FDAAA differs by exempting the so-called phase I and some device trials from compulsory registration. Consequently, ClinicalTrials.gov offers fields for such trials, but their registration is voluntary. There are also initiatives to develop regional registries and software that will facilitate development of individual country registries in a given region such as in the Americas [30].

Creation and Management of a Trial Registry: The User Perspective

Design of Trial Registries

As mentioned earlier, every primary trial registry now contains fields for a 24-item minimum data set as defined by the international standards and usually a few additional ones. These include the fields for the ID assigned by any other registry, the unique trial registration number (UTRN) assigned by WHO, trial website URL,

publications, etc. The required items are often expanded in several fields. For example, there may be special fields to indicate whether healthy volunteers are being recruited or to specify which participants are blinded. In parallel with registration of a minimum data set, arguments have been built for publishing the full protocol, and some journals have already started doing so. It will be particularly useful to have publicly available electronic versions of structured protocols, following SPIRIT guidelines. However, even if and when that happens, the data provided in trial registries will be useful as a summary of the protocol. These two major tools of protocol transparency (trial registry and publicly available SPIRIT-based protocol) each attract different users but undoubtedly will provide a foundation for a number of navigation and analytic tools directed toward researchers, consumers, and policy-makers.

International Standards

International standards were the major impetus for the development of trial registries. Among other advantages, standards ensure the trustworthiness of data and comparability among registries. It is important that data provided is precise and meaningful, which depends on the precision of instructions for registration and also on the fields [24]. These instructions, inspired by the WHO standards, might be developed by regulators in combination with the registry and/or journal editors as for example the Australian Clinical Trail Toolkit [38]. Registries usually have levels of compulsory completion of fields that cannot be skipped. Furthermore, they might indicate which fields or items are required by the WHO standards and/or by the appropriate national regulator. It is important to note that at this time, there are no standards for registration of observational studies, so currently registries use the trial fields and allow other descriptive data to be added.

Data Fields

The design of fields for trial registries is extremely important. Possibilities include free-text, drop-down, or predefined entries. It is advisable to define which data is needed and develop a drop-down list whenever possible. Such a drop-down list should include all known possibilities and the category "other" with text field to elaborate. Considering the rapidly developing field of clinical trials, it is necessary to anticipate additional items in a drop-down list.

Well-defined fields are prerequisite to obtain high-quality protocol data in trial registries. For example, if a registry field is free text and the data entry prompt reads *type of trial*, the answer will likely be simply "randomized controlled trial" or "randomized clinical trial" or even just the acronym "RCT." However, the registry might prespecify in a drop-down list whether the trial is controlled or uncontrolled and whether it is an RCT and whether its design is parallel, crossover, etc.

Although phases I–IV are still in use as descriptive terms, they will probably be replaced with more specific descriptions of studies in the future. Elaboration of those numbered phases is already taking place: the phase 0 has been added, and existing phases are subdivided into a, b, and c (e.g., phase II a, b, etc.). In some cases, two phases are streamlined into one study (e.g., I/II or II/III).

Other examples of terminology issues arise within the *Study Design* field, which might include allocation concealment (nonrandomized or randomized) control, endpoint classification, intervention model, masking or blinding, and who is blinded. Thus, in the case of RCTs, the trial registry data will not simply classify a study as an RCT but will also indicate if it is a parallel or crossover trial, which participants are blinded, whether the trial is one center or multicenter, and if the latter plans to recruit in one or several countries.

Data Quality
In order to ensure the quality of data entered, instructions in the form of guidelines or learning modules are needed. Registries are developing such instructions to help researchers achieve better quality of data submitted. For example, the Australian New Zealand Clinical Trial Registry developed "data item definition and explanation" [39]. International standards, the two countries' regulations, funders, and registries' policies all inform the content of this tool. Initial analysis of data entry in existing acceptable registries showed that a substantial amount of meaningless information was entered in open-ended text fields [40], but it has also shown improvement in this area over time [31, 41]. Finding the balance between general versus specific information is important. For example, indicating that the trial is blinded or double-blinded is much less informative than specifying who is blinded.

Many registrants will do only what is required, which is often determined by regulations, policies of funders, or simply recommended by WHO international standards and ICMJE instructions. The following is one potential look at levels of required data fields.

First-Level Fields First-level fields are required by the regulator. For example, ClinicalTrials.gov has fields that cannot be skipped because the FDAAA requires them; ISRCTN also has fields that cannot be skipped, which are aligned with the WHO international standards. While designing a registry, one should keep in mind the possibility of expansion and provide a few fields for such unexpected information.

Second-Level Fields Second-level fields are not made compulsory by some registries but are required by others. For example, because public funders or journal editors may require additional information beyond the international standards, there is an expectation that the relevant information will be provided by registrants; however, registries themselves cannot necessarily make these fields compulsory on their end, and consequently, some registries might not have these fields. Because adding fields to registries can sometimes be difficult, posting such additionally required information elsewhere in the registry is allowed. It may be placed along with or below other information or in the *Other* or *Additional information* field. For this reason, it is necessary to anticipate creation of such fields. For example, Canadian Institutes of Health Research (CIHR) requires the explicit reporting and public visibility of the ethics approval and confirmation of the systematic review justifying the trial.

Third-Level Fields Third-level fields are optional and contain information that might be suggested by the registry, research groups, or offered by the researcher as important for a given trial. Such third-level data are usually entered in the *Additional information* field. This variation in fields means that, although there are international standards, there are differences among registries, specifically in the number of fields and their elaboration. The current stage of trial registries might be considered the initial learning stage, and the analysis and evaluation of current practices will point to better policies and practices for the future.

Maintenance of Trial Registries

The researcher or sponsor of a trial provides annual updates of the trial record, and all of these updates should be displayed in the registry. These updates aim at capturing all amendments (i.e., changes of the protocol, the stage of trial implementation, eventual early stopping, etc.). It is important that these updates have dedicated fields and do not overwrite previous information. Such an approach enables the identification of changes and tracks the flow of the trial implementation. The registry can be designed so that a reminder is sent automatically to registrants so that they can obtain the annual update. As mentioned earlier, registries develop special mechanisms of deduplication within the registry and with other registries.

Results Databases

Traditionally the main vehicle to disseminate trial results and findings in a trustworthy way has been via publication in a peer-reviewed journal. Due to publication and outcome reporting bias and the availability of the Internet, there is a growing international discussion about Internet-based databases of summary results. Public disclosure of results in such databases will complement publication in peer-reviewed journals, and it is an integral part of the transparency tool set.

Theoretically results databases are complex, and they might include aggregate data, metadata, and analyzable data sets. Clinical trial databases in public domain are being developed by trial registries. Currently three registries developed them: ClinicalTrials.gov, European clinical trial registry, and the Japanese UMIN. Similarly, to trial registries, results databases are expected to build hyperlinks, the most important ones being between the given trial in the registry and related publications or systematic reviews and meta-analysis. As of 2018, results databases and repositories are far less developed than trial registries. As identified by the international meeting of the Public Reporting Of Clinical Trials Outcomes and Results (PROCTOR) group in 2008 [42], and discussed later on by us [10] especially in the IMPACT Observatory [43], and by others [44], there are numerous issues to be resolved in order to get the results data, especially microlevel data sets, publicly disclosed.

Standards

There are no international standards for public disclosure of trial results, and there are no standards for preparing and use of the analyzable data sets, based on cleaned, anonymized individual participant data (IPD) and adjacent needed documentation (metadata, dictionary, etc.). However, there is much discussion on how these should be designed, and some initiatives have been contributing to accumulation of experience [28, 42, 45]. In 2010, the journal *Trials* started posting them on the Internet as the series "Sharing clinical research data," edited by Andrew Vickers. The topic of results disclosure actually includes a spectrum of information from aggregate (summary) data to fully analyzable, i.e., IPD-based data sets. In 2017, following several years of consensus building process that involved participants from various areas and backgrounds, the ECRIN leg of the CORBEL project developed a set of recommendations regarding clinical trial data sharing [44]. Of note, clinical trial registries generally only enable the public disclosure of summary data and findings of clinical trials many of which are also published in peer-reviewed journals, while the IPD-based analyzable data sets are published in repositories.

Some of the outstanding challenges and disclosure issues regarding summary results and analyzable data are comparable to those of trial registries. These include the need to develop international standards, quality and completeness of data, timing of reporting, and standardization of terms. Other issues are more specific to the practical details of public disclosure of analyzable data sets. Those include the cleaning of data, quality of data, accountability, defining which adjacent documentation is needed, who is the guarantor of truth, privacy issues/anonymization, intellectual property rights, and issues related to anonymization efforts [46].

Many of these issues suggest a need to develop levels of detail related to levels of access. In the era of electronic data management, some of these steps, such as cleaning of raw data, are becoming less of an issue as they take place simultaneously with the data collection. Much can be learned from other areas especially from the experience of genome data sharing, for which many have shown that data sharing has boosted the development of the field [47, 48].

A lot has changed since the first version of this chapter published in 2012 [11], when these data were either protected in the hands of regulators or might have been shared with systematic reviewers only upon request and only under certain conditions. Meanwhile many constituencies engaged in making data available, especially in order to facilitate systematic reviews that include of IPD data sets (meta-analyses). For example, journal editors are increasingly encouraging data sharing upon publication of trial findings in their respective journals [49].

Data sharing is becoming more and more appealing to all stakeholders [50–53]. Earlier hesitation has been gradually lightening, and we are witnessing increased transparency and a consecutive change of the research paradigm. Although many issues have yet to be resolved, this area is constantly and rapidly evolving, and by the

time this book is printed, there will likely be more progress. However, several dilemmas and issues are still present and will require research and resolution. These include the lack of standards on how to prepare data sets for public sharing, heterogeneity of repositories, and finding the balance of privacy versus transparency [43]. All of these elements create specific challenges, require interdisciplinary work, and present an opportunity for clinical research informatics and information technology experts.

Repositories

Repositories, i.e., research data repositories, are electronic databases hosting research raw data and facilitating their reuse. They are the newest research transparency tool complementing trial registries and results databases.

As mentioned earlier when talking about data sharing from clinical trials, we are talking about the cleaned anonymized individual participant data (IPD) sets and adjacent documentation forming the analyzable data.

Repositories can be classified by the scientific area they cover or the level (university, region, country, international) at which they are organized. Re3data [54] (described below) classifies them into disciplinary, institutional, and other. Some of repositories hosting clinical trial data are based at universities and accept data only from researchers from a given university or consortium, such as Edinburgh DataShare or DRUM (Data Repository for the University of Minnesota). Figshare, on the other hand, accepts data from anywhere. Dryad accepts data if the research is published. Most general open-access repositories in public domain host data from any research. Their number is growing, and as of June 2018, there were 2109 repositories registered in re3data. However, only a small portion of them host clinical trial data. In our ongoing study we identified about a dozen general open access repositories in public domain that also host clinical trial data and analyzed their basic features [43, 55–57]. However, besides general research data repositories, there are also disease-specific repositories and research data repositories organized by funders, such as several repositories run by the NIH institutes.

With the exception of the Japanese register UMIN [58] that hosts clinical trial data of trials that are already registered in it, there is currently no domain repository in public domain, i.e., repository devoted to hosting exclusively clinical trial data.

It is important to note that the data management should begin at data collection, and public funders are increasingly demanding that the data management plan be developed up front. This leads to the understanding that the data preservation and storage of academic trials starts at the academia, that the institution – academia conducting a trial should anticipate data sharing and act accordingly – preferably develop a database and then might send data to established repositories. Indeed, several universities have been doing this. One of the first was the Edinburgh University that established Edinburgh DataShare repository which also hosts clinical trial data. It started with a JISK project led by

Edinburgh University in partnership with two other UK universities (Oxford and Southampton). While it initially hosted data from the international stoke trial, it is now hosting data from other studies conducted at the Edinburgh University [59]. The key role in setting and running of this repository has been played by research librarians. Actually, management and storage of research data have become a field of interest of research librarians, and they are increasingly engaged in this field.

Some repositories hosting clinical trial data might limit the uploading of data to members of a given university or consortium, but all of them enable open access to data for secondary use. There is a limited control of data quality at entry and no curatorship of data already in the repository. Basically, repositories rely on the clinician trialist – data provider to clean, anonymize, and organize data for publication.

Several specific projects and software have been influencing development of this field. One of them is Dataverse, which is an open-source web application to share, preserve, cite, explore, and analyze research data [60]. A Dataverse repository is the software installation, which then hosts multiple virtual archives called Dataverses. Each Dataverse contains data sets, and each data set contains descriptive metadata and data files (including documentation and code that accompany the data). As an organizing method, Dataverses may also contain other Dataverses. There are 33 Dataverse repositories (installations) around the word, and one of them, Harvard Dataverse, also hosts CT data [61].

It is important to point to the Research Data Alliance (RDA) which aims at building the social and technical infrastructure to enable open sharing of data. It functions through interest and working groups that elaborate specific topics and provide recommendations for the community [62].

Few related tools to data sharing by repositories include persistent identifiers/PID, DataCite, re3data, and the CoreTrustSeal of certification organization [63].

Re3data is a registry of research data repositories from various academic disciplines. In 2014 it merged with another similar tool, Databib, and it is now managed by DataCite. Re3data registers repositories from various disciplines and describes basic features of each of them. "It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. re3data.org promotes a culture of sharing, increased access and better visibility of research data. The registry went live in autumn 2012 and is funded by the German Research Foundation (DFG)" [54, 64].

Citability and findability of published data are very important. Among other benefits, they stimulate public data sharing. Citability and to certain extend findability are achieved by assigning the *persistent identifier (PI or PID)* to published data sets [43]. PID is a long-lasting reference to a document, file, web page, or other object. The term "persistent identifier" is usually used in the context of digital objects that are accessible over the Internet. Once plugged in the web browser, it will link to related data sets which enables citation of given

data sets [65]. Persistent identifiers help the research community locate, identify, and cite research data with confidence.

DataCite is a leading global nonprofit organization that provides persistent identifiers (DOIs) for research data [66]. DataCite assigns DOI persistent identifier to each repository registered in re3data. Repositories in turn assign persistent identifier to hosted data sets, i.e., data sets published in them. In our ongoing scanning of general repositories within the IMPACT Observatory we noticed that most of the open access general repositories in public domain that host clinical trial data assign DOI, or some other PID [57].

The research community realized the importance of ensuring the quality of repositories, and in 2017, the *CoreTrustSeal* certification organization was established, developed by *the ICSU World Data System (WDS) and the Data Seal of Approval (DSA)* under the umbrella of RDA. The CoreTrustSeal has a set of criteria that a given repository has to meet [63]. *The* re3data indicates for each indexed repository whether it is certified or whether it supports repository standards.

The User Perspective

Some of repositories that host clinical trial data are open for hosting of data from certain groups of researchers, usually those linked to a given university, or area, but all of them allow open access to data they host. The lack of standards and heterogeneity of repositories makes the analysis of hosted data across several repositories very difficult if not impossible, without contacting the original data provider. It can be expected that the interest and the need for reanalysis will trigger development of needed standards. Such standards should be developed by the research community, not by repository. Ideally, internationally renowned organizations, such as WHO, will lead standard development and include key stakeholders in the consensus building process, as was the case with development of the trial registration standards.

Summary and Future

The future of clinical research and informatics is closely interwoven, and it can be expected that these evolving fields will mutually inform and influence each other. Clinical trial transparency and especially sharing of analyzable data sets are lagging behind most other research areas. There are barriers to overcome, some of which are specific for clinical trials, and they will probably continue presenting exciting challenges for researchers, information technology (IT) experts, and in fact all interested to further existing tools and figure out the sustainable strategies for public disclosure of trial information – from protocol via results to data, including the stewardship and reuse of such data in knowledge creation which will in turn speed development of new and more powerful diagnostics and therapeutics.
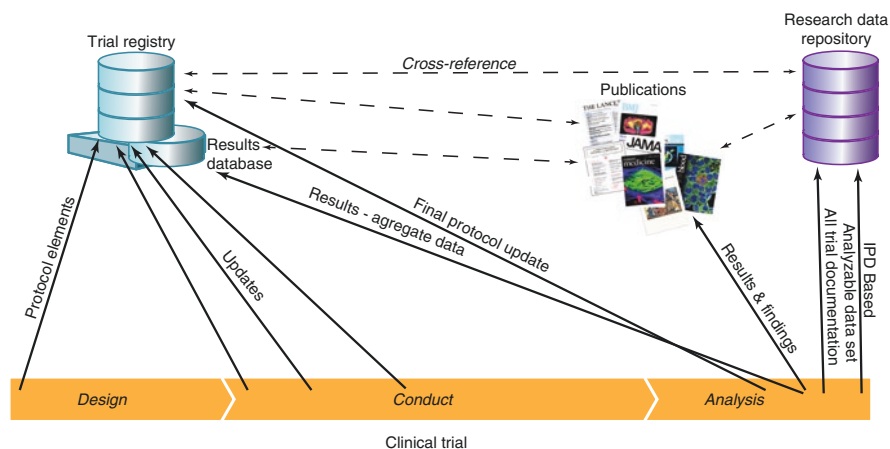
Fig. 21.3 Anticipated flow of data from clinical trial to public domain. Please note that while all parts of the data flow have evolved since 2012, the major change of this flow of data took place by the establishment of open-access research data repositories in public domain

It is anticipated that data flow from trials to the public domain and the linking and cross-referencing of related data will create a more efficient system of information sharing and knowledge creation (Fig. 21.3). Although it has not yet been completely accomplished, there is a clear tendency to move in that direction, which will ensure a high level of transparency, getting closer to open data and open science.

Furthermore, it is expected that existing systematic reviews will be updated with the meta-analysis of IPD-based analyzable data to inform various levels of decision-making with the updated evidence. Finally, in an ongoing effort to increase transparency of research and to build on the experience of trial registries, other types of studies are being registered in trial registries, and other types of research registries are being developed. However, although there are no standards and guidelines for the preparation of clinical trial data for public release and although repositories are heterogenous, the existence of open-access repositories is a big step forward toward opening of clinical trial data.

Trial registries host defined protocol items, and they are in constant evolution, from the elaboration of fields to the establishment of hyperlinks. It can be expected that the analysis and evaluation of the existing primary registries' experience will inform the best practice and potential expansion of the data included, like adding fields to host more data than required by the initial 20-item international standards. This has already taking place, and, for example, WHO recently revised standards (version 1.3.1.) include four more protocol items: ethics review, completion date, summary results, and IPD sharing plan [23].

Furthermore, there is a strong push for publication of the full protocol, either in the registry or elsewhere. It will certainly be particularly useful to have publicly available electronic versions of structured protocols, following SPIRIT guidelines. If this were to happen, the protocol data set that is available in registries will

continue to provide valuable summaries of protocols with links to other trial related information including the full protocol, publications, trial website, systematic review, meta-analysis, results databases and research data repositories and thus continue to play an important role in achieving trial transparency.

Results databases are in their early stage of development, and they currently lack international standards. They are being formed by trial registries and aim at providing summary/aggregate results data of registered trials in predefined tables. Out of 17 general open-access registries in public domain that are linked to the WHO, only 3 developed summary clinical trial results databases: ClinicalTrials.gov, EU CRT (European Clinical Trial Register, https://www.clinicaltrialsregister.eu/ and Japanese registry, UMIN. As mentioned earlier, UMIN also displays IPDs. These databases differ. Each of them follows the rules of their respective countries, and at the same time, they are meeting the WHO and ICMJE request to register and share summary results. Apparently, the need to synchronize has been understood, and it seems that ClinicalTrials.gov and EMA/European Clinical Trial Registry are working on developing comparable data fields which might inform future development of international standards of data sharing.

Open-access research data repositories in public domain are certainly the most important tool for data opening and can play a major role in enabling public availability of research data. However, they are heterogenous, and there are still no international standards to govern the public disclosure of analyzable data sets which include cleaned, anonymized IPDs (i.e., usually numeric or encoded) and documentation sufficient to make the data reusable.

Development of such standards will require participation of all interested constituencies in thorough planning, analysis of quality control, resources, as well as dealing with specific issues, such as privacy, i.e., anonymization methods and practices. It is important to note that although there are no standards and guidelines for the preparation of clinical trial data for public release and although repositories are heterogenous, the existence of open-access repositories and a possibility to publish data in them are a big step forward toward opening of clinical trial data.

The progress achieved as well as the interest and expectations this data opening process has created so far is encouraging, but still a lot needs to be done. As mentioned earlier, there are numerous initiatives contributing to increasing the transparency of clinical trials and opening of its data beyond described in this chapter. There are also initiatives and projects addressing the needed standards development as mentioned CORBEL project [44]. It can be expected that this process will be observed and supported in various ways by key players at various levels, including regulators, public funders, clinicians, academia, pharmacists, journal editors, industry, patients, consumers, consumer advocates, and general public. Thus, researchers and IT experts will not be alone in this process as the clinical trials and their contribution to creation of the evidence needed for decisions in health are of paramount interests to numerous stakeholders.

The dynamics of the process are so immense and complex that they merit assessment of actions, initiatives, and practice of various players and their interactions. It is equally important to assess the impact of these dynamics on opening of

analyzable data for reuse, on the consequent transformation of clinical trial research all adjacent issues. An observatory or natural experiment is the methodology of choice to collect, assess, and disseminate such data and thus inform the process and indicate trends. The IMPACT Observatory aims to do just that and become a tool, a hub, informing the process of opening of trial data [43].

Disclaimer The views expressed here are the author's and do not represent the views of the MedILS or any other organization.

References

1. Vickers AJ. Sharing raw data from clinical trials: what progress since we first asked "Whose data set is it anyway?" Trials [Internet]. 2016;17:227. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4855346/.
2. Krleža-Jerić K. Clinical trial registration: the differing views of industry, the WHO, and the Ottawa Group. PLoS Med. 2005;2:1093–7.
3. Clinical Study Data Request [Internet]. [cited 2016 Sep 1]. Available from: https://www.clinicalstudydatarequest.com/.
4. YODA Project [Internet]. [cited 2016 Jul 19]. Available from: http://yoda.yale.edu/.
5. Project Data Sphere [Internet]. [cited 2016 Aug 1]. Available from: https://www.projectdatasphere.org/.
6. Vivli-A global clinical trial data sharing platform: proposal, definition and scope background and objectives; 2016 June.
7. The Vivli Platform is live – Vivli [Internet]. [cited 2018 Aug 23]. Available from: https://vivli.org/news/the-vivli-platform-is-live/.
8. European Medicines Agency – clinical data publication – documents from advisory groups on clinical-trial data [Internet]. [cited 2018 Aug 21]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/document_listing/document_listing_000368.jsp&mid=WC0b01ac05809f3f12#section1.
9. Krleža-Jerić K. Sharing of data from clinical trials and research integrity. In: Steneck N, Anderson M, Kleinert S, Mayer T, editors. Integrity in the Global Research Arena; Proceedings of the World Conference on Research Integrity. 3rd ed. Montreal, Quebec, Singapore: World Scientific Publishing; 2013. p. 91.
10. Krleža-Jerić K. Sharing of clinical trial data and research integrity. Period Biol. 2014;116(4):337–9.
11. Krleža-Jerić K. Clinical trials registries and results databases. In: Ritchesson RL, Andrews JE, editors. Clinical research informatics. London: Springer; 2012. p. 389–408.
12. Bass A. Side effects; a prosecutor, whistelblower, and a bestselling antidepressant on trial. 1st ed. Chapel Hill: Algonquin Books; 2008. 260 p.
13. Gibson L. GlaxoSmithKline to publish clinical trials after US lawsuit. BMJ. 2004;328(7455):1513.
14. DeAngelis CD, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the international committee of medical journal editors. JAMA [Internet]. 2004 [cited 2016 Jul 13];292(11):1363–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15355936.
15. Krleža-Jerić K, Chan A-W, Dickersin K, Sim I, Grimshaw J, Gluud C. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). BMJ [Internet]. 2005 [cited 2016 Jul 14];330(7497):956–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15845980.

16. The Mexico Statement on Health Research [Internet]. Mexico City; Available from: http://www.who.int/rpc/summit/agenda/Mexico_Statement-English.pdf.

17. International Clinical Trials Registry Platform (ICRTP) [Internet]. [cited 2016 Jul 12]. Available from: http://www.who.int/ictrp/en.

18. Mahmić-Kaknjo M, Šimić J, Krleža-Jerić K. Setting the impact (improve access to clinical trial data) observatory baseline. Biochem Med. 2018;28(1):7–15. 010201. https://doi.org/10.11613/BM.2018.010201.

19. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med. 2013;158(3):200–7.

20. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ [Internet]. 2013;346:e7586. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3541470&tool=pmcentrez&rendertype=abstract.

21. The SPIRIT Statement [Internet]. [cited 2018 Aug 17]. Available from: http://www.spirit-statement.org/spirit-statement/.

22. El Emam K, Jonker E, Sampson M, Krleža-Jerić K, Neisa A. The use of electronic data capture tools in clinical trials: web-survey of 259 Canadian trials. J Med Internet Res [Internet]. 2009;11(1):e8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2762772&tool=pmcentrez&rendertype=abstract.

23. WHO Trial registration Data set version 1.3.1. [Internet]. [cited 2018 Aug 19]. Available from: http://www.who.int/ictrp/network/trds/en/.

24. Reveiz L, Chan A-W, Krleža-Jerić K, Granados CE, Pinart M, Etxeandia I, et al. Reporting of methodologic information on trial registries for quality assessment: a study of trial records retrieved from the WHO search portal. PLoS One [Internet]. 2010;5(8):e12484. Available from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0012484.

25. WHO|The WHO Registry Network [Internet]. WHO. World Health Organization; 2016 [cited 2018 Aug 17]. Available from: http://www.who.int/ictrp/network/en/.

26. Reveiz L, Krleža-Jerić K, Chan A-W, de Aguiar S. Do trialists endorse clinical trial registration? Survey of a Pubmed sample. Trials [Internet]. 2007;8:30. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2147029&tool=pmcentrez&rendertype=abstract.

27. WMA Declaration of Helsinki – ethical principles for medical research involving human subjects – WMA – The World Medical Association [Internet]. 2013 [cited 2018 Aug 20]. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.

28. Krleža-Jerić K, Lemmens T. 7th revision of the Declaration of Helsinki: good news for the transparency of clinical trials. Croat Med J [Internet]. 2009 [cited 2016 Jul 14];50(2):105–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19399942.

29. Goodyear MDE, Krleza-Jeric K, Lemmens T. The Declaration of Helsinki. BMJ [Internet]. 2007 [cited 2016 Jul 14];335(7621):624–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17901471.

30. Krleža-Jerić K, Lemmens T, Reveiz L, Cuervo LG, Bero LA. Prospective registration and results disclosure of clinical trials in the Americas: a roadmap toward transparency. Rev Panam Salud Publica [Internet]. 2011 [cited 2016 Jun 16];30(1):87–96. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22159656.

31. Zarin DA, Tse T, Williams RJ, Rajakannan T. The status of trial registration eleven years after the ICMJE policy. N Engl J Med [Internet]. 2017;376(4):383–91. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5813248/.

32. Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the food and drug administration: review of publication and presentation. Ioannidis J, editor. PLoS Med [Internet]. 2008;5(11):e217. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2586350/.

33. CDISC [Internet]. [cited 2018 Aug 20]. Available from: https://www.cdisc.org/.

34. Harriman SL, Patel J. When are clinical trials registered? An analysis of prospective versus retrospective registration. Trials [Internet]. 2016;17:187. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4832501/.

35. Viergever RF, Karam G, Reis A, Ghersi D. The quality of registration of clinical trials: still a problem. Scherer RW, editor. PLoS One [Internet]. 2014;9(1):e84727. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3888400/.

36. Food and Drug Administration Amendments Act (FDAAA) of 2007 [Internet]. Office of the Commissioner; [cited 2018 Aug 22]. Available from: https://www.fda.gov/regulatoryinformation/lawsenforcedbyfda/significantamendmentstothefdcact/foodanddrugadministration-amendmentsactof2007/default.htm.

37. Prospero-International prospective register of systematic reviews [Internet]. [cited 2018 Aug 21]. Available from: https://www.crd.york.ac.uk/prospero/.

38. Australia Clinical Trials Toolkit|Australian Clinical Trials [Internet]. [cited 2018 Aug 22]. Available from: https://www.australianclinicaltrials.gov.au/clinical-trials-toolkit#overlay-context=home.

39. Australia New Zealand Clinical Trials Registry. Data item definition/explanation [Internet]. [cited 2018 Aug 14]. Available from: http://www.anzctr.org.au/docs/ANZCTR%20Data%20field%20explanation.pdf.

40. Zarin DA, Tse T, Ide NC. Trial registration at ClinicalTrials.gov between May and October 2005. N Engl J Med [Internet]. 2005;353(26):2779–87. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1568386/.

41. Askie LM, Hunter KE, Berber S, Langford A, Tan-Koay AG, Vu T, Sausa R, Seidler AL, Ko H SR. The clinical trials landscape in Australia 2006–2015 [Internet]. Sydney: Australian New Zealand Clinical Trials Registry; 2017 [cited 2018 Aug 19]. 83 p. Available from: http://www.anzctr.org.au/docs/ClinicalTrialsInAustralia2006-2015.pdf#page=1&zoom=auto,557,766.

42. Krleža-Jerić K. International dialogue on the public reporting of clinical trial outcome and results – PROCTOR meeting. Croat Med J. 2008;49:267–8.

43. Krleža-Jerić K, Gabelica M, Banzi R, Martinić MK, Pulido B, Mahmić-Kaknjo M, et al. IMPACT Observatory: tracking the evolution of clinical trial data sharing and research integrity. Biochem Medica [Internet]. 2016;26(3):308–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27812300%0A, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5082220.

44. Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open [Internet]. 2017;7(12):e018647. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5736032/.

45. Bian Z-X, Wu T-X. Legislation for trial registration and data transparency. Trials [Internet]. 2010;11(1):64. Available from: https://doi.org/10.1186/1745-6215-11-64.

46. El Emam K, Rodgers S, Malin B. Anonymising and sharing individual patient data. BMJ [Internet]. 2015;350:h1139. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707567/.

47. Collins F. Has the revolution arrived? Nature [Internet]. 2010;464(7289):674–5. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5101928/.

48. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature [Internet]. 2003;422:835. Available from: https://doi.org/10.1038/nature01626.

49. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. PLoS Med [Internet]. 2017;14(6):e1002315. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5459581/.

50. Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. Trials [Internet]. 2011 [cited 2018 Aug 20];12(1):249. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22112900.

51. Zarin DA, Tse T. Sharing Individual Participant Data (IPD) within the Context of the Trial Reporting System (TRS). PLoS Med [Internet]. 2016;13(1):e1001946. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4718525/.

52. Rockhold F, Nisen P, Freeman A. Data sharing at a crossroads. N Engl J Med [Internet]. 2016 [cited 2018 Aug 13];375(12):1115–7. Available from: http://www.nejm.org/doi/10.1056/NEJMp1608086.

53. Eichler H-G, Abadie E, Breckenridge A, Leufkens H, Rasi G, Doshi P, et al. Open clinical trial data for all? A view from regulators. PLoS Med [Internet]. 2012 [cited 2016 Jul 14];9(4):e1001202. Available from: http://dx.plos.org/10.1371/journal.pmed.1001202.

54. Re3Data; Registry of Research Data Repositories [Internet]. [cited 2018 Aug 12]. Available from: www.Re3data.org.

55. Krleza-Jeric K, Hrynaszkiewicz I. Environmental Scan of Repositories of Clinical Research Data: How Far Have We Got With Public Disclosure of Trial Data? [Internet]. figshare; 2018. Available from: https://figshare.com/articles/Environmental_Scan_of_Repositories_of_Clinical_Research_Data_How_Far_Have_We_Got_With_Public_Disclosure_of_Trial_Data_/5755386.

56. Krleza-Jeric K, Gabelica M, Mahmic-Kaknjo M, Malicki M, Utrobicic A, Simic J, et al. Setting of an Observatory of clinical trial transition regarding data sharing; IMPACT Observatory. Poster, Cochrane Colloquium Vienna, 2015. Available from: https://figshare.com/articles/Setting_of_an_Observatory_of_clinical_trial_transition_regarding_data_sharing_IMPACT_Observatory/5753226.

57. Gabelica M, Martinic MK, Luksic D, Krleza-Jeric K. Clinical trial transparency and data repositories; an environmental scan of the IMPACT (Improving Access to Clinical Trial Data) Observatory. Poster, 8th Croatian Cochrane Symposium, Split. 2016. https://doi.org/10.6084/m9.figshare.7390559.v1. Available from: https://figshare.com/articles/Clinical_trial_transparency_and_data_repositories_an_environmental_scan_of_the_IMPACT_Improving_Access_to_Clinical_Trial_Data_Observatory/7390559.

58. UMIN-ICDR Individual Case data repository [Internet]. [cited 2018 Aug 12]. Available from: http://www.umin.ac.jp/icdr/index.html.

59. Edinburgh DataShare [Internet]. [cited 2018 Aug 19]. Available from: https://datashare.is.ed.ac.uk/.

60. The Dataverse Project [Internet]. [cited 2018 Aug 19]. Available from: https://dataverse.org/.

61. Harvard Dataverse [Internet]. [cited 2018 Aug 20]. Available from: https://dataverse.harvard.edu/.

62. Research Data Alliance RDA [Internet]. [cited 2018 Aug 17]. Available from: https://www.rd-alliance.org/about-rda/who-rda.htmlNo.Title.

63. CoreTrustSeal [Internet]. [cited 2018 Jun 28]. Available from: https://www.coretrustseal.org/about/.

64. Pampel H, Vierkant P, Scholze F, Bertelmann R, Kindling M, Klump J, et al. Making research data repositories visible: the re3data.org registry. PLoS One. 2013;8(11):e78080.

65. Persistent Identifier [Internet]. [cited 2018 Aug 19]. Available from: https://en.wikipedia.org/wiki/Persistent_identifier.

66. DataCite [Internet]. [cited 2018 Aug 12]. Available from: https://www.datacite.org/index.html.