synchrogenix

A CERTARA COMPANY

Synchrogenix ClinGenuity Redaction Management Solution (CRMS)

Artificial Intelligence for Transparency and Disclosure Initiatives

Transparency and disclosure in clinical trials

The issue of transparency and disclosure of clinical trial data has been growing in importance over the past few years. Clinical trials are essential to offering new therapies to patients. However, the individuals that participate in those trials put themselves at risk, and it is up to the medical community to derive as much benefit from that risk as possible. Disclosing clinical trials information and creating transparency around the data are a key step toward maximizing that benefit, increasing trust between the public and the industry, providing data for further work by additional researchers that could spur new products or therapeutic approaches, and potentially avoiding unnecessary trials. However, data sharing carries its own risk, especially with regard to protected personal data (PPD; such as subject or patient ID numbers, event profiles, etc) and company confidential information (CCI; such as trade secrets and protected intellectual property not intended for public consumption).

> "Greater data sharing could enhance public well-being by accelerating the drug discovery and development process, reducing redundant research, and facilitating scientific innovation."
>
> – Institute of Medicine

On January 14, 2015, the Institute of Medicine (IOM) released its report, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*, which outlines principles and a practical framework for the responsible sharing of clinical trial data. In the report, the IOM committee identified the optimal stage in the clinical trial lifecycle at which each data type should be shared and under what conditions. The committee recommends that all data should be published no more than 18 months after completion of a trial; however, if the trial is part of a submission to a regulatory agency for approval, the data should be shared no later than 30 days after that approval or 18 months after product abandonment. When trial findings are published before the 18-month period has passed, the committee recommends that the supporting analytic dataset be shared within six months of publication.

Regulatory drivers

The European Medicines Agency (EMA) Policy 43, *Access to Documents*, adopted on December 1, 2010 provides a process by which individuals can request access to all EMA business related documents unless the document falls into an exception listed in Article 4 of Regulation (EC) 1049/2001. Once a request has been made, the EMA approves or rejects the request. If the request is approved, the EMA reviews the requested documents for PPD and CCI and redacts this information if found. If the document was created by a sponsor, the EMA may consult with the sponsor to allow them to provide feedback on what information was or wasn't redacted.

These consultations can come with very short notice and short turn arounds, requiring sponsors to have to react quickly with a review of the proposed redacted documents from EMA. The sponsor must compile a justification table outlining any requested changes to the EMA's proposed redacted document to include a rationale for each requested change.

The EMA Policy 70, *Publication of Clinical Data for Medicinal Products for Human Use*, adopted on January 1, 2015, set the stage for a new era in transparency and disclosure. Under Phase I of this new regulation, the EMA will make available clinical reports to include clinical summaries, clinical overviews, clinical study reports (to include the report body, the protocol and amendments, the sample case report form and documentation of statistical methods) that support an approved marketing authorization for submissions made after January 1, 2015. The EMA will make available applications for additional drug indications or line extensions of indications for applications submitted after July 1, 2015. In order to prepare these documents to be made publically available, the reports must first be anonymized by the sponsor. Redacting, referred to as masking by EMA, is an acceptable form of anonymization. However, redaction limits the data utility of the information and so the EMA is recommending two additional anonymization options: randomization (eg, noise addition and permutation) and generalization (eg, aggregation and k-anonymity). Randomization and generalization both require incorporation at the time the report is being written or retrospectively changing the data and in essence producing a second report. As a result of the cost to create a second report, most sponsors are considering redaction as the preferred method for Policy 70 submissions in the short term.

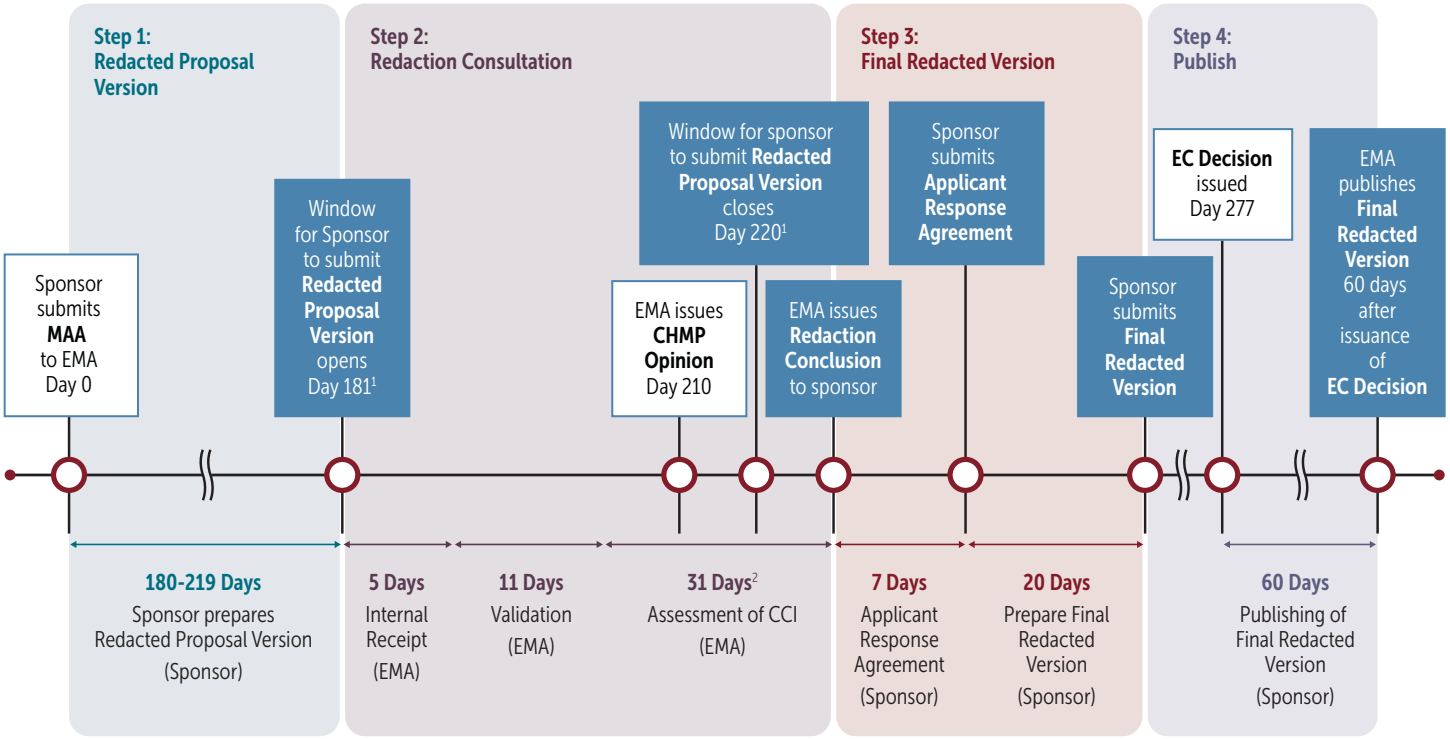
As detailed in EMA's newly published implementation guidance, *External guidance on the implantation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use*, sponsors will be required to submit two packages to the EMA. The first package will contain the required submission documents with proposed redactions identified, but not applied, a justification table that lists all CCI redactions and anonymization report that will outline the rules by which the sponsor anonymized the report. The second package will contain the final redacted version of the required submission documents. The second package will be the version made publically available. Once the first package is submitted, the EMA will assess the CCI redactions and review the sponsor's anonymization report outlining the rules by which they anonymized the documents for PPD. The EMA will utilize the justification table to communicate which CCI redactions they reject or request to be altered and will communicate back to the sponsors if any rules defined in the anonymization report are not found to be acceptable. Once the sponsor receives the feedback, they must make the requested changes and submit the second and final package.

For Phase II of Policy 70, the EMA plans to also make individual patient data available, after assurances that the privacy of patients is adequately protected before their data are released. Policy 70 is the first key step toward a new EU Clinical Trials Regulation expected in 2016. The US Food and Drug Administration (FDA), along with the National Institutes of Health, are also addressing the need for improved transparency and disclosure. The 2012 FDA Safety and Innovation Act requires this issue to be addressed.

Industry perspective

The pharmaceutical industry's largest trade organizations, Pharmaceutical Research and Manufacturers of America in the US and European Federation of Pharmaceutical Industries and Associations in the EU, are committed to *Principles for Responsible Clinical Trial Data Sharing*. Under these principles, companies agree to enhance public access to clinical study information for any submission following a January 1, 2014, US or EU approval. At a minimum, companies agree to provide the synopses of CSRs to participating patients. Companies have also agreed to provide the full CSR to researchers/investigators, as requested for legitimate uses, as they do for requests to access patient-level data, study-level data, and protocols.

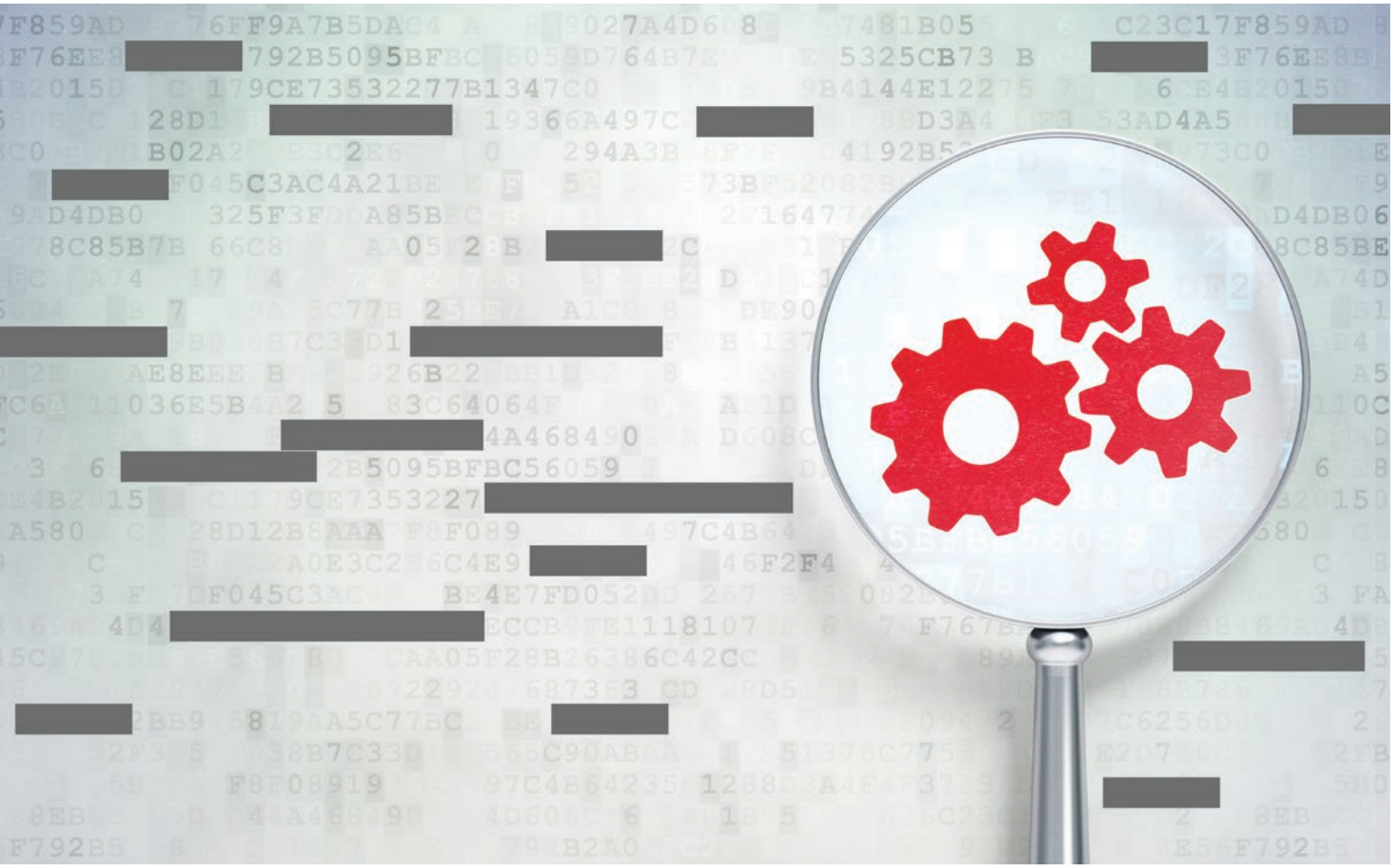
In late 2014, the industry group TransCelerate BioPharma published its *Clinical Study Reports Approach to Protection of Personal Data*, which outlined its commitment to enhancing public health and medical and scientific knowledge through the sharing and transparency of clinical trial information. The TransCelerate members have agreed that privacy concerns are paramount in the context of public disclosure of Clinical Reports and have set very specific guidelines for the members of their organization, including the requirement for full redaction of private information.



AI technology to meet transparency and disclosure demands

Clinical trial data are contained within regulatory documents, such as CSRs, marketing application submission documents (New Drug Applications [NDAs], Marketing Authorization Applications [MAAs], Biologics License Applications [BLAs], etc), and others.

Prior to complying with transparency and disclosure initiatives and providing these documents to the public, it is necessary to first accurately identify and securely redact information that could lead to the identification of an individual from these regulatory documents. The two critical categories of sensitive information (PPD and CCI) must be identified and then either redacted or removed prior to releasing the documents to the public.



Given the EMA's acknowledgement in the implementation guidance that masking is the most likely technique to be used by sponsors in the short term, there are essentially three options for pharmaceutical companies to choose from for immediate transparency compliance: 1) fully manual process; 2) manual process aided by minimal technology; 3) artificial intelligence (AI)-enabled automated redaction.

1. Manual. The manual process involves reading through regulatory documents and manually redacting sensitive information either with black marker or using Adobe® Pro redaction tools.

Issue: High risk of accidental disclosure

CSRs and other submission documents can be up to 50,000 pages in length. On average, there are 600 words per page or approximately 30 million words per CSR. Even a small, standard transparency initiative would include roughly 25 documents per year. All 30 million words in each document must be reviewed to identify, on a case-by-case basis, sensitive information, as defined previously. On an annual basis, that would require reviewing 750 million words accurately. This is not a logical option and has proven to be wrought with error.

Issue: Highly resource dependent
A substantial resource base is needed, both within the client for direct oversight and also within the vendor, to perform the manual functions.

Issue: Variability in deployment
Multiple resources are not likely to do the same task exactly the same as one another. As such, redaction in this scenario is highly variable and not consistent.

Issue: Not a durable solution long-term
Because the process is entirely manual, it is likely that there will be high turnover in staff internally and externally, requiring constant modification to the process and continued, dedicated, internal oversight.

Issue: Slow and therefore high cost
Manually, each CSR can take 40 hours to redact by hand.

2. Manual with minimal technology. Several companies have attempted to approach this problem by assisting the manual process with rudimentary technology "tools". These tools almost always consist of Adobe® Pro redaction tools with minor modifications. At best, they incorporate a pattern-matching, regular expression tool. They also tend to promote the ability to search and redact. However, even with these "tools", the overall approach is still the same as the fully manual process. Pattern matching and/or search functionality without AI essentially reduces the tool to a manual process. For pattern matching and search functions to work effectively, the individual driving the process would need to pre-emptively know every possible name, phone number, email address, subject ID, event term, etc, to enable pattern matching or search functions to accurately locate and redact this information. For example, with a search and redact mechanism, an individual user begins by entering the term for the tool to search for and then redact consistently throughout the document. Even the manual process assisted with rudimentary tools is still essentially a fully manual process and prone to the same pitfalls.

Issues: The same as mentioned for the manual process.

3. Automated Redaction Using AI. Synchrogenix's unique AI engine, has created an engine that is built on natural language processing and recognition. As such, the engine is able to identify individual words, parts of speech, word combinations, and phrasing combinations automatically to determine context. Because of the power behind our unique AI, the engine can be configured to identify PPD and CCI the same way an individual would be trained on these definitions. The process of identifying and redacting sensitive information becomes automated and significantly more accurate. Synchrogenix's Redaction Management Solution (CRMS), is not limited to volume constraints as described in the manual option and can be deployed as a durable, long-term solution without the need for significant oversight by the client.

How it works

Redaction Requirements

Synchrogenix works with sponsors to define the criteria for redaction, including reviewing transparency legislation and company confidentiality policies to determine what information is to be retained and what is to be redacted. Synchrogenix has developed a base line set of rules from its vast client base experience as well as its familiarity with the regulatory policies. This can be provided to a client as the initial draft to significantly shorten the rule definition timeframe during the CRMS

onboarding process. We work with each sponsor to prepare a formal redaction requirements document that defines precisely what constitutes company-specific sensitive information to be redacted. Not surprisingly, redaction requirements can differ significantly from sponsor to sponsor. This is why we configure our AI engine specifically for each client.

Configuration

The CRMS is configured on a customer-by-customer basis. First, we configure the system based on your individual redaction requirements. A customized workflow is designed to meet sponsors' individual processes. Once the workflow is approved, we work in unison with sponsors on implementation strategies that fit their specific needs and goals.

How the Engine Works

The CRMS is built on a comprehensive AI engine that requires statistical models to determine the data to be redacted. Essentially, the system acts like a human would when you teach them the process. You provide the system general rules, as identified in your redaction requirements document. The system is then trained on your specific redaction requirements and learns specifically what needs to be redacted. Far different than human redaction, the CRMS never needs to be retrained or replaced. Additionally, the CRMS is not limited in its capabilities and is able to redact hundreds of studies per day, consistently, accurately, and efficiently.

Quality Control

The CRMS does not stop there. Our team of experts performs a thorough quality control (QC) review of each and every document, ensuring the highest level of accuracy. Unlike other vendors, our resources are domain-specific experts. Any findings, including resolution, are documented on a QC checklist form either provided by us or the sponsor. The QC review findings are seamlessly integrated into the system and used in refining the accuracy of the tool in real time.

Publishing

Sponsors publish their clinical reports according to the Electronicc Common Technical Document (eCTD) guidelines. As a result, the archived version of the clinical reports are often comprised of modular, individual components. Once the documents are fully redacted and redaction review completed, the system republishes the appropriate documents that are ready to post to your transparency website.

Conversely, some clients maintain an archived version of their clinical reports. For Policy 70, EMA is requiring that Sponsors upload their redaction documents for the first and second packages in the eCTD modular format. If the sponsor provides an uploaded super PDF, our system can publish the document back out into modular appendices.

Value Proposition

The risk to any pharmaceutical company of mistakenly posting PPD publicly is significant and potentially debilitating. The liability involved in transparency initiatives can be paralyzing to consider. Understanding the limitations of humans, the complexity of large- or small-scale redaction efforts, and the implications if errors are made, hinder the implementation and successful realization of agency and corporate transparency initiatives. Synchrogenix's robust AI engine, significantly reduces those risks. Our CRMS consistently and accurately identifies and redacts sensitive information that manual options simply cannot match.

Internal Oversight

Not only do we reduce your risk, we also drastically reduce your internal project management burden. Consider a standard transparency initiative involving millions of pages redacted and the staff required for that effort. Contracting that project to a standard CRO with manual options is still a complicated maze of oversight wrought with potential disasters. Our CRMS and QC process reduces the need for managing a large group of possibly-changing staff to coordinate the redaction efforts.

Other Uses

Synchrogenix's AI engine is seemingly limitless in its functionality. In addition to CRMS automation, our AI engine is currently used for:

- CSRs
- Submission documents (CTD 2.7.3, 2.7.4, etc)
- Patient narratives

About Certara

Certara is a leading provider of decision support technology and consulting services for optimizing drug development and improving health outcomes. Certara's solutions, which span the drug development and patient care lifecycle, help increase the probability of regulatory and commercial success by using the most scientifically advanced modeling and simulation technologies and regulatory strategies. Its clients include hundreds of global biopharmaceutical companies, leading academic institutions and key regulatory agencies.

For more information visit www.certara.com or email sales@certara.com.

© Copyright Certara 2016

BR Jun 2016 v2 082216