

Truth Be Told: Where AI and Automation Can Really Take Us

AI and automation are both exciting and daunting prospects in the future of medical writing, but the advantages of these tools should be scrutinised, as well as what they mean for medical writing professionals

Julia Forjanic Klapproth at Trilogy Writing & Consulting and Angela Russell Winnier at Pfizer

Anyone actively involved in the writing of regulatory documents needed for clinical drug development knows that there are large content redundancies across those documents. From the introduction of the first study protocol to the clinical overview of the final dossier submitted for marketing approval, we are telling the same story over and over again. New data are added and the messaging gets more refined, but, overall, many of the pieces are repeated across the documents. Different teams rework similar text in overlapping work streams. With the increasing number of concurrent studies being conducted, we need to be able to match increased scale with increased productivity in communication of results. We could add more people to tackle this scale or we could explore solutions that amplify the efficiency of our existing staff, while also reducing the risk of errors and inconsistencies in how information is communicated.

As a result, the idea of automating the reuse of this repetitive information across documents arose decades ago in an attempt to save time and cost by eliminating those inefficiencies and risks. However, initial versions of automated reuse software were cumbersome and restrictive, and most authors found it simply faster and more effective to continue to

copy and paste the material they needed. Today, the technical options are much more advanced and offer some truly meaningful alternatives to speed up the writing of these documents.

Technically, we now have a spectrum of different tools being developed that can be used in combination to

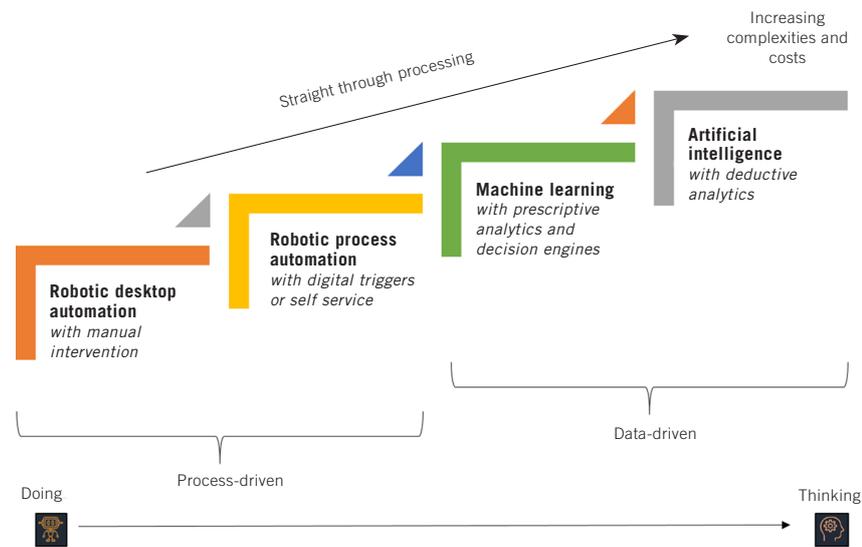


Figure 1: Technical spectrum of writing automation and AI processes

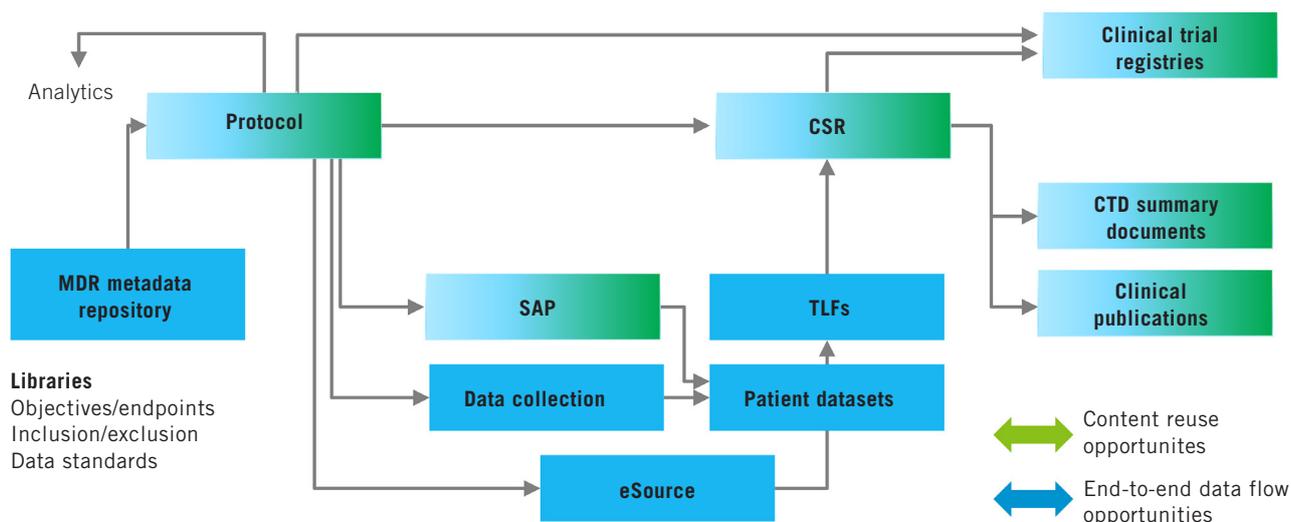


Figure 2: Documents generated using intelligent templates for automation and content reuse. CSR = clinical study report, SAP = statistical analysis plan, and TLFs = tables, listings, and figures

assimilate, in an automated way, fairly complete drafts of documents. Many in our industry are working towards an AI solution for document development. Such a tool would be able to understand, reason, learn, and interact with humans in a natural language. However, to get to a true AI, we typically start with process-driven tools to assist in document content generation and reuse. We create bots that mimic human action by automating routine work. In this scenario, every activity has to be explicitly programmed. We set the rules for what actions follow a manual intervention or an automated trigger. The robotic process-driven solutions are quicker to implement with lower complexity and cost compared with AI data-driven solutions, so it is somewhat natural to start with this end of the paradigm shift (see **Figure 1**). Evolving, then, from the process driven tools are solutions that include machine learning capability and AI that can apply prescriptive and deductive analytics to select the information desired for a particular topic area. Process automation pulls predefined tables or text blocks (e.g., from libraries) based on programmed content creation or the use of intelligent templates. In comparison, machine learning and AI use natural

language solutions to auto-generate *de novo* text and tables based on teaching the tool how to find the right information and extract what is needed.

The majority of the technical tools currently in use utilise process automation based on programmed content creation. This involves template-based rules for content selection and generation, pulling from content libraries, datasets, and even from the documents available in a company's internal document management system. If metadata content repositories can be created, say for a specific new medicinal product in development, then these tools can pull from such a repository to populate several documents over the course of the development programme: clinical protocols, study reports, the statistical analysis plans, summary documents for the common technical document (CTD) dossier, and even clinical publications (see **Figure 2**). The TransCelerate tech-enabled clinical protocol template has been designed to facilitate exactly this kind of automation and content reuse, and, used in combination with the TransCelerate clinical study report tech-enabled template, these tools will rapidly change the way these

documents are generated (1). Effective content reuse benefits from the application of a foundational content model, which provides an end-to-end document map of where all information will be used across the documents of the clinical programme (see **Figure 3, page 32**). It also identifies the specific content components that will be reused, the sources of that content, and defines any relationships and interdependencies. For example, some texts evolve over the course of a clinical programme. This is a natural and necessary part of how messaging develops over time, both due to new information becoming available that refines and changes the clinical story, and also due to the natural crafting and building of ideas that occur over time. Evolutions in clinical messaging are valuable and need to be reflected in documents that are written later in the developmental programme compared to earlier documents. As a result of this, any automation tools must reflect interdependencies across documents, which may specify that later permutations of a text block supersede earlier versions. A content model compares documents across a programme, reading each sentence written and identifying verbatim and non-verbatim matches across these.



It is precisely this intellectual input that no AI tool can deliver, and it means it will be a long time before good medical writers are out of a job



standardised and similar, when an AI tool begins to analyse them for similarities, we see just how different each study report is from the next. Even study reports written in the same company and for the same therapeutic area often have considerable-enough differences in structure and the way that information is communicated that the current AI technology is unable to recognise meaningful patterns among them. Without feeding the system tens of thousands of any type of document, true AI will not be really applicable.

So, how do we tap into the AI technology without pooling all the study reports of all the pharmaceutical companies? We apply machine learning that uses domain knowledge. What this means is there is a combination of natural language techniques and rule-defining. In other words, we help teach the system what patterns to identify. This is very similar to teaching a small child how to play a sport. You explain to them the basic rules, then you let them have a go at it; you see what they do not understand, then you explain some more to help them understand. It takes patience and lots of practice. Think of the 10,000 hour rule to master something (2) – teaching an AI system is very similar. With time, the system gets more and more refined in what it can produce. However, what everyone has to remember is the system is not thinking. It is getting better at recognising what we want to put where.

This is an important concept to understand in the face of socialisation challenges associated

with using AI tools. Many medical writers have only heard that the AI is going to write reports from scratch in a matter of minutes, which, to many, sounds like a job-elimination programme. The truth of the matter is that the output of the AI tools will be an initial draft that pulls together content from source documents (e.g., the protocol and the SAP) and from source data. It will get all the pieces in place to then do the really intellectually interesting part of medical writing, namely, working with the authoring team to craft the storyline and refine the messaging. It is precisely this intellectual input that no AI tool can deliver, and it means it will be a long time before good medical writers are out of a job.

Overall, the advantages of all of these technical tools are the speed of generation and the reduction of errors that would be related to manual data entry. This allows the medical writers and other authors to focus on the messaging with a much faster turnaround once sources are available. In the interest of getting medications to market faster, these tools will mean we can get submissions and pharmacovigilance documents out the door faster and enable all of the people involved in producing these deliverables (whether it is quality control specialists, clinical leads, or medical writers) to process a greater number of them in a shorter timeframe with higher accuracy and correctness. Truth be told, this is a good thing.

References

1. Visit: blissintegrate.staging.wpengine.com/assets/clinical-content-reuse-assets

2. Gladwell M, *Outliers: The Story of Success, 1st edition, Little, Brown and Company: 2008*



After receiving her PhD in developmental neurobiology, **Julia Forjanic Klapproth** started her career as a medical writer in the regulatory group at Hoechst Marion Roussel (later Sanofi) in 1997. Since then, she has been President of the European Medical Writers Association twice (2001-2002, 2007-2009). In 2002, Julia Co-Founded **Trilogy Writing & Consulting**, a company specialised in providing regulatory medical writing. In addition to managing the company as President/Senior Partner, she writes a wide array of clinical documents including study protocols, study reports, and is specialised in the clinical parts of CTD submission dossiers.



Angela Russell Winnier has almost 20 years in technical/regulatory writing and clinical sciences experience, including 15 years of medical writing team management in the CRO and pharmaceutical company settings. Angela is currently a Senior Director and Therapeutic Area Lead for Immunology and Inflammation Submission Medical Writing at **Pfizer**, in addition to the lead for medical writing automation strategies. Prior to her time at Pfizer, she had the privilege of managing a large contract medical writing team that supported pharmaceutical companies of all sizes, ranging from small biotech to the top 5 pharma. Angela is a graduate of Emory University (BS in Biology and Art History) and Vanderbilt University (PhD in Cell Biology).