

Using Natural Language Processing to Create Subject Safety Narratives

Frank Meloni, Regulatory Medical Writing
Tim Willittes, R&D Intelligent Automation

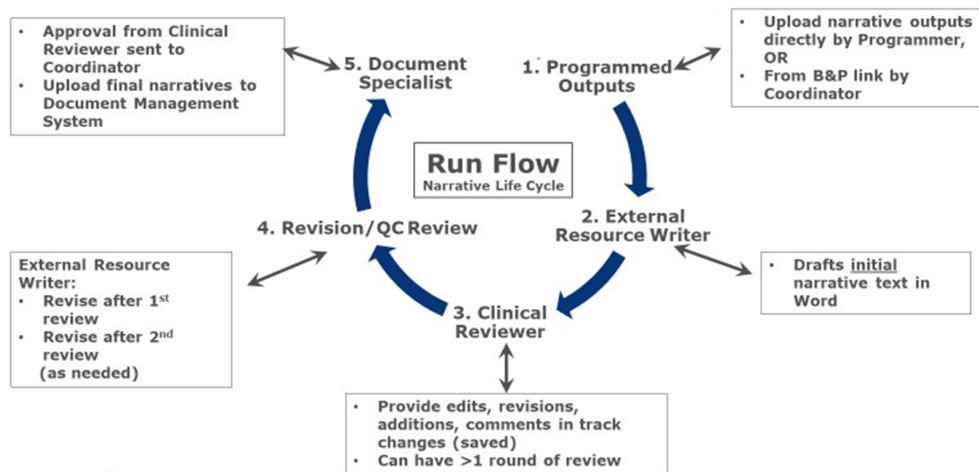
Subject Safety Narratives – Overview

Janssen prepares thousands of narratives each year in support of regulatory submissions to health authorities

Subject Safety Narratives detail information about **Adverse Events** of special clinical interest (SAEs, Deaths, etc.) and form an important component of **Clinical Study Reports**.

Clinical Narrative authoring involves communicating the messages clearly and effectively by **pulling relevant information from Clinical data** and liaising with experts in the Therapeutic Areas.

Currently, Clinical Narrative authoring is manual, time consuming and difficult to **track and manage** reviews and updates – especially for large studies.



Value of automated first narrative drafts

Using technology, we can realize efficiencies and reimagine the narrative authoring process to achieve critical business outcomes

Efficiency

Automation provides significant cycle time reduction vs. current manual process

Consistency

Narrative drafts follow a consistent format and structure

Flexibility

Ability to scale with demand and process many trials in parallel

Accuracy

Information is extracted direct from source

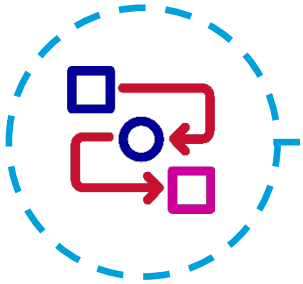
Speed to innovation

How an agile methodology played a key role in the success of this opportunity



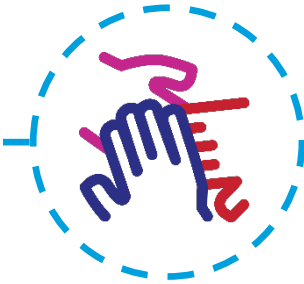
Technical Prototype

Short and focused effort to assess the technical feasibility of NLP and automation as a viable solution



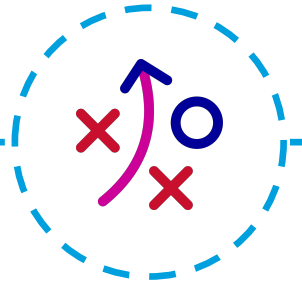
Process Prototype

Building on the technical prototype to automate the end-to-end process including exception scenarios



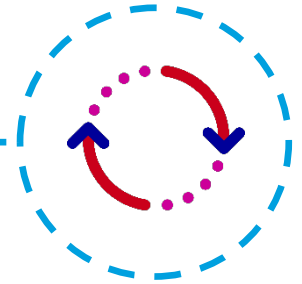
Stakeholder Alignment

Using the process prototype as a foundation to educate and gain alignment on the full capability



Full Delivery

Agile sprints plus regular feedback resulted in higher quality at velocity



Continuous improvement

Ability to update rules engine or NLP model based on experience

Multi-tool approach

By combining multiple tools, each based on their own strengths, we were able to deliver an effective solution



Natural Language Processing

Using a multi-task Convolutional Neural Network trained for Part of Speech tagging, parsing, Named Entity Recognition and vectorization, information is **broken down** into individual sentences and words and then **scored based on relevance**.



Robotic Process Automation

RPA creates additional rules-based content within the first narrative draft and allows the team to impact parameters for the NLP model while maintaining their current business process.



Wrapper Functions

Additional functionality wrapped around NLP and RPA served as the crucial fabric for these technologies to operate seamlessly and craft narratives in a rapid and effective manner

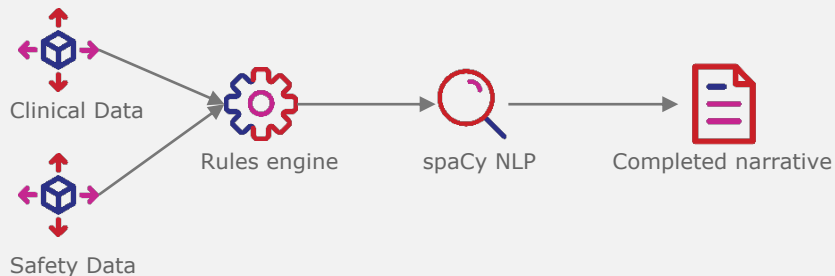


spaCy

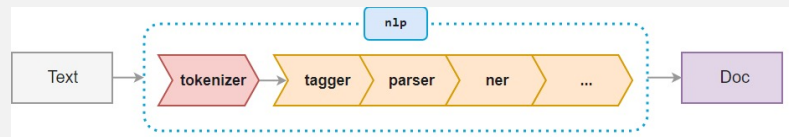
Using open-source tech to accelerate results

Open-source NLP models served as a strong foundation to rapidly produce initial outputs. With subsequent iterations, the NLP model was customized and model inputs were structured to make this open-source technology truly viable in a clinical setting.

Data Flow Diagram



spaCy NLP Pipeline



Model Customization

Specific functions within the model pipeline were customized to achieve:

1. More robust sentence splitting
2. Better chronological ordering of sentences
3. Rewording phrases to help anonymize patient information

Example Automated Output (with additional anonymization)

■ NARRATIVE TEXT

The subject was randomly assigned to receive GOODDRUG 50 mg subcutaneously (SC) every 4 weeks (q4w) and received her first dose of GOODDRUG on 27 Jun 20XX. Additional medical history not listed in the table above but mentioned in the CIOMS report included urticaria and osteoporosis.

SAE: Cataract (Study Day 112)

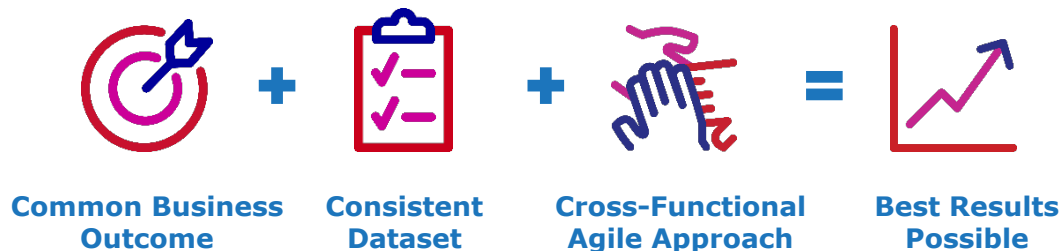
According to the CIOMS report, on 16 Oct 20XX, the subject had a strange feeling in both eyes and was diagnosed with cataract. She was hospitalized on 28 Jan 20XX she underwent lens reconstruction and intraocular lens implantation for right cataract on 29 Jan 20XX and for left cataract on 31 Jan 20XX. Additional treatment included pirenixine, ceftriaxone sodium hydrate, levofloxacin, and cefcapene **pivoxil** hydrochloride for prevention of postoperative infection. The subject was discharged on 04 Feb 20XX. She recovered from cataract on 22 Jan 20XX.

Potential expansion opportunities

Using NLP and automation can positively impact many areas across clinical development

Full range of possibility to achieve better outcomes:

- Clinical documents (CSR, Protocol, ICD etc.)
- Statistical analysis report summary





PHARMACEUTICAL COMPANIES OF

Johnson & Johnson