Lionbridge
LIFE SCIENCES

Expert
Language
Translation.
Global
Success.

Scaling to meet your translation needs

Plus ça change: Document Revision Control in Life Sciences Translation

Introduction

Translating content for the life sciences industry presents a number of challenges. Some of these are due to the nature of the subject matter itself; scientific and medical documents demand much higher levels of expertise and precision than are typically required in mainstream commercial translation.

However, the challenges are technical as well as linguistic. As electronic records and systems continue to gain acceptance among industry leaders and regulators alike, information technology is playing an increasingly important role in the way life sciences companies produce their content and ensure its compliance with regulations.

Document revision control is one of the emerging challenges of this transformation. This article looks at how revision control is used by life sciences companies and regulators, and examines some of the key challenges language service providers (LSPs) must consider when integrating revision control with existing translation technology and processes.

What is document revision control?

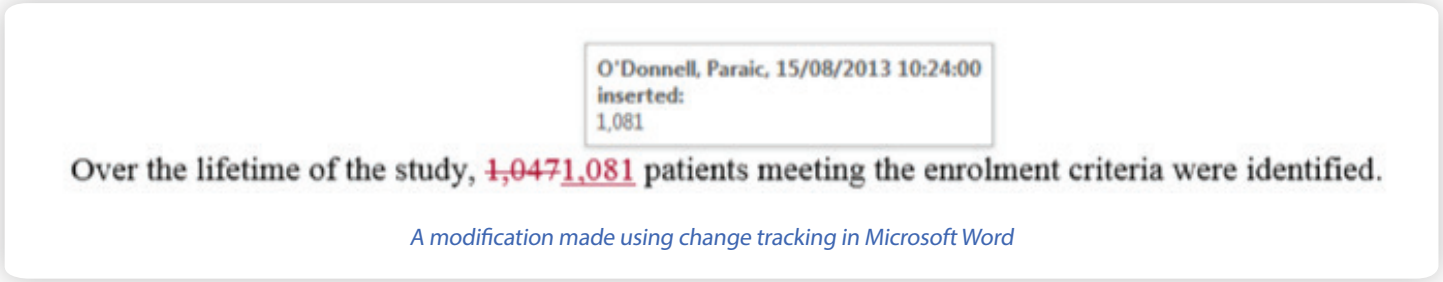
For most of us, editing and saving electronic documents is something we do so often that we rarely think about it. We add a few paragraphs to a report, correct some typos and click on Save. A week later, we might want to share that version of the report with a colleague, so we check the date and time on the file. Occasionally, we may need to preserve a particular version of a file, so we give it a suitable descriptive name ("Performance Report with Comments – August 2013") and save a copy.

It's not a sophisticated approach, but it works for most personal documents. Even for many business users, this is all the revision control they will ever need. In regulated industries, however, revisions to documents must be controlled and captured with much greater precision and transparency.

Typically, the minimum requirements for document revision control will include the following.

- Each individual modification to a document must be captured, whether an entire sentence was changed or just a single character.
- For each modification, the system must retain the identity of the user who made the change.
- For each modification, the system must retain the exact date and time when it was made.
- A user reviewing the document must be able to interpret each modification easily. In other words, they must be able to tell what was there before and what exactly has been changed.
- It must be possible to "reject" changes (go back to the "old version") or accept them.

There are many software systems available that provide these features. Some more sophisticated systems offer even more detail and control, or are designed for special purposes in particular industries. By far the most widely used of these are the change tracking features included in Microsoft Word. In fact, because Microsoft Word documents with tracked changes are required for submissions to regulatory bodies like the European Medicines Agency[i], this system has become a *de facto* standard in the life sciences industry.



A modification made using change tracking in Microsoft Word

Revision Control and Regulatory Compliance

The question of whether particular processes that rely on software systems are regulated by statutory authorities is one that often causes confusion. While a comprehensive discussion of this area is beyond the scope of this article, there are a number of general observations that should be kept in mind.

Firstly, there are the regulations themselves. In the case of the life sciences industry, by far the most widely applicable regulations are the following.

> Code of Federal Regulations Title 21, Part 11 - Often abbreviated to CFR 21 Part 11, these regulations are set out and enforced by the US Food and Drug Administration (FDA).
>
> Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use, Annex 11: Computerised Systems – Commonly shortened to GMP Annex 11, this body of regulations is closely modelled on CFR 21 Part 11 and is enforced by the European Medicines Agency (EMA).

These regulations are the subject of a great deal of misconception, some of which may be attributed to the claims that commercial providers make for particular software systems, suggesting either that their system is generally compliant or that its deployment will somehow guarantee compliance on the part of the user. Needless to say, these claims tend to be made in promotion of one product or another, and should be treated with caution.

While it is certainly possible to use software systems to achieve compliance in particular activities, the systems themselves are not directly regulated. Equally, it is possible to use even the best designed of systems in ways that violate regulatory requirements.

The regulations reflect this reality. In fact, their scope is deliberately and carefully limited so that only the use of systems to satisfy certain requirements in the course of regulated activities can be said to comply with them. Much of the general intent and origins of CFR 21 Part 11, for instance, can be understood from its very first article.

> *Subpart A--General Provisions*
>
> *Sec. 11.1 Scope.*
>
> *(a) The regulations in this part set forth the criteria under which the agency considers electronic records, electronic signatures, and handwritten signatures executed to electronic records to be trustworthy, reliable, and generally equivalent to paper records and handwritten signatures executed on paper.*

Title 21 dates from 1997, though it has since undergone several revisions and clarifications. At that time, records that were maintained or submitted by regulated organisations were predominantly paper-based. Information technology, though, was beginning to have a transformative effective on many industries and it was clear that life sciences companies could not stand still for long.

With these changes in mind, the FDA was attempting to formulate precisely the criteria under which it would treat electronic records such as computer files in the same way as paper documents bearing physical signatures. In other words, the regulations are concerned with the systems that produce electronic records only insofar as the systems can be shown to guarantee the reliability of those records.

Of course, the regulations don't stop there. They go on to set out the particular characteristics that a computer system must have for its records to be considered trustworthy. Again, this article will not attempt to explain these exhaustively. For the purposes of understanding how document revision control can help to meet regulatory obligations, only a few of these characteristics are of interest.

"[T]he ability to discern invalid or altered records"

-- 21 CFR Part 11, subpart B, article (a).

"The ability to generate accurate and complete copies of records in both human readable and electronic form suitable for inspection, review, and copying by the agency"

-- 21 CFR Part 11, subpart B, article (b).

"Use of secure, time-stamped audit trails"

-- 21 CFR Part 11, subpart B, article (e).

"Revision and change control procedures to maintain an audit trail that documents time-sequenced development and modification of

Before moving on, it's worth making some general remarks about the regulations cited above. Firstly, they are set out in clear, simple language that a non-specialist can easily understand.

Secondly, they deliberately avoid addressing any of the details of how their requirements are to be implemented. The FDA is interested only in high-level characteristics and does not favour one approach over another. It certainly does not endorse one commercial offering over another, contrary to the claims of some of their manufacturers.

Finally, and perhaps most importantly, it is vital to understand that the FDA interprets these provisions narrowly[ii]. It recognises that life sciences companies operate in a highly competetive environment and must employ information technology widely in order to innovate and survive. Specifically, it recognises that not all of the uses of particular computer systems constitute regulated activities.

Revision Control and Translation Technology

For life sciences companies and LSPs managing the translation of regulated content, the technical and regulatory considerations described above take on an additional dimension of complexity. We mentioned that Microsoft Word has become a *de facto* standard for document revision control. Until relatively recently, it occupied a similar position in the translation industry.

The developers of the first generation of computer-assisted translation (CAT) tools recognised that most of the professional linguists for whom their products were designed were already working in Microsoft Word and were familiar with its editing environment. Designing CAT tools that were closely integrated with that environment made obvious commercial sense, and this model dominated the translation technology landscape for many years.

However, a number of factors have contributed to the emergence of an entirely new model. Perhaps the most important of these has been the development of Extensible Markup Language (XML) and the extent to which structured documents built on its principles have become pervasive.

For the developers of translation technology, one of the most attractive features of XML has been the way in which it separates content (in this case text) from formatting (there are many kinds of formatting, but simple examples include fonts, colours and so on). In the old days, CAT tools relied on Microsoft Word's rich text format (RTF), in which formatting information was deeply embedded in the text.

What makes CAT tools useful to translators is that they "remember" how a sentence was translated last time by looking it up in a database called a translation memory (TM). To do this, they need to break documents up using a process called segmentation.

The problem with RTF is that, even after segmentation, the chunks of text that end up in TMs are full of proprietary formatting code. That's fine if all you want to do is translate other RTF files, but for other forms of content, it greatly reduces the number and quality of matches a translator will get. The TM is effectively "polluted."

The move towards structured documents has led to significant changes in the way that CAT technology is designed. The abstraction of the text for translation allows it to be manipulated with less effort, and in much more powerful ways. This has permitted the development of rich terminology identification features, the assembly of very large cross-platform corpora, the application of automated quality checks and many other improvements.

While these have been positive developments, they have had the effect of breaking the close links between Microsoft Word and CAT tools. Word documents are now routinely converted to proprietary structured formats for translation and converted back prior to delivery to customers.

In most circumstances, this has no undesirable effects. For life sciences customers, however, and for the specialist LSPs they work with, questions are emerging about how widely used forms of document revision control can be accommodated in this new model.

Emerging solutions

LSPs and translation technology providers are beginning to offer solutions that bridge the gaps between document revision control and CAT tools. Especially among specialist life sciences providers, there is increasing recognition of the need for tracked changes to be preserved throughout the translation life cycle.

In navigating this territory, life sciences companies should bear in mind the points discussed above in relation to the scope and interpretation of regulations. There are many valid technical approaches to these problems, but ultimately what matters to regulators reviewing electronic documents is that the system that produced those documents can be shown to satisfy a few relatively simple criteria.

Does it make it easy to discern changes? Does it generate a secure, time-stamped audit trail? Are the files it produces accurate and complete records, both of the clinical or scientific data being reported and of the authoring and revision activities that produced those reports?

Of course, decision-makers in life sciences will also evaluate the features and performance of particular commercial solutions, but when making strategic choices, most will also look for partners who offer not just a technical solution but a deep understanding of regulated processes in large, transnational supply chains.

In these environments, the "system" is more than just the software; it is a wide-spectrum solution depending on many disciplines. It encompasses not just proven technology, but also domain expertise, linguistic authority and IT security. It is providers building on these foundations that will lead the way in integrating compliant revision control with excellence in life sciences translation.

To learn more about document revision control, or to speak with a Lionbridge Life Sciences expert, contact us today at http://www.lionbridgelifesciences.com/contact/.

References

Code of Federal Regulations Title 21 Part 11 http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?cfrpart=11

Guidance for Industry, Part 11, Electronic Records; Electronic Signatures – Scope and Application http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm125125.pdf

Good Manufacturing Practice, Medicinal Products for Human and Veterinary Use – Annex 11: Computerised Systems http://ec.europa.eu/health/files/eudralex/vol-4/annex11_01-2011_en.pdf

---

[i] Specifically, the EMA requires that all submitted documents be in Microsoft Office 2003 format. See: http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/landing/human_medicines_regulatory.jsp&mid=WC0b01ac058001ff89

[ii] Guidance for Industry, Part 11, Electronic Records; Electronic Signatures – Scope and Application. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm125125.pdf