In April 2019, zbMATH was completely transformed from \TeX{} to \LaTeX{} sources. On this occasion, we give a brief history of typesetting Zentralblatt volumes, and describe the challenges, methods and benefits of the transition.

**A very short overview of typesetting Zentralblatt**


\TeX{} and \LaTeX{} have been the standard tools for creating documents for at least two generations of mathematicians. Today it is almost inconceivable that mathematical content was typeset before. Indeed, mathematical typesetting has a much longer history at zbMATH. zbMATH has existed for about 150 years if one includes its printed predecessors, *Jahrbuch über die Fortschritte der Mathematik* and *Zentralblatt für Mathematik*. Although it would be somewhat tempting to retell the history of mathematical typesetting in these periods based on the appearance of the *Jahrbuch* and *Zentralblatt* volumes in a similar manner as, e.g., [5], this would go beyond the scope of the present article. Instead, the figure below provides impressions of volumes based on various lead types, linotypes, or phototypesetting using IBM golfball typewriters.

In general, mathematics was always very expensive to typeset.\(^1\) The various developments until the 1970s aimed to make this process more efficient. Famously, the quality of the more cost-efficient phototypesetting technique decreased in comparison to classical lead techniques [5]. Indeed, in the *Zentralblatt* volumes from the years of phototypesetting one can observe several formulae which needed to be added by hand, before one could start to create the phototypesetting master by cutting and gluing.

However, at *Zentralblatt*, the shift to phototypesetting was inevitable due to the growing amount of content, which resulted in the production of immense register volumes. This created the most obvious and urgent need for further digitisation.

**From the beginnings of \TeX{} to its adaptation**

Donald Knuth developed the \TeX{} system to overcome the limitations of phototypesetting. As a welcome side effect, the fact that \TeX{} is an open system established the autonomy of mathematical writing. With hindsight, this appears to be an inevitable development, though in the transition period it was not. For instance, a subject of detailed debate was whether it would be more efficient to employ scientists as their typesetters, instead of having specialised staff [9].

A major milestone in this process was the \TeX{}82 version with both its improvements and stability. The 1982 meeting of the \TeX{} Users Group at Stanford was the first TUG meeting lasting for about a week. Incidentally, a *Zentralblatt* editor – who was involved at that time in the editors’ exchange program with Mathematical Reviews\(^2\) – took part, and advocated later in Berlin the adaptation of \TeX{} for the production of the *Zentralblatt* volumes. At Mathematical Reviews \TeX{} became fully productive as of 1985, after several years of preparation [7]. In contrast, it took until 1992 before the advent of \TeX{} at *Zentralblatt*. In the meantime, the phototypesetting technique became no longer sustainable. In 1984, a proprietary, internal Springer system was employed that resembled many of the typesetting functions of \TeX{}. The commands of the Springer system were designed for specialised technical staff instead of self-use. A key argument for the Springer system in contrast to \TeX{} was the resulting lower number keystrokes for volume production.

From today’s viewpoint, the greatest advantage of moving away from phototypesetting was that at least the texts and formulae were available in a digital form for the volumes 531–734. However, the disadvantage was that after the Springer system became outdated in the early 1990s, the introduction of \TeX{} required conversion: simultaneously to the Springer-based production plan. Since the transition to \TeX{} also included a migration from Springer servers to FIZ Karlsruhe, the expected delay occurred.\(^3\) Fortunately, the \TeX{} expertise acquired in the meantime allowed for a successful transition. It turned out that even most formulae could be translated automatically, though some constructions

---

\(^1\) Typically, there were only a few highly specialised typesetters involved. One anecdote that has passed through generations was that one single typesetter was responsible for producing *Zentralblatt* volumes for many years. Eventually, he was able to spot errors in mathematical formulae without any semantical knowledge.

\(^2\) That was also the time when a merger of both services was actively pursued.

\(^3\) The 3-month hiatus between Vol. 734 and 735 is by far the largest post-war gap.
had inherent problems; e.g., matrices needed conversion from column-based encoding to row-based encoding.\footnote{By looking at the sources of these, a reader can easily indicate that they were not genuine \TeX-coded; e.g., single variables were not set in formula italics in the old system, so they lack conversion until now.}

**From \TeX to \LaTeX**

For some years after the switch to \TeX, the production of printed volumes had been the main objective. The main objective was the appropriate presentation. A standardised encoding, which is desirable from an information retrieval viewpoint, was less relevant. This changed in the second half of the 1990s after the online database became the primary objective. In 2004, adapting to these needs, the database production switched to a PostgreSQL system. This stored all available information in ASCII-coded \TeX\,texts. This framework hadn’t changed significantly until recently.

Of course, despite the impressive robustness and stability of \TeX\,technical development didn’t stop in the 80s. Probably already in the mid-90s \LaTeX\,was the preferred dialect for many users. Today \LaTeX\,accounts for more than 90% of zbMATH review submissions. Since the database routines previously interacted with \TeX, reviews in \LaTeX\,required a re-standardisation. Furthermore, an increasing amount of data is provided in the UTF-8 format. Its conversion to \TeX\,encoding for internal storage and later reconversion for online presentation may cause errors and a loss of information. This pertains especially to bibliographic reference data which may include the need to encode native script such as Arabic, Chinese, Farsi, Hebrew, Japanese, Korean, or Russian. This makes it desirable to have the \XML expansion available.

**How does one convert ~20 million formulae?**

Therefore, a switch of the production system underlying zbMATH to \XML was in preparation for several years. The initial work-intense step was made when MathML was introduced in 2010 \cite{Baier2010}. Standard tools for MathML conversion employed by zbMATH, such as Tralics \cite{LIDM3}, require \LaTeX source. Thus, it was necessary to convert different \TeX\,commands to \LaTeX – at least, those commands which could be processed by MathML converters. This part of the conversion could be amply addressed by regular expressions. However, it turned out to be necessary to build the full expression tree for \TeX\,formulae. This step, which was finally taken in April of this year, was preceded by an upgrade of PostgreSQL, which allows for UTF-8 handling and a 14-hour routine that converted approximately 18 million standard inline mathematical expressions and 600,000 displaystyle formulae. These included more sophisticated environments, like \texttt{alignat} or \texttt{gathered}. Fortunately, only a few environments, e.g., commutative diagrams, require additional manual transition. Overall, the introduction of \LaTeX resulted in a pause of zbMATH updates of about a month. In contrast to the introduction of \TeX, which took three months, this is a significant improvement. The most visible difference for zbMATH users looking at the review sources is the replacement of $...$ by \texttt{(...)}. Standardisation will allow for much easier integrity checks, and for a much more seamless integration of the submitted reviews. Additionally, considerably better presentation is possible by new functions available via \LaTeX\,packages. MathJax \cite{MathJax} (needed for maths presentation in browsers not capable of MathML) works better on widely used \LaTeX\,commands instead of less frequent \TeX\,commands.

Another advantage pertains to the MathML generation; while Tralics works well for presentation purposes, it hasn’t been developed further for some years, so it is reasonable to look for alternative options like \LaTeXXML \cite{LIDM3}. The availability of \LaTeX\,code makes such alternative implementations much more feasible.

**Paving ground for future developments**

Even more importantly, further developments in mathematical information retrieval and processing will most likely be based on \LaTeX. \LaTeX\,became the de facto standard in mathematics, and most working mathematicians use \LaTeX\,in their publication workflows. Moreover, most websites use \LaTeX\,as an input language for mathematical formulae. To make mathematical content better discoverable, multiple approaches exist for enhancing semantics in mathematical formulae. For example, the NIST Digital Library of Mathematical Functions developed a set of semantic \LaTeX macros for mathematics. These macros are easy to use for mathematicians fluent with \LaTeX. By using these macros, with minimal overhead, information retrieval systems would be able to better disambiguate the syntax and semantics for mathematical expressions. Eventually, this will provide better search and recommendation functionality for users of mathematical digital libraries \cite{Cohl2016}. To some extent, such approaches have already been applied to realise the zbMATH formula search \cite{Baier2019,Muller2006}, but having standardised \LaTeX sources at hand will certainly make further developments in this direction much more feasible.

**References**

\begin{itemize}
  \item \cite{Baier2010}, P. Baier, O. Teschke, Zentralblatt MATHMLized, Eur. Math. Soc. Newsnl. 76, 55–57 (2010; Zbl 1278.68088)
  \item \cite{Cohl2016}, H.S. Cohl et al., Growing the digital repository of mathematical formulae with generic \LaTeX sources, Lect. Notes Comput. Sci. 9150, 280–287 (2015; Zbl 06512424)
  \item \cite{Muller2006}, F. Müller, O. Teschke, Full text formula search in zbMATH, Eur. Math. Soc. Newsnl. 102, 51 (2016; Zbl 1366.68355)
  \item \cite{Mohlhase2013}, M. Mohlhase, H. Mihaljević-Brandt, W. Sperber, O. Teschke, Mathematical formula search, Eur. Math. Soc. Newsnl. 89, 56–58 (2013; Zbl 1310.68217)
  \item \cite{Knuth1979}, D.E. Knuth, Mathematical typography, Bull. Am. Math. Soc., New Ser. 1, 337–372 (1979; Zbl 0404.92025)
  \item \cite{MathJax}, \LaTeX\,XML, a \LaTeX to XML/HTML/MathML converter, https://dlmf.nist.gov/LaTeXML/
  \item \cite{LIDM3}, D.C. Latterner, W.B. Wolf, \TeX at Mathematical Reviews, Tugboat 10, No. 4, 639–654 (1989)
  \item \cite{Letters2013}, Tralics: a \LaTeX\,to XML translator, https://www-sop.inria.fr/marelle/tralics/
  \item \cite{Letters2013}, Letters et alia, Tugboat 4, No. 2, 81–102 (1983), https://www.tug.org/TUGboat/tb04-2/tb08letters.pdf
\end{itemize}
Moritz Schubotz [moritz.schubotz@fiz-karlsruhe.de] is a senior researcher for mathematical information retrieval and open science. He maintains the support for mathematical formulae in Wikipedia and is off-site collaborator at NIST.