

Planning a new type of literary edition: the Thomas Mann Project

Ingrid Schmidt

VIA, Heidelberg, Germany
schmidt@via-iscm.de
<http://www.via-iscm.de>

Carolin Müller

VIA, Heidelberg, Germany
mueller@via-iscm.de
<http://www.via-iscm.de>

Abstract:

The constantly changing requirements of today's media landscape demand a new concept for literary editions. Such a forward-looking model should be SGML/XML-based, and should acknowledge the central importance of topic maps. In this respect, the Thomas Mann project combines in a unique way the work of one of the most famous authors of the 20th century with an innovative way of information organization.

Introduction

Cross-media publishing, multimedia publishing, electronic publishing: new terms are being added almost every day. But what is behind all this? Are these just buzz words or really new developments? At first glance, it is hard to distinguish between them. When asked directly, everyone seems to understand something different by it; the scope is huge, and it's hard to define the borders. Does it help us any further to determine the differences between these terms and their common ground? Or is this simply art for art's sake? Isn't the point rather to comprehend the phenomenon as such?

With regard to the publishing industry, a more detailed analysis cannot be divided from the following questions: What are the needs of the market and how can publishers respond by developing forward-looking strategies? It is not always a question of reacting fast and at low cost to market requirements, nor does it make any sense to undertake a long and costly planning phase and thereby lose touch with the target product. The goal is to develop strategies based on constructive interaction between theory and practice. However, there is no patent remedy; it is always necessary to take into account the specific circumstances of a particular publishing house.

Above all, strategies are also needed because developments do not always gather momentum of their own accord. It's not just a question of reacting to market needs but also of actively affecting them. Even though the field of new media is a fast-moving market, developments can be recognized and influenced by including them in one's own planning.

Against this background, the Thomas Mann project is an example of a forward-looking publishing concept. We first describe the starting point of the project, discuss the project requirements established on this basis, and then present their conceptual realization. In the final part, we indicate the new perspectives offered by this concept and show how its universal approach can be applied to other projects.

Starting point

Thomas Mann is one of the most significant German authors of the early 20th century. His extensive works reflect many of the major concerns of his age, and have a firm place in classic German literature. To this day, the author Thomas Mann generates the biggest turnover for the German publisher S. Fischer. Fischer holds the exclusive publishing rights of his works until 2025. Between 2001 and 2015 they are planning to bring out a 58-volume reference edition of his works, letters and diaries at a rate of 2-3 volumes a year. This edition, called GKFA (Große kommentierte Frankfurter Ausgabe) , will be edited and include about 20 volumes of comprehensive annotations by at least 15 international renowned researchers.

Fischer's decision to publish this new reference edition - which is intended to provide unique access to Thomas Mann's works for many decades - was also based on a number of other considerations. It was obvious that a project planned for such a long period of time should be published not only in the form of printed books, but also electronically. After having examined the available literary editions already published in electronic form, it became clear that they did not resemble what was envisaged for the GKFA. It was, however, not specified what the electronic version of the GKFA should look like; but it should on no account be merely a copy of the print version. It had to be guaranteed that the quality of the electronic version should meet the high standards set for the printed version. In this context, high quality meant not only that of the content, but also in particular its realization in a form appropriate to the medium. Moreover, as part of this huge edition, it was required that the content should be prepared in such a way as to make it a resource for long-term use. Another big project at Fischer had proved over several years that SGML/XML was a good investment for long-term archiving of data. It, therefore, seemed a good idea to use these standards for Thomas Mann, too.

With this in mind, two central project requirements could be set, and these will be discussed in the next section:

1. The GKFA will be published in printed and electronic form.
2. The texts of the GKFA will be stored as SGML/XML.

Project requirements

Printed and electronic version

The first project requirement was to publish a printed and an electronic edition of the GKFA. This requirement, however, said nothing about the relationship between the two editions. Comparable literary editions published to date normally show not only that the electronic version has been created on the basis of the book, but also that it follows it closely.

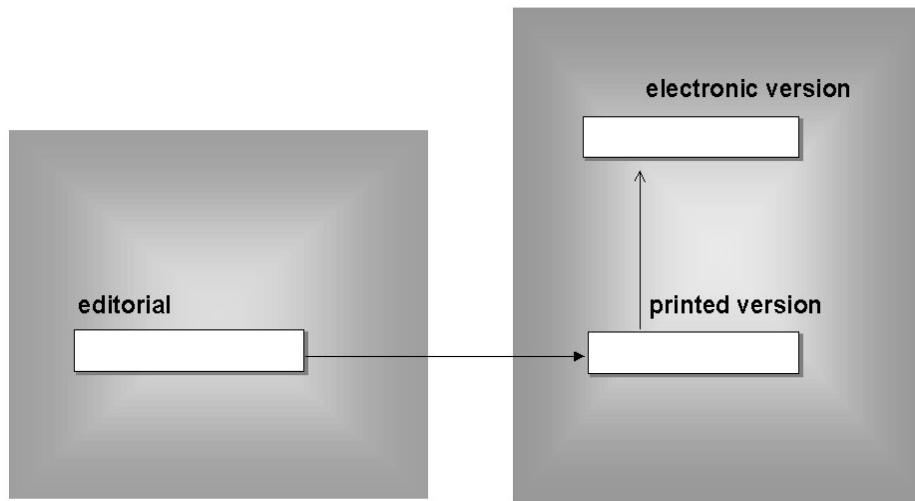


Figure 1. Electronic version as replica of the printed version

Most electronic editions take over the content prepared for the book almost unchanged and thereby imitate the printed version. The media-specific qualities of the book such as the pleasure of reading and the feel of the paper cannot be transferred to the electronic medium. The specific features offered by the electronic medium, however, are the various, fast ways it provides for accessing the content. For this reason, the electronic version must be a convincing working tool, and not merely a replica of an edition designed for pleasurable reading. Nevertheless, many available electronic editions do not provide any considerably different retrieval possibilities apart from full-text searches. This functionality is of course a valuable asset, but it by no means utilizes the power of the electronic medium to the full. Full-text retrieval usually leaves the user's expectations unfulfilled: typically, too many hits are listed in uncategorized form; important content references are found only if the search term appears explicitly in the text. If we apply this to the works of Thomas Mann, "Schiller" and "Goethe" would not be found in the novel *A Weary Hour*, and it be established that the novel *Buddenbrooks* is set in Lubeck, because none of these terms occur in the text.

But how can the potential of the electronic medium be fully utilized? Aside from the development of new ways of retrieving content, it is a reasonable idea to add multimedia features. But this in turn begs the question as to which multimedia enhancements are appropriate for an academic edition. At first, there were no limits set on ideas for the Thomas Mann edition: there existed various portrait photographs of the Mann family, film shots of the Nobel prize ceremony, the seating plan of the Buddenbrook family Christmas dinner, bible facsimiles with Thomas Mann's notes for his Joseph novel, music pieces, recordings of BBC radio addresses spoken by Mann himself, a documentation of Thomas Mann's record collection, recordings of public readings, replies to his letters, or his notebooks. At first glance, a whole host of possibilities opened up - brave new world! But was this new world really that brave? Already the second glance put a considerable blight on it. It was unclear how the multimedia enhancements related to the academic demands of the edition. Including sources independent of their medium, seemed to be unproblematic. But how, for

example, should the documentary film of the Nobel prize presentation be handled? In a strict sense, it is not part of an academic edition. One possibility that suggested itself was to relax the strict definition of the academic nature of the work. But then, how should the character of the electronic version relate to that of the book? The book as an academic edition in the strict sense of the term, and the electronic edition as a lesser academic edition, but with more fun features? Many open questions needed to be answered. Moreover, one important point was still unresolved: How did all this tally with the role of the editors? Their names stand for the concept of the printed version. Can this concept simply be altered for the electronic edition? Can the editors then also be regarded as editors of the electronic version?

The central requirement of the GKFA is to be the new reference edition for Thomas Mann. A reference edition is the binding edition for quoting Thomas Mann. But if the concept of the printed edition differs from the electronic one, which part of the edition can then be called the reference edition? Traditionally, it is the book. But should it not be the electronic version, if this includes more Thomas Mann texts than the book itself? Would it not rather be desirable to keep it as **one reference edition**, realized in **two media**? What is the position regarding the quotability of the electronic medium? An academic edition must provide a means of quoting passages precisely. This is possible in the book by using page and line numbers; there is as yet no corresponding mechanism electronically.

The decision to publish the GKFA in two media left a lot of detailed questions unanswered; for the planning phase, however, the following points could be extracted:

What is the relationship between the book and the electronic edition?

How extensive should the electronic version be?

How should the question of reference edition be resolved?

Data storage in SGML/XML

The second project requirement for the edition was that the Thomas Mann texts should be prepared in such a way as to make them a resource for long-term use. Now, what does this mean exactly? The basic prerequisite is that the data can be used for a long time without any need for conversion, and therefore need to be stored in a system-independent format. The most obvious solution is to use the international standard of SGML/XML. The main motivation for building up such a resource was to ensure a business advantage for Fischer that extended beyond the year 2025, when the exclusive publishing rights run out. This goal can be reached if the content prepared for the GKFA can also be used for other in-house projects and products. First of all, it seemed reasonable to think about possible spin-offs, e.g., a school textbook edition of Buddenbrooks that made use of the explanation of terms in the annotated GKFA, or a volume about "Thomas Mann and Switzerland". But how would the content have to be prepared in order to make this possible? And what is required in terms of time and money to actually generate such products? What would such a publication process look like? Furthermore, it is conceivable that other material could be included that complements the works of Thomas Mann, without being part of the GKFA, or the works of other authors could be affiliated, e.g., those of his brother Heinrich Mann. What does that mean for organizing data storage? Which of the structures developed for the GKFA could be reused in such a case?

With regard to this second project requirement, namely, to structure the texts in SGML/XML, the concept has to provide answers to the following questions:

What is the publication model for the Thomas Mann project?

What is the concept for archiving the content?

How is the DTD model built up?

In the previous sections, we have shown which considerations accompanied the initial project requirements. In this discussion, it became clear that we could not fall back on any existing concept. The next step, therefore, was to develop a concept for the Thomas Mann project based on the above requirements. This concept consists of one basic concept and further concepts for the information pool, for the GKFA, and for structuring.

Concept

Basic concept

The basic concept of the Thomas Mann project is the component scheme of a SGML/XML-based publishing process. This scheme facilitates orientation in the overall project and shows the relationship between the two project requirements.

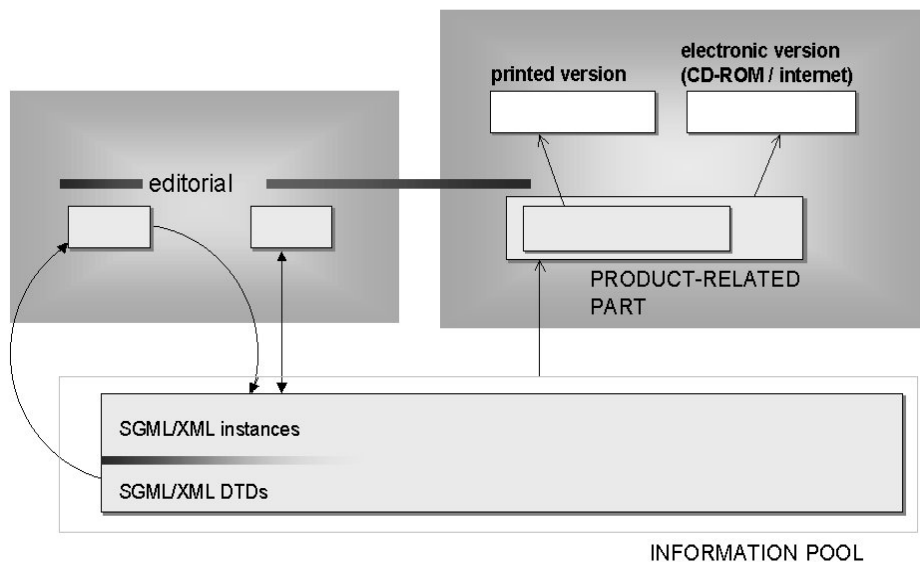


Figure 2. Scheme of components in an SGML/XML-based publication process

The core of the schema is the information pool; it contains all the content, including SGML/XML-structured texts, photographs, and films. In addition to the information pool, there are the areas of editing and product generation. On the left-hand side, two editorial tasks are shown: structuring of the content according to the SGML/XML DTDs, and editing previously structured documents that are stored in the information pool. Another editorial job, namely, the extraction of content for a specific product, i.e., the definition of a product-related part, is illustrated on the right-hand side. Examples of such a part could be an annotated Thomas Mann novel out of the GKFA, a collection of letters by Thomas and Heinrich Mann, or other editions that

can be extracted from the data source of the information pool. On the basis of this kind of product-related part, both a book and an electronic edition can be developed. As regards content, the basis for the electronic presentation can be more extensive than the one for the book. Nevertheless, one content-related basis serves for various media-specific presentations. In this case, the content is not primarily prepared for the book, and thereby restricted in the ways they can be converted for electronic presentation. Electronic presentations derived in this way generally cannot come up to the quality standards of the book, since the power of the electronic medium cannot be fully utilized any more. On the contrary, the result of the model shown here is **one edition in two media**.

This envisaged SGML/XML-based publication model is not only applicable to the GKFA, but can also be utilized at S. Fischer for the complete works of Thomas Mann and all related material. Furthermore, it is transferable to any publishing project that can be based on a SGML/XML-structured information pool. It is thereby not important whether the edition is published only in book form, only in electronic form, or in both media.

Concept of the information pool

All content is retained in the information pool. Most of it is SGML/XML-structured, other parts (e.g., graphics or films) are stored in their various native formats. The information pool is used for structuring and for editorial preparation of the content as well as for defining and extracting the "content packages" of the product-related part. The editorial work proceeds from different viewpoints, e.g., whether all names of persons, locations, and non-Thomas-Mann-works are consistently marked, or whether all material and translated text passages related to one work are correctly linked. It, therefore, has to be feasible to search the whole host of information with regard to content criteria in order to access it specifically. The definition of a product-related part requires a similar functionality. Here, too, relations regarding content must be retrievable if volumes on "Thomas Mann and Switzerland" or "Thomas Mann and Goethe" are planned. In order to provide for such retrieval questions, the data stored in the information pool has to be structured and highly interlinked with regard to content. Hence, the link concept has to go beyond hyperlinks. It would require a lot of personal resources to carry along all name variants, e.g., of every marked person, and make them consistent throughout. Besides, the possibilities to reproduce the complex relations with regard to content are limited. Using only hyperlinks, it would not be viable to search for 19th century Russian authors with whom Thomas Mann concerned himself. Thus a semantic net is required as a metastructure. Semantic nets will not be discussed in this context. Briefly, the principle is to define objects by their characteristics, and interlink them with relations of different types. The information pool is, therefore, divided into a text basis and a metastructure.

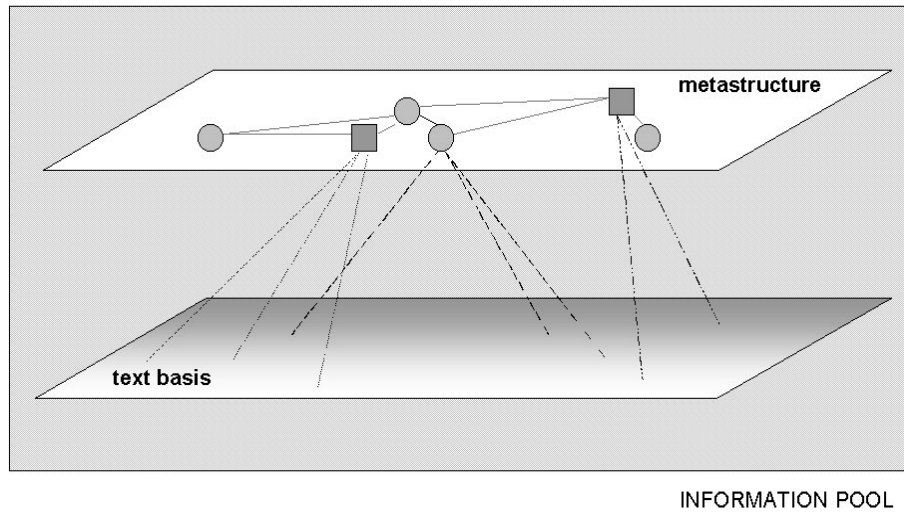


Figure 3. Areas of the information pool

The figure illustrates the various object types of the metastructure by means of different graphic symbols. The objects themselves are linked by relations of different types and, in the same way, are also interlinked with the content of the text basis. The structure of the text basis and the metastructure are SGML/XML-based; the metastructure is realized as a topic map.

This structure of the information pool offers the perspective of adding other Fischer authors. They can be interlinked with the works of Thomas Mann and still remain clearly differentiated. The information pool is structured system and media independent. As a result, this means that the publisher stops archiving content in the form of typeset data, thus freeing it from ties to a specific book format, which often has to be converted to another typographical format if a new edition is prepared. Content is now prepared as variously usable resources. These support, on the one hand, the editorial tasks and make them more efficient. On the other hand, they can be used by the publisher to react fast and flexibly to market requirements. They thus provide the basis for a publisher to not only publish books, but also to expand its traditional business to other media.

Building up such an information pool is, therefore, allied to a change in the publishing strategy: the publisher no longer concentrates on the printed form, but focuses on content, in order to publish it in whatever medium is appropriate.

Concept of the GKFA

As the discussion of the project requirements has shown, there is an area of conflict in the development of an academic electronic edition between the apparently limitless possibilities of the electronic medium - such as unlimited space, inclusion of multimedia elements - and the necessity of limiting an edition in order to preserve its unity with regard to content, and not distort its character with regard to the printed version.

But can it really be the solution to abstain completely from the possibilities of the electronic medium by adding additional content, in order to retain the strict academic character? This certainly cannot be the solution, because the publisher would give away the big opportunity of developing an electronic edition that is attractive for the consumer. Which Thomas Mann expert would not be glad to have a handwritten outline of the Christmas dinner of the Buddenbrook family, or listen to the famous BBC radio addresses by Mann himself? On the other hand, it is also not a solution to extend the electronic edition to such a degree that the print and electronic version diverge so completely that they are unrecognizable as a single edition. In that case, it would probably only be the book that followed the rules of academic editing and, thus, alone would be regarded as the reference edition. This would lead to the publisher gambling away the chance to sell the electronic version to libraries and other academic institutions. Especially in times when cash is short, publishers will not spend money in producing an unquotable non-reference edition with a lot of fun features, if they need then an extra expensive book edition in order to provide the current reference edition for Thomas Mann.

The concept developed for the GKFA, therefore, combines the advantages of possible enhancements in the electronic medium with the conceptual unity for both publication media with regard to content.

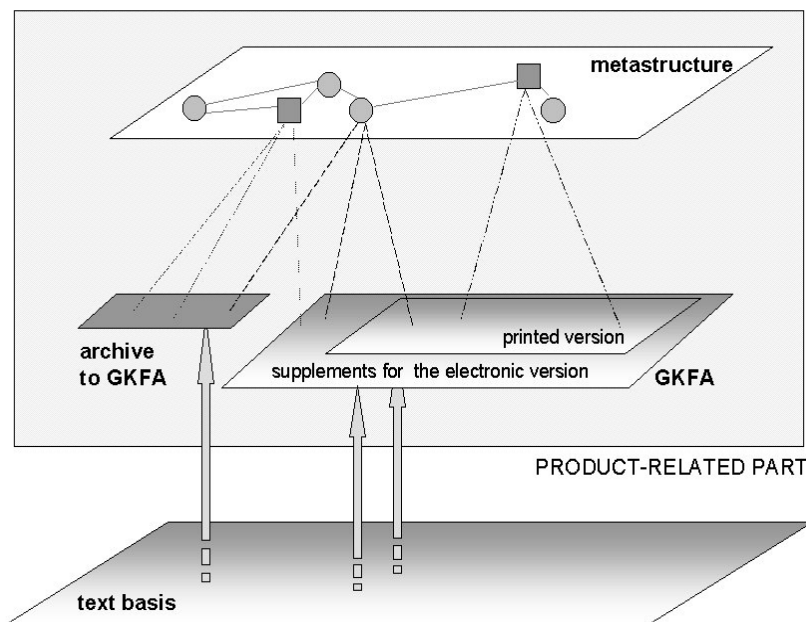


Figure 4. Areas of the product-related part

The text basis for the book and electronic edition is divided into the GKFA and an archive to GKFA. The GKFA consists of a book edition with possible minor enhancements for the electronic version. These enhancements are mostly variant text passages of a certain length, that did not fit into the book for reasons of space, or a limited amount of sources that could not be published in the book for media reasons. They obey the same editing rules as the book edition and are solely decided by the editors. The enhancements are tied in a quotable way to the edition. The final achievement of all this is that both parts of the edition together, book and electronic, constitute the new reference edition for Thomas Mann. With regard to content, they

both comply with the same concept, are both academic editions, and are both answered for by the editors. Each version of the edition, however, satisfies different user needs: the book those of an academic reading version and the electronic version that of a functional working tool. In addition to the GKFA, there is an archive which is the publisher's responsibility. This archive includes documents and texts pertaining in some way to the works of Thomas Mann. It uses the above-mentioned possibilities of the electronic medium to present sound documents, facsimiles that are difficult to access for a broad audience, or to make Thomas Mann's letters more vivid by including the replies from his correspondents. The archive and the GKFA are linked by the metastructure. With the archive, the publisher has the opportunity to add rare and attractive material in its own right to the electronic version and thus create an incentive for buying it. As the archive is clearly separate from the GKFA, these enhancements do not go against the concept of an academic edition of the complete works. Therefore, this separation of archive and GKFA on the CD-ROM is of central importance. The user of the electronic version is always aware of whether he is working with the GKFA and thus using quotable texts selected and edited by the academic editors, or whether he is rummaging through the archive that is not part of the GKFA as such. The two areas of the electronic edition each has a different functionality. Much higher retrieval demands are placed on the GKFA than on the archive. With this in mind, it becomes obvious that the content-related concept influences the structure, thereby making different functionality possible, and moreover, determining the shaping of the user interface.

DTD concept

The aim of the modeling concept of the DTDs is to represent the content units and their interrelation, not the hierarchical structuring of documents. Consequently, the content is stored in a structure that keeps it independent of any media-specific presentation. The DTDs are closely related to the content concept of the relevant version. They represent the interface between the concept and its realization, by modeling the content requirements of the version in a formal language and thereby creating the basis for technical translation. The DTD concept itself is interwoven with those portions of the concept described previously; the DTD concept also distinguishes between the text basis and the metastructure.

Each text in the text basis must be SGML/XML-structured. Texts that later flow into the GKFA are structured more deeply than those assigned to the archive. Structures used in several contexts should only be defined once. In addition, it is also required that the DTDs developed in the course of the Thomas Mann project should also be applicable to other authors or comparable projects. The DTD library thus developed satisfies both conditions. It, therefore, has a modular structure in a double sense that provides for two module types:

1. **Cross-DTD modules** In these modules, elements, attributes, entities, content models etc. that are used in different DTD contexts are bundled together. A Thomas Mann novel, for example, is divided into sections and paragraphs in the same way as the general commentary of the editors.
2. **Modules for structure levels** These modules identify the different levels that make up the text structure. The structural framework of a novel, for example, is distinguished from the semantic elements to mark up persons, locations, and works as well as from the link elements.

The cross-DTD modules are needed to handle similar structures in the text basis accordingly in order to achieve formal consistency in the markup. This is important because the DTD structures are usually not fixed for ever, but change over time. With this in mind, the DTD library should be built up in such a way that any necessary structural changes can be made in one place and then affect all relevant parts of the text basis. In this way, the initial consistency of the structures is retained, even though the structures themselves are changed. The modules for structure levels allow a text to be structured with increasing depth. This ensures that the text basis can be built up in several structuring phases and remains a self-contained unit at the end of

each phase. Another major principle of the module system is its transferability to other projects. According to the project requirements, individual structure levels can be chosen and applied; if the requirements are enhanced, further structure levels can be added.

The metastructure is a semantic net built as a topic map. Fundamental object types in literary editions are persons, locations, and works. The metastructure concept strives to achieve a distinction between different groups of objects, such as:

Objects that model immutable facts, e.g., Thomas Mann was born in Lubeck, Thomas Mann is the author of *Buddenbrooks*, or Heinrich and Thomas Mann were brothers;

Objects that are connected by interpreting relations, e.g., Thomas Mann was a great admirer of Dostoyevsky, or Erika was Thomas Mann's favorite daughter;

Objects with a historical dimension, e.g., "Germany" means different geographical areas in different historical periods.

In this way, a user of the electronic version can clearly distinguish interpretation from pure facts. The metastructures play an important role in the overall concept of the Thomas Mann project. In the final section, we show the perspectives offered by the project.

New perspectives

If one departs from conventional approaches to producing an edition in printed and electronic form, a new concept is required. The Thomas Mann project shows an innovative way to publish one edition in two media and its underlying strategy.

Within the framework of this strategy, an information pool was designed to constitute the publisher's core of information for the works of Thomas Mann. It constitutes part of an overall information architecture that incurs high investment costs in the initial phase. In the long run, these costs are justified by the increased speed in editing and making it more efficient, and also by the opportunity to create spin-off products at lower cost. An information architecture such as this opens up the content - and thereby also the basic publisher's task - for different publication media. It means the future safeguarding of Thomas Mann's works and offers a new way of organizing the edition.

Not only are the printed and electronic versions published from one data source, but they also follow one common content concept that makes them **one edition in two media versions**. Furthermore, each version fully utilizes the specific qualities of its medium, leading to the clear distinction of well-defined user needs. The book is still the best version for pleasurable reading, whereas the electronic version has a clear advantages in information retrieval. The electronic version of the GKFA as a working tool stands out against currently available electronic versions and, therefore, sets new standards. For example, it will be possible to retrieve information on the following questions:

Which newspapers published reviews by Thomas Mann?

In which essays did Thomas Mann give his opinion on other contemporary authors?

In which letters did he write about his Magic Mountain?

Which authors are quoted most by Thomas Mann?

In addition, it will also be possible to retrieve single terms such as "Lubeck". This search possibility goes beyond the potential of simple full-text retrieval by finding all occurrences of "Lubeck", even though the word "Lubeck" is not explicitly mentioned in the text. The hits then can be shown in a categorized list, e.g.,:

fictional location in literature

Thomas Mann's residence

sender location of letters

The tool character of the electronic version is strengthened by the fact that the user can, for example, define electronic index cards and notes attached to text passages which can later be included in search routines.

A concept such as that of the Thomas Mann project is related to a new strategy. The aim of this strategy is to create a basis from which the publisher can react flexibly to the rapidly changing demands of the market. This does not just mean a publication in different media, but also the opportunity to put together and publish at short notice and low cost. In this connection, new types of editorial work are created, which have to be integrated into the publisher's range of tasks both in terms of content and technology.

The Thomas Mann project represents a new starting point for planning literary editions. As we have illustrated, the basic concept is not restricted to literary editions, but can also be transferred to other types of publication projects.

Authors

Ingrid Schmidt

Senior Information Architect
VIA
Postal Address:
Parkstraße 7
69126 Heidelberg
Germany
Telephon: + 49.(0)6221.373266
Fax: +49.(0)6221.374679
E-mail: schmidt@via-iscm.de
Web: www.via-iscm.de

Ingrid Schmidt - In January 2000 Ingrid and Carolin founded VIA. By providing consultancy as a main focus, VIA accompanies firms (especially publishing houses) that want their content resources to outlive rapidly outdated media and thus be prepared for the constantly shifting requirements of the media landscape. Within this context VIA emphasises on the design and realisation of forward-looking SGML/XML-based information architectures.

Ingrid worked since 1993 as independant consultant, information architect, and trainer in the field of SGML/XML-based (and related standards) publishing, for both, industry and research. Moreover, she is regularly teaching classes on SGML/XML at the German Linguistic and the Computational Linguistic department of Heidelberg University. From 1993 to 1998 she worked as free-lancer on different projects with the Institute for Integrated Publication and Information Systems of GMD in Darmstadt. Research focuses were knowledge-based information access for hypermedia reference works and the evaluation of machine-aided, automized semantic encoding possibilities based on an object-oriented database system. From 1991 to 1993 she worked with Texcel as application developer and consultant. From 1988 to 1991 she worked, also in the field of application development and consulting, with Manfred Krüger, later MID/Information Logistics Group in Heidelberg.

Carolin Müller

Information Architect
VIA
Postal Address:
Parkstraße 7
69126 Heidelberg
Germany
Telephon: + 49.(0)6221.373266
Fax: +49.(0)6221.374679
E-mail: mueller@via-iscm.de
Web: www.via-iscm.de

Carolin Müller - In January 2000 Carolin and Ingrid founded VIA. By providing consultancy as a main focus, VIA accompanies firms (especially publishing houses) that want their content resources to outlive rapidly outdated media and thus be prepared for the constantly shifting requirements of the media landscape. Within this context VIA emphasises on the design and realisation of forward-looking SGML/XML-based information architectures.

Carolin works since 1998 as an independant consultant for SGML/XML-based publication environments. Before this she was involved in linguistic research with the main focus on lexikography. Apart from her own studies, she participated in the design of a XML-based lexicological-lexicographical information system, developed by the Institut für Deutsche Sprache (IDS) in Mannheim. Besides she worked for a dictionary project covering the different language stages of Early New High German.