

Aufwand und Tücken der Migration digitaler Editionen auf der Basis eines nicht-standardisierten Datenformats

On the Migration of a Digital Scholarly Edition: Effort and Perils of a non-standardized data format.

Harald Jele

Die „Klagenfurter Ausgabe Robert Musil“ erschien 2009 als digitale Edition. Sie wurde mit der Software FolioViews als eine installierbare Desktop-Anwendung erstellt und auf DVD ausgeliefert.

In den letzten Jahren wurden einige Anstrengungen unternommen, die Inhalte dieser Edition schrittweise in ein standardisiertes Datenformat umzuwandeln. Als Quellformat diente dabei im Wesentlichen das FolioViews eigene Exportformat. Als Zielformat wurde XML-TEI definiert. Bei der Migration wurde darauf geachtet, gespeicherte Informationen möglichst vollständig in das Zielformat zu übernehmen.

Dieser Beitrag beschreibt den Aufwand und die Tücken, die mit einem solchen Migrationsvorhaben verbunden sind. Der Aufwand wird dabei jenem gegenübergestellt, der entsteht, wenn umgekehrt XML-TEI als Ausgangsformat zur weiteren Verarbeitung herangezogen wird. Dieses Format war bereits vorhanden und einigermaßen bekannt, als man sich für eine Umsetzung der Edition mittels FolioViews entschieden hatte. Aus verschiedenen Gründen hatte man sich in der Entscheidungsfindung jedoch gegen einen Einsatz von XML-TEI ausgesprochen.

Das Ergebnis dieses Vergleichs wird nicht überraschen: Der Aufwand, der entsteht, wenn ein nicht standardisiertes Datenformat gewählt wird, ist ungleich höher, wesentlich komplexer und letztlich deutlich kostspieliger. Zudem bleibt zu beachten, dass der dauerhaften Speicherung und Verfügbarmhaltung (im Sinne einer Langzeitarchivierung) solcher Informationen sehr enge Grenzen gesetzt sind.

The release of the „Klagenfurter Ausgabe Robert Musil“ was published in 2009 as a digital scholarly edition, created with the software FolioViews as a typical desktop application and delivered on DVD.

During the last years substantial efforts have been made to convert the contents of this edition, step by step, into a standardized data format. The FolioViews own export format was used as the essential source format. As target format XML-TEI has been defined.

During the migration process, the circumstance not to leave any stored information behind was taken very seriously.

This article describes the effort and the perils associated with such a migration project.

The effort is contrasted with that, which arises when in the reverse case TEI-XML is used as the initial format for further processing. The XML-TEI format was already invented and fairly well known when the decision was made to realize the edition with FolioViews and, for various reasons, against the use of a standardized format like XML-TEI.

The result of this comparison is not surprising: The effort, which arises if a non-standardized data format is selected, is much higher, more complex and finally much more expensive. It also remains to be noted, that the permanent, sustainable storage and availability of such information (in terms of long-term archiving) is restricted to very close limits.
