

Abstract

Metadata is data about data. It provides a brief and accurate description of some artifact or person in one or more specific dimensions. It can be an extract of the base data, i.e. it is already contained within it, but in a more useful form, for instance the topic of a document. It can also be an addition, when the information is not present in the base entity, like the last modification time of a document or the country of living of a person. As metadata is typically formalized to support automatic processing, many tools can be based on it, which would not be able to provide their functionality with the "plain" base data alone, respectively not at all regarding additional information.

In spite of this potential usefulness, metadata is often largely or completely missing. Because of this common lack of prerequisite information in a directly usable form, applications using metadata are rare too. A need exists to improve this situation in both aspects: increasing the presence of metadata, and offering applications providing direct benefits to both creators and users through employing the metadata.

The most important step to fulfill this need is taking into account the complete metadata lifecycle for which a model is proposed in this thesis. Often metadata is just created according to some schema and used. But when including other lifecycle stages too, e.g. evaluation, archiving, and re-purposing, application utility for users and perceived metadata value for authors improve. Through this, metadata presence is increased because it is created and maintained in a systematic process instead of ad hoc. Additionally, incentives for creating tools using metadata increase too, as input data supply is expanded, and result quality improved through defined semantics. This in turn encourages adding metadata, reinforcing the approach.

In this thesis, both aspects are exemplified by several publications, mostly in the area of E-Learning. They highlight possibilities for automatic derivation/harvesting of metadata, for instance deducing user's interests or intentions from metadata keywords of learning material, which are themselves derived automatically from the material's base text if missing. Other examples presented are ideas and implementations for extended metadata usage, e.g. generating graphical course representations or filtering course content to create material subsets.

Another important aim is investigating legal issues connected with metadata, consisting of two aspects: metadata in the legal area (legal texts as base data), and legal regulations of metadata (metadata and its usage as subject of legal discussion). The former is examined through attribute certificates for legal persons and the legal relevance of metadata contained therein. For the latter, the focus here is privacy: Which metadata may be gathered when or under which conditions, and what legally may be done with it. Especially in E-Learning technically much data on learners can be gathered, which might be used to derive metadata on them as individuals or groups for providing services like recommendations or personalization. However, other and possibly unwanted uses, e.g. tracking users across several learning platforms or passing personal information to other learners/teachers, are possible too. The balance between those metadata usage options is based on the law and must be heeded in the implementation of ICT systems.

Through this two-pronged approach it is argued, that both taking into account the whole metadata lifecycle and considering legal aspects are important prerequisites to extend metadata presence and usage. The technical background is necessary for implementing software systems, which additionally must conform to legal regulations to be acceptable for users. Through this, metadata presence and usage can be extended, with the end result of enhanced utility of applications for end-users.