Metainformation scenarios in Digital Humanities: Characterisation and conceptual modelling strategies
Requirements for the analysis, interpretation and reuse of information are becoming more and more ambitious as we generate larger and more complex datasets. This is leading to the development and widespread use of information about information, often called metainformation (or metadata) in most disciplines. The Digital Humanities are not an exception. We often assume that metainformation helps us in documenting information for future reference by recording who has created it, when and how, among other aspects. We also assume that recording metainformation will facilitate the tasks of interpreting information at later stages. However, some works have identified some issues with existing metadata approaches, related to 1) the proliferation of too many “standards” and difficulties to choose between them; 2) the generalized assumption that metadata and data (or metainformation and information) are essentially different, and the subsequent development of separate sets of languages and tools for each (introducing redundant models); and 3) the combination of conceptual and implementation concerns within most approaches, violating basic engineering principles of modularity and separation of concerns. Some of these problems are especially relevant in Digital Humanities.

In addition, we argue here that the lack of characterization of the scenarios in which metainformation plays a relevant role in humanistic projects often results in metainformation being recorded and managed without a specific purpose in mind. In turn, this hinders the process of decision making on issues such as what metainformation must be recorded in a specific project, and how it must be conceptualized, stored and managed. This paper presents a review of the most used metadata approaches in Digital Humanities and, taking a conceptual modelling perspective, analyses their major issues as outlined above. It also describes what the most common scenarios for the use of metainformation in Digital Humanities are, presenting a characterization that can assist in the setting of goals for metainformation recording and management in each case. Based on these two aspects, a new approach is proposed for the conceptualization, recording and management of metainformation in the Digital Humanities, using the ConML conceptual modelling language, and adopting the overall view that metainformation is not essentially different to information. The proposal is validated in Digital Humanities scenarios through case studies employing real-world datasets.