Supporting an interdisciplinary research agenda through meta-modelling. The case of Living with machines

Giorgia Tolfo*[1] and Timothy Hobson*[2]

[1]British Library – United Kingdom
[2]The Alan Turing Institute – United Kingdom

Abstract

Living with Machines is an interdisciplinary research project focusing on the interactions between machines and humans during the long nineteenth century. The heterogeneity of the source materials (digitised newspapers, books, census data, maps, parliamentary papers), the variety of enrichments and annotations produced, the asynchronous availability and access to the data and the diverse research interests of scholars within the project pose a number of different problems when approaching data modelling. We therefore need or both a methodology and a meta-model that reconcile these different needs keeping in sight the traditional FAIR data principles.
Who (or what) is an *agent* for a research software engineer, a computational linguist or a historian? How can we model *place* when we have both newspapers with places of publication and areas of diffusion, census data with places of birth and death or maps with place names that may vary over time? How can we model a newspaper article when semantically it's a unit of meaning and physically a group of boxes spanning over multiple pages? The *LWM metamodel* that we will present in this poster is an attempt to answer such questions. It is a conceptual data model which attempts to support/incorporate and integrate in a single model the different datasets to improve reusability and sustainability and to prevent research tasks and outcomes from becoming siloed. Building on concepts like *intellectual entity* and *representation* borrowed from the PREMIS standard, our metamodel aims to describe digitised documents from multiple perspectives – physical characteristics, semantic structure and metadata concerning the processes through which the representation came into being – and to use this tripartite structure to integrate other currently available standards. The purpose of the metamodel itself is to promote the integration of research activity and the interoperability of software and data, both within the project and beyond, but also to accommodate the ambiguity, uncertainty and incompleteness that typically characterise these datasets and outputs.

The poster delineates the current status of the metamodel and the methodology devised in order to produce it. It will also point to specific challenges and limits of modelling in a context where inputs and outputs can't be determined and fully outlined at the beginning of a digital humanities project.

*Speaker

Keywords: metamodel, interoperability, standard, PREMIS, data reconciliation, newspapers, representation, conceptual data model