

Using Semantic Technologies to Create Virtual Families from Historical Vital Records

Christophe Debruyne^{1,2}, Oya Beyan¹, Stefan Decker¹ and Sandra Collins²

¹Insight @ NUI Galway, Galway, Ireland

²Digital Repository of Ireland, Royal Irish Academy, Dublin, Ireland

We report on the semantic architecture and ontology creation of the Irish Record Linkage 1864-1913 project, which aims to create a platform to reconstitute families and create longitudinal health histories by applying semantic technologies to annotate, store and analyze the data contained in historical birth-, marriage- and death records. This enables researchers to, for instance, investigate to what extent maternal and infant mortality rates were underreported.

We make a clear distinction between (i) the curation of the encoded data contained in the records as well as the long-term preservation thereof, managed by a digital archivist, and (ii) the analysis and interpretation of that data to answer specific questions for historians and researchers. To support the two processes and to maintain the clear separation of concerns, two distinct but interrelated knowledge bases (KBs) are developed. The knowledge engineer is responsible for setting up the semantic infrastructure, and interprets the research questions of the historians into queries for the enriched KB.

A first KB is setup to contain the in RDF encoded records using “flat” ontologies that capture the information contained in those records in a lexical manner. While encoding, any noise such as errors or missing values will be respected to preserve the original historical record and provenance. The forms of these records were “lifted” into an ontology, as available ontologies for vital records were, to the best of our knowledge, non-existent.

We then populate more expressive ontologies capturing domain knowledge for creating the second and richer KB from the RDF records. Existing ontologies such as the Persona Vocabulary were adopted and extensions specific to this project were developed. This KB was partly populated by adopting ontology matching techniques to detect correspondences across different records and by creating links to external datasets for additional contextual information (e.g., Logainm.ie, the authority database for historical and contemporary Irish place names). Additional knowledge includes medical information expressed as rules based on the International Classification of Diseases. This allows historians to analyze the content in the records using different parameters and different assumptions, e.g., by using different or newer bodies of medical knowledge.

We thank the Registrar General of Ireland for permitting us to use this rich digital content contained in the vital records. This publication has emanated from research conducted within the Irish Record Linkage, 1864-1913 project supported by the RPG2013-3; Irish Research Council Interdisciplinary Research Project Grant and within the Science Foundation Ireland Funded Insight Research Centre (SFI/12/rc/2289). The Digital Repository of Ireland (formerly NAVR) gratefully acknowledges funding from the Irish HEA PRTL program.