

Are Online Genealogical Data Suitable for Demographic Research? Evidence from FamiLinx

Andrea Colasurdo^{1,2,*} and Riccardo Omenti^{3,†}

¹Max Planck Institute for Demography Research

²University of Groningen

³University of Bologna

*corresponding email: colasurdo@demogr.mpg.de

†corresponding email: riccardo.omenti2@unibo.it

Abstract

FamiLinx is an online genealogical dataset with information on profiles and their kinship ties spanning multiple centuries and countries, providing key demographic information from historical periods lacking ground-truth demographic data. Nonetheless, the dataset is affected by several weaknesses, including a high heterogeneity in the completeness and quality of demographic variables and family-related information as well as numerous biases arising from the construction of genealogical trees. This study aims to assess the distribution of missing values and the accuracy of demographic information, at both individual and family network levels, to highlight the major potential of online genealogical data for demographic research. We observe 1) positive correlation in the availability of different demographic variables, and 2) higher completeness and quality in kinship networks of profiles with non-missing demographic information and more precise dates. Our findings provide valuable insights to researchers willing to use FamiLinx for their own studies.