PLAYFAIR PROJECT: CHALLENGES AND SOLUTIONS TOWARDS BUILDING AND VISUALISING FAIR DATA FOR TRADITIONAL GAMES

PAPER SUBMISSIONS FOR DARIAH ANNUAL EVENT 2022: STORYTELLING

Carlos Utrilla Guerrero*
Institute of Data Science (IDS)
Maastricht University (UM)
Paul-Henri Spaaklaan 1
c.utrillaguerrero@maastrichtuniversity.nl

Vincent Emonet
Institute of Data Science (IDS)
Maastricht University (UM)
Paul-Henri Spaaklaan 1
vincent.emonet@maastrichtuniversity.nl

Matthew Stephenson
Department Knowledge and Engineering
Maastricht University (UM)
Paul-Henri Spaaklaan 1
matthew.stephsenson@maastrichtuniversity.nl

Walter Crist
Department Knowledge and Engineering
Maastricht University (UM)
Paul-Henri Spaaklaan 1
walter.crist@maastrichtuniversity.nl

Cameron Browne
Department Knowledge and Engineering
Maastricht University (UM)
Paul-Henri Spaaklaan 1
cameron.browne@maastrichtuniversity.nl

Michel Dumontier
Institute of Data Science (IDS)
Maastricht University (UM)
Paul-Henri Spaaklaan 1
michel.dumontier@maastrichtuniversity.nl

March 7, 2022

*Submitting Author

ABSTRACT

Like almost all research disciplines, digital humanities is poised to enter an era of unprecedented large scale analysis powered by massive amounts of (public) digital collections and hundreds of millions of records on the web [1]. However, this rising amount of humanities data is largely unstructured, making it nearly impossible to connect to other datasets for better analysis, and in some cases even a shortage of usefulness or reusability.

An often invisible, but crucial part of the data storytelling lifecycle is the transformation of digital rudimentary data into intelligible information, namely data modeling. Historians, academics and data scientists are stymied in their ability to find, access and reuse this digital gold to produce statistical analyses and visualisations. Several studies have proposed semantic web technologies [2] and FAIR [3] approaches as a set of recommended solutions supporting better computational approaches, data storytelling and reuse. PLAYFAIR [4] is concerned with how semantic web technologies can facilitate statistical analysis and visualisations on traditional games from various available sources and formats, in a universal and FAIR manner, and subsequently enhance data published on the Web in digital humanities for data storytelling.

This paper aims to present our chosen methodology and the challenges we faced when building a Knowledge Graph (KG) from the ERC-Digital Ludeme Project (DLP) [5] using semantic tools such as CLARIAH public services [6]. The DLP is constructing a unique database of historical evidence for traditional games, which can be used to model the evolution of games throughout history. The games within this database are described in terms of distinct ludemes, alongside auxiliary information such as rulesets, periods, evidences, regions and categories [7].

We specifically describe the process of building an online resource to explore the historical context of traditional games. We introduce the data model using established standards, in particular *CIDOC-CRM* and *schema.org* ontologies, for supporting data interoperability and longevity, as well as providing stable digital representations of traditional games. We then present the design and implementation of the KG using CLARIAH tools, which allow historians to transform source data into computable representation as well as link these to other resources (i.e. British Museum digital collection). We also present the online visualisation tool [8] utilised to analyse and visualise clusters of ancient games. Entirely based on semantic web technologies, this tool is used to publish, access, and visualise research outputs. It also exposes a SPARQL endpoint and several other services to generate and share queries (e.g. data exploration and visualising high-dimensional data), enabling the 'data storytelling' to be FAIR. The KG and related visualisation tool turned data into a online storytelling resource that can be located by the community and thus maximise reusability. We conclude with a discussion of challenges we are facing at each step of the process and related methodology implementation. Additionally, we propose design recommendations for effective methods for data modeling and visualisation, which can be used to facilitate FAIR storytelling and subsequent data reuse. We believe that this paper will be of interest to humanities projects that use visual analytics as part of their research process. The methodology outlined here takes the realities of many digital humanities research projects into account and could act as a blueprint for further discussions rooted in the digital humanities community.

This work is funded by the CLARIAH-PLUS W4 *(Cp-21-f-i5)* national project and aims at the realisation of a common infrastructure for the humanities and social sciences.

References

[1] INTEF. Big data, for better or worse: 90 percentage of world's data generated over last two years. *www.sciencedaily.com/releases/2013/05/130522085217.htm*, 2013.

[2] Bernhard Haslhofer, Antoine Isaac, and Rainer Simon. Knowledge graphs in the libraries and digital humanities domain. *Encyclopedia of Big Data Technologies*, page 1–8, 2018.

[3] Mark D Wilkinson, Michel Dumontier, and et al. Aalbersberg. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3, 2016.

[4] Grant number: Clariah-plus wp4: Cp-21-f-i 5 - fair data for historical games. Read the Interview: http://lnkd.in/dwWm8pT4.

[5] C Browne. Ai for ancient games. https://doi.org/10.1007/s13218-019-00600-6, 2020.

[6] Bernhard Haslhofer, Antoine Isaac, and Rainer Simon. Clariah tools reference. *Encyclopedia of Big Data Technologies*, page 1–8, 2018.

[7] Matthew Stephenson et al. Digital ludeme project database guide - (erc consolidator grant 771292). https://ludii.games/downloads, 2021.

[8] Rinke Hoekstra, Albert Meroño-Peñuela, Auke Rijpma, Richard Zijdeman, Ashkan Ashkpour, Kathrin Dentler, Ivo Zandhuis, and Laurens Rietveld. The datalegend ecosystem for historical statistics. *Journal of Web Semantics*, 50:49–61, 2018.