

# Improving metadata quality assessment in public health and epidemiology

Christiana McMahon<sup>a,b,1</sup>, Tito Castillo<sup>a,c</sup>, Dennis Kehoe<sup>d</sup>, Spiros Denaxas<sup>a,b</sup>

<sup>a</sup>*Farr Institute of Health Informatics Research at UCL, London, United Kingdom*

<sup>b</sup>*Institute of Health Informatics, University College London, United Kingdom*

<sup>c</sup>*Xperimint Ltd. London, United Kingdom*

<sup>d</sup>*AIMES Grid Services CIC, Liverpool, United Kingdom*

**Keywords.** Public Health Informatics, Epidemiology, Quality Improvement

The increase in generation of, and access to, public health and epidemiology data necessitates robust mechanisms to produce data documentation - metadata. Metadata often describes research data and the process through which the data was collected. Secondary use of clinical data sourced from electronic health records can greatly benefit from access to metadata and other such documentation.<sup>1</sup> However, metadata quality assessments often fail to form a routine part of metadata administration.<sup>2</sup> The aim of this work is to review different methods of metadata quality assessment and improve understanding of current practices in public health and epidemiology research.

The review focused on identifying pre-existing methods of metadata quality assessment including approaches to defining quality dimensions.

A total of 11 approaches to metadata evaluation were identified during the review with totality, accuracy and accessibility being the three most commonly identified principles of quality. A review conducted in 2013 by Tani and Candela<sup>3</sup> was also identified.

Results of the review indicate whilst there are pre-existing approaches to metadata quality assurance, scope remains for public health and epidemiology-specific guidance.

The next step is to circulate a web-based survey aimed at stakeholders in public health and epidemiology to identify current challenges and areas of improvement. This will inform development of a conceptual model to support improved metadata quality assessment.

1. Pathak, J., et al., Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPN consortium. *Journal of the American Medical Informatics Association*, 2013. **20**(e2): p. e341-e348.
2. Hillmann, D.I., Metadata quality: From evaluation to augmentation. *Cataloging & Classification Quarterly*, 2008. **46**(1): p. 65-80.
3. Tani, A., L. Candela, and D. Castelli, Dealing with metadata quality: The legacy of digital library efforts. *Information Processing & Management*, 2013. **49**(6): p. 1194-1205.

---

<sup>1</sup> Corresponding Author: Address: 222 Euston Road, London, NW1 2DA, United Kingdom. Tel: +44 (0)203 549 5338. Email: christiana.mcmahon.11@ucl.ac.uk