

## Population Data Science: The science of data about people

McGrail, K<sup>1</sup> and Jones, K<sup>2</sup>

<sup>1</sup>University of British Columbia

<sup>2</sup>Swansea University

### Introduction

Societal and individual benefits of data-intensive science are substantial but raise challenges of balancing individual privacy and public good, while building appropriate governance and socio-technical systems to support data-intensive science. We set out to define a new field of inquiry to move collective interests forward.

### Objectives and Approach

Our objectives were: 1. To create a concise definition of the emerging field of Population Data Science; 2. To highlight the characteristics and challenges of Population Data Science; 3. To differentiate Population Data Science from existing fields of data science and informatics; and 4. To discuss the implications and future opportunities for Population Data Science. Objectives 1 and 2 were met largely through International Population Data Linkage Network (IPDLN) member engagement, Objective 3 was evaluated via literature review, and Objective 4 was achieved through iterative and collective work on a peer-reviewed position paper.

### Results

We define Population Data Science succinctly as the science of data about people. It is related to, but distinct from, the fields of data science and informatics. A broader definition includes four characteristics of: i) data use for positive impact on individuals and populations; ii) bringing together and analyzing data from multiple sources; iii) identifying population-level insights; and iv) developing safe, privacy-sensitive and ethical infrastructure to support research. One implication of these characteristics is that few individuals or organisations possess all of the requisite knowledge and skills comprising Population Data Science, so this is by nature a multi-disciplinary “team science” field. There is a need to advance various aspects of science, such as data linkage technology, various forms of analytics, and methods of public engagement.

### Conclusion/Implications

These implications are the beginnings of a research agenda for Population Data Science, which if approached as a collective field, will catalyze significant advances in our understanding of society, health, and human behavior and increase the impact of our research.

