Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 633–640

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Workshop on Privacy and Security in Healthcare (PSCare 2017)

Privacy preserving distributed computation of community health research data

Anders Andersen[a], Merete Saus[b]

[a]*Department of Computer Science, UiT The Arctic University of Norway, 9037 Tromsø, Norway*
[b]*RKBU North, UiT The Arctic University of Norway, 9037 Tromsø, Norway*

Abstract

Research in community health introduces challenges regarding analysis of the research data. It involves multiple actors in a varity of arenas, and it is often directed towards the local community and children and their families. The legal, ethical and privacy issues involved introduce constraints upon the analysis performed. SNOOP combined with the D^{2}Worm declarative modelling and infrastructure architecture is a promising approach to support a wide range of possible privacy preserving analysis in community health research.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Community health data; Analysis; Distributed Computations; Secure Multiparty Computations; PKI

1. Introduction

Community health research is a complex research field, due to the numbers of different possible factors concerning one case. The number of actors involved might differ, the multiple variations of relevant arenas, and even the informants' involvements and role might vary within and among the cases. Typical in community health research, the problem is about prevention or rehabilitation, and the health community directed their intervention towards both the patients and the surroundings. It might be the family, the network, the health and social services, or community based institutions. Community health services aimed at children will often involve schools, kinder gardens, and even leisure activities. All of these arenas, with all of these different actors, might be resources in community health research. It means that there are legal, ethical, practical and privacy issues involved when collecting and analyzing data in health community research, and these issues introduce a set of constraints upon the computations and implementations [1]. In this paper, we will suggest a solution that can handles this complexity. We will focus on the practical approach to meet these challenges using SNOOP and data-centric workflow modelling (SNOOP is just a name and not an abbreviation for anything). We will first introduce SNOOP as a privacy conserving distributed computation platform that can be used to perform SMC (Secure Multiparty Computation) algorithms to analyse community health data. We will

* Corresponding author. Tel.: +47-776-44703 ; fax: +47-776-44580.
E-mail address: Anders.Andersen@uit.no

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.319

then discuss how workflow modelling and D²Worm (Distributed Data-centric WORkflow Management system) can introduce a high level data-centric modelling of such computations and the challenges related to privacy preserving processing with such an approach.

2. Community health research

Internationally there is a growing interest in community health services at the expenses of specialised health service[2,3]. A simplified, but descriptive classification is that the specialized health services focus on treatments and targeted interventions while the community health service focus on prevention and rehabilitation. A specialised health service system is typically individual oriented and takes place in hospitals or other treatments centers. Prevention and rehabilitation in community health, however, take place locally, where the people live their lives. Intervention for prevention and rehabilitation is often oriented towards the public, and aims often collective. Despite that specialized health workers might outline the intervention in community health service, the performer of the health services might varies. It can be the teachers, the social workers, or nurses in community health centers that actually implements the health interventions. The community health research have to mirror these variations of possible actors and arenas when studying these interventions.

In community health, there has been an emphasis towards the policy of investments of the wellbeing for children, and the children's prospects to a healthy, productive, and meaningful life where they can fulfil their potential[4]. One of the reason is that is has proven economical profitable for a society to invest in children: *"The evidence is quite clear that inequality in the development of human capabilities produces negative social and economic outcomes that can and should be prevented with investments in early childhood education, particularly targeted toward disadvantaged children and their families"*[5]. Investments in early childhood is also demonstrated as efficient for adult health[6]. To prevent unwanted prospects is cheaper than the price of treating and caring for a life that does not fulfill its potential. For community health research, this insight have directed the research towards the intervention that it targeted to the youngest population. In doing so, the research face a number of challenges that make the data computation complicated. The childrens age might make it necessary to involve others informants on their behalf. When and how the children have to be involved are controlled by different ethical regimes. The difference in the childrens age might also affect the way the inquiry is outlines, meaning that addressing the same issues might need multiple questionnaires. This is a complexity in doing community health research that makes it resource demanding in addition to the ethical challenges.

In sum, the local focus in community health and that prevention is targeting children as a population, are aspects that provides challenges for research in community health. Collecting and analysing data is more challenging. In specialized health service in hospitals, the patients come to the researchers. In community health research, the researchers have to go to local communities and visit the patients in their own environments. The researchers have to handle that the possible informants for an evaluation of an intervention might be differ and that the questionnaire might be in multiple variations. These challenges are the starting point for our discussion concerning infrastructure and services to facilitate research in the complexity of community health service.

3. Data computation in community health research

Important aspects of data in community health research are that (i) the data might contain *sensitive personal information*, (ii) the data is collected from *a wide range of sources*, and (iii) the data at-rest is *distributed*. Data with sensitive information about patients, research subjects, or informants raises privacy concerns, and access to the data has to be tightly controlled. By combining data from several sources, more knowledge about individuals and groups of people can be gained. Both this new knowledge and intermediate results from such computations might be sensitive and should be included when privacy concerns are analysed.

Data about a single patient, a research subject, or an informant, might be distributed among several nodes (data servers with vertically partitioned datasets). With community health research data, vertically partitioned datasets at-rest are typically distributed over a wide range of institutions, including hospitals, general practitioners, specialist, labs, and social service offices. In horizontally partitioned datasets, one type of data about a large number of patients, research subjects, or informants might be distributed among several nodes. A typical example is data collected by