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.

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 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.319

then discuss how workflow modelling and D^{2}Worm (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

Table 1. The notation used for messages, encryption and signing in this text and in earlier papers on SNOOP[16,17].

(a, b, c)	A group with the elements a, b, c	$\{m\}^n$	m signed by n
$\{m\}$	A message containing m	$\{m\}_p^n$	m signed by n and encrypted with public key p
$\mathrm{s}\{m\}$	m encrypted with secret key s	$\{n, p\}^c$	CA c binds public key p to identity n
$\{m\}_p$	m encrypted with public key p	$A \rightarrow B : \{m\}$	Message $\{m\}$ sent from A to B

general practitioners. Each patient is bound to one general practitioner, but for a large number of patients many general practitioners might be the source of that type of data. In real examples, both vertically and horizontally partitioned datasets might exist, and legal, ethical, and privacy aspects of managing those datasets have to be respected. These aspects might enforce local processing of data at-rest at the node.

Privacy preserving distributed analysis of community health research data can be achieved by combining cryptography, suitable algorithms, constrain specification and enforcement, explicit workflow models (that can be analysed for privacy concerns), and carefully designed workflow run-times. Cryptography is used to protect data at-rest and in-transit. Symmetric encryption with unique encryption keys is a flexible and efficient way to ensure confidentiality. Combined with public-key encryption and Public Key Infrastructure (PKI), the confidentiality provided by symmetric encryption can be extended with data integrity and secure sharing of data[7]. The algorithms used in the data analysis can contribute to the privacy preserving part of data processing. In the book chapter *Privacy preserving personalisation in complex ecosystems*[8], the privacy preserving processing in the context of personalisation is discussed. This can be directly mapped to privacy preserving processing in the context of community health research data. Local processing of data at-rest, de-identifying data (e.g. differential privacy[9,10]), Secure Multiparty Computation (SMC)[11], and homomorphic encryption[12], are examples of algorithmic approaches to privacy preserving data analysis. These approaches can be used individually or combined. In some cases, constrains have to be specified and enforced to obstruct privacy violations. An example is re-identifying data when the number of individuals in the dataset is small.

The process of analysing distributed community health research data can be modeled as workflows[13]. Data-centric workflow modeling (as opposed to flow-based) promises a flexible and adaptable approach to model and create such processes[14]. The data focus of data-centric approaches is a good basis for conducting the legal, ethical and privacy concerns of community health data research. Since the focus is on the data, constraints and protective operations related to these data could be included in the workflow modeling.

SNOOP is a middleware built to support the constructions, deployment and execution of applications performing analysis of sensitive distributed data. SNOOP supports contract based deployment of components in SNOOP run-times. The contracts are in SNOOP used to match the software component requirements with the run-time resources and requirements. At deploy time the component and the run-time tries to fulfill the contract. If succeeded the component is deployed and activated. The contract is also used to explicit specify what data, services and resources the component in the given context can access at the host it is deployed. Operations executed at a single host are a subset of a complete data analysis. The host is typically a general practice or a hospital, and it is a participant in the data analysis. The contract of a component includes a signed delegation from an approved authority that in a given context grants access to the specified data to perform the operations executed by the component.

The D^{2}Worm[14] infrastructure is used to model data-centric workflows. It is based on the Guard-Stage-Milestone (GSM)[15] meta-model for lifecycles. Current data-centric modelling approaches (including GSM) do not provide syntactical mechanisms to restrict data exchange across organisational boundaries and existing workflow management systems for data-centric workflows are not capable to enforce data privacy.

4. SNOOP

Before we continue, a short introduction to SNOOP is given. A more detailed introduction to SNOOP is available in other papers[16,17]. With SNOOP, a typical approach to fulfill the privacy requirements is a combination of SMC algorithms and careful usage of cryptography. It is based on a coordinator that prepares the computation and a set of sub-processes representing the parties in the multi-party computation. The coordinator and the sub-processes are nodes in a computation graph. The directed edges of the graph are the messages sent between the nodes. Each node has an identifier (address) and a unique public/private encryption key pair.

The combination of SMC algorithms and public-key encryption (in combination with symmetric key encryption) ensure that each node is unable to learn about the other nodes local data, input data and intermediate results. PKI and its certificate authorities (CAs) ensure that the participants can distribute and trust public keys. PKI enables public-keys as the tool to authenticate participants and maintain the integrity and privacy of the data exchanged.

4.1. Computing graph

The computing graph for a computation is represented as a set of layered messages, where each layer in these messages exposes the next edges in the directed graph. The computation is initiated by the coordinator sending these messages to the first set of nodes. At each node one layer of the received message is decrypted exposing both the input data set for the calculation performed at this node, and the identifier and public-key of the next nodes in the computing graph. The calculation is performed using the input data set and local data available at this node.

When the calculation is done the node generates a set of messages forwarded to the next nodes in the computing graph. The public-keys are used to encrypt the messages. The data sets included in these messages are based on the result of the performed calculation. The notation used is described in Table 1.

A node n_a will receive and unwrap a set of messages signed by the senders n_i and containing input data sets I'_a and data blobs B'_a:

$$n_i \rightarrow n_a : \{\, \{I'_a\}_{p_a},\ B'_a \,\}^{n_i}$$

The input data set I'_a is encrypted with the public key p_a of n_a is a subset of I_a. I_a represents sufficient data to perform the calculation at node n_a. It might be aggregated from a set of input messages containing subsets I'_a of the data. The data blobs B'_a originate at the coordinator c. B_a represents the view of the computing graph from node n_a. It might be aggregated from a set of input messages containing subsets B'_a of the data. In most cases, a single B'_a is equal to the complete view B_a. The data blobs have the following structure:

$$B'_a = \{\, \mathit{input\ node\ list},\ \mathit{output\ node\ list},\ \mathit{meta\ data} \,\}^{c}_{p_a}$$

The *input node list* includes all the nodes that node n_a should expect input from. In many use-cases this list contains a single node, the node that this message was received from. The input node list is used for two things: (i) to verify that the coordinator intended this input to the node, and (ii) to inform the node what input to wait for before the computation is performed. If node n_a in our example should only expect input from node n_i, the input node list would be the single element (n_i, p_i). This is used in node n_a to verify that it was the coordinator's intention that node n_a should receive this input from node n_i.

The *output node list* describes the next nodes in the computing graph. It lists the nodes receiving the intermediate results calculated in this node. For each node a data blob generated specifically for that node by the coordinator is also included. If node n_a in our example is supposed to forward its intermediate results to the three nodes n_o, n_p, and n_q, the output node list will be this:

$$(\, (n_o, p_o, B_o),\ (n_p, p_p, B_p),\ (n_q, p_q, B_q) \,)$$

Figure 1 shows node n_a with an *input node list* with a single element n_i and an *output node list* with the element n_o, n_p, and n_q.

The *meta data* contains information needed to perform the computation at this node and to ensure progress if anythings fails. The meta data received at node n_a is denoted E_a. More details on the significance and usage of the meta-data are found in other SNOOP papers[16,17].

Based on the example described above, a B'_a will have the following structure:

$$B'_a = \Big\{\, (n_i, p_i),\ (\, (n_o, p_o, B_o),\ (n_p, p_p, B_p),\ (n_q, p_q, B_q) \,),\ E_a \,\Big\}^{n_c}_{p_a}$$

4.2. Processing

Each node n_a will perform its calculation f using the received data set I_a and its local data set:

$$R_a = f(I_a)$$

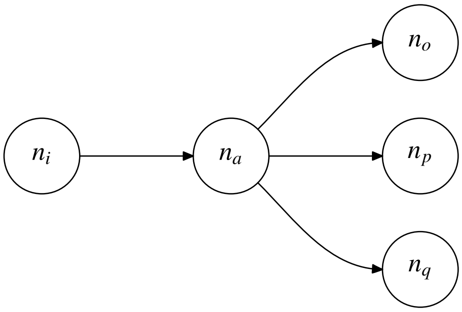

Fig. 1. The *input node list* of n_a is n_i and the *output node list* of n_a is n_o, n_p, and n_q.

F is a filter function that removes data from a data set that should not be forwarded to a given node. $F(R_a, n_o)$ produces a new data set I'_o where only the data that should be available for node n_o is present:

$$I'_o = F(R_a, n_o)$$

From node n_a all the next nodes n_x in the computing graph (n_o, n_p, and n_q in the example above) are forwarded the following message:

$$\{\,\{I'_x\}_{P_x},\ B_x\,\}^{n_a}$$

A signed message where the filtered intermediate results I'_x from the calculation on this node are the input data for the next nodes n_x. The input data is encrypted with the public key of the receivers. All data blobs B_x originate from the coordinator and are forwarded unmodified to the nodes.

An example SMC-based privacy preserving computation using a coordinator n_c and three nodes n_1, n_2, and n_3 are shown in Figure 2. The x values are the local values and the c values are used to count the number of values that is part of the mean calculation. Each node has a sensitive local data value (x_1, x_2, and x_3 respectively) that should participate in the calculation of the mean value m. The coordinator starts by generating two large random numbers r_0 and c_0. The large random number r_0, combined with the encryption of the input data to the nodes, ensure that participating nodes (and others) are unable to deduce the sensitive local data from previous nodes in the computing graph. The large number c_0 is used to hide what number the current node is in the computation chain and the total number of nodes involved (this is not necessary to protect the local data at each node and is used only to reduce the information spread about the current computation). The actual computing graph G for the calculation of the mean value m specified in the format of the data blobs is:

$$G = \{\,\varnothing,\ (n_1, p_1, B_1),\ E_G\,\}^{n_c}_{p_c}$$

$$\begin{array}{rcl|rcl} B_1 &=& \{\,(n_c, p_c),\ (n_2, p_2, B_2),\ E_1\,\}^{n_c}_{p_1} & B_3 &=& \{\,(n_2, p_2),\ (n_c, p_c, B_c),\ E_3\,\}^{n_c}_{p_3} \\ B_2 &=& \{\,(n_1, p_1),\ (n_3, p_3, B_3),\ E_2\,\}^{n_c}_{p_2} & B_c &=& \{\,(n_3, p_3),\ \varnothing,\ E_c\,\}^{n_c}_{p_c} \end{array}$$

Each blob contains the three values *input node list*, *output node list* and *meta data*. G is the overall graph and also represents the starting point of the computation. Therefor, its *input node list* is empty. B_c represents the end point of the computation and its *output node list* is empty. Each subpart (data blob) of the computing graph is signed by the coordinator and encrypted with the public key of the node that has to access (and interpret) this information.

5. Workflow modeling

The focus on data in data-centric workflow modeling is a good basis for our focus on privacy preserved analysis of community health research data. The D²Worm infrastructure's Guard-Stage-Milestone (GSM) approach to model workflows compromises of a logical *information model* and a declarative *lifecycle model*. The information model contains two distinct sets of attribute types: (1) data attributes represent application-level data, and (2) status attributes

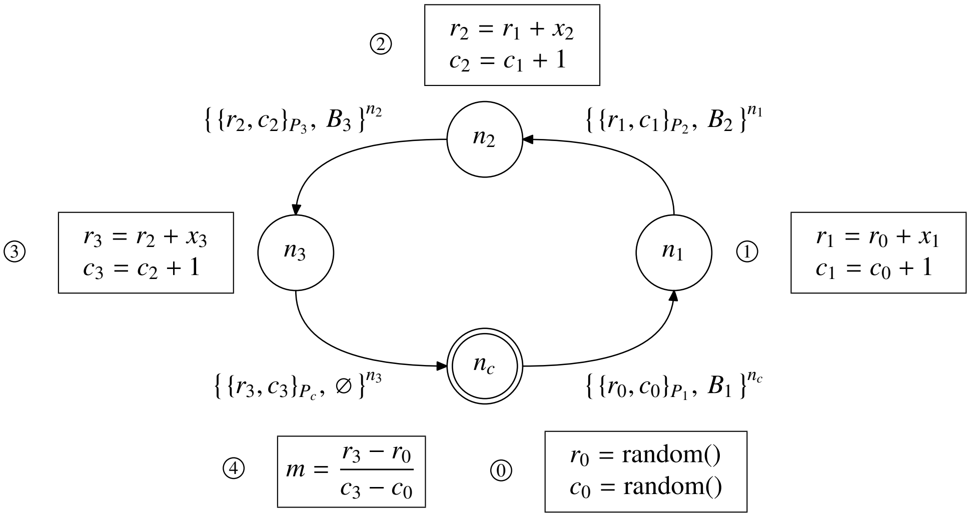
②
$r_2 = r_1 + x_2$
$c_2 = c_1 + 1$
$\{\{r_2, c_2\}_{P_3}, B_3\}^{n_2}$
n_2
$\{\{r_1, c_1\}_{P_2}, B_2\}^{n_1}$
③
$r_3 = r_2 + x_3$
$c_3 = c_2 + 1$
n_3
n_1
①
$r_1 = r_0 + x_1$
$c_1 = c_0 + 1$
n_c
$\{\{r_3, c_3\}_{P_c}, \varnothing\}^{n_3}$
$\{\{r_0, c_0\}_{P_1}, B_1\}^{n_c}$
④
$m = \frac{r_3 - r_0}{c_3 - c_0}$
⓪
$r_0 = \text{random}()$
$c_0 = \text{random}()$

Fig. 2. Privacy preserving calculation of the mean value m from the sensitive local data x_1, x_2, and x_3 at nodes n_1, n_2, and n_3, respectively.

describe the current state of the process according to its lifecycle. For the specification of the lifecycle model, GSM provides three major building blocks: (1) *stages* hierarchically cluster the individual process tasks (aka activities). A task definition in GSM requires the specification of input and output parameters, both taken from the information model. Every stage can have two distinct states, *opened* and *closed*. A task enclosed within a stage can be only executed if the stage is opened. Every stage has at least a single (2) *guard* that control when to open it. An opened stage is intended to achieve one of the (3) *milestones* associated with it. Milestones represent business-relevant objectives that can have two distinct states, *achieved* or *invalidated*.

The following example is based on the SMC-based privacy preserving example discussed above and presented in Figure 2. One problem with the previous example is that the number of participating nodes are low. Statistical analysis on small samples of data increase the possibility to use analytics to expose single sample values. To avoid such privacy concerns we can introduce threshold values on the number of samples in the data set before we are allowed to perform statistical analysis on them. In D^{2}Worm we introduce conditions in the guard of stages to avoid performing the calculation when the number of samples are to small.

Instead of using the standard graphical notation for GSM modelling, we have created our own more compact notation that more easily can be used in SNOOP context.

The example includes one stage S_i, where $i \in [1 .. (n-1)]$, for each such node. In addition, one initial stage S_0 and one final stage S_n are necessary in the example. The example use the following notation for each stage:

$$\underline{S : \langle g \rangle \rightsquigarrow M : (m)}$$

S are the label of the stage, g is the guard, M is the label of the milestone, and m is the milestone values. A stage can have multiple guards and milestones:

$$\underline{S : \langle g_1 \rangle \mid \langle g_2 \rangle \rightsquigarrow M_1 : (m_1) \mid M_2 : (m_2) \mid M_3 : (m_3)}$$

In each stage the actual computation is listed. The conditions of the guards (sentries) are boolean expressions. If a stage is *reached* or *completed* can be included in these expressions. $\oplus S$ is true if we have reached stage S, and $\ominus S$ is true if stage S is completed.

Variables in bold font are representing local (and possible sensitive) data. Examples of such data are the private encryption key of the node and local data that might be accessed to perform the computation (see for example $\mathbf{Priv}_1$ and $\mathbf{v}_1$ in Figure 4).

Figures 3, 4 and 5 illustrates how the calculation of the mean value m could be modelled in D^{2}Worm. In this example we are able to introduce a threshold value for the number of participating nodes (number of stages) that has to be reached before the mean value is calculated. For privacy concerns, this approach can be used to avoid performing statistical analysis on to small data sets.

$$\underline{S_0 : \langle \text{calculate} \rangle \rightsquigarrow M_0 : (x_1^e, c_1^e, x_0^e, c_0^e)}$$

$$\begin{aligned}
x_0 &= \text{random}()\\
c_0 &= \text{random}()\\
x_1^e &= \text{encrypt}(\text{Pub}_1, x_0)\\
c_1^e &= \text{encrypt}(\text{Pub}_1, c_0)\\
x_0^e &= \text{encrypt}(\text{Pub}_{n'}, x_0)\\
c_0^e &= \text{encrypt}(\text{Pub}_n, c_0)
\end{aligned}$$

Fig. 3. Initial stage S_n of privacy preserving calculation of mean value.

$$\underline{S_1 : \langle \ominus S_0 \rangle \rightsquigarrow M_1 : (x_2^e, c_2^e)}$$

$$\begin{aligned}
x_1 &= \text{decrypt}(\mathbf{Priv}_1, x_1^e) + \textstyle\sum \mathbf{v}_1\\
c_1 &= \text{decrypt}(\mathbf{Priv}_1, c_1^e) + |\mathbf{v}_1|\\
x_2^e &= \text{encrypt}(\text{Pub}_2, x_1)\\
c_2^e &= \text{encrypt}(\text{Pub}_2, c_1)
\end{aligned}$$

$$\underline{S_i : \langle \ominus S_{i-1} \rangle \rightsquigarrow M_i : (x_{i+1}^e, c_{i+1}^e)}$$

$$\begin{aligned}
x_i &= \text{decrypt}(\mathbf{Priv}_i, x_i^e) + \textstyle\sum \mathbf{v}_i\\
c_i &= \text{decrypt}(\mathbf{Priv}_i, c_i^e) + |\mathbf{v}_i|\\
x_{i+1}^e &= \text{encrypt}(\text{Pub}_{i+1}, x_i)\\
c_{i+1}^e &= \text{encrypt}(\text{Pub}_{i+1}, c_i)
\end{aligned}$$

Fig. 4. Stage S_1 and S_i, where $i \in [2 \mathinner{..} (n-1)]$.

$$\underline{S_n : \langle \ominus S_{n-1} \rangle \rightsquigarrow M_{n_1} : (\text{error}) \mid M_{n_2} : (m)}$$

$$\underline{S_{n_0} : \langle \oplus S_n \rangle \rightsquigarrow M_{n_0} : (c_n)}$$

$$c_n = \text{decrypt}(\mathbf{Priv}_n, c_n^e) - \text{decrypt}(\mathbf{Priv}_n, c_0^e)$$

$$\underline{S_{n_1} : \langle \ominus S_{n_0} \wedge c_n < \text{threshold} \rangle \rightsquigarrow M_{n_1} : (\text{error})}$$

$$\text{error} = \text{true}$$

$$\underline{S_{n_2} : \langle \ominus S_{n_0} \wedge c_n \geq \text{threshold} \rangle \rightsquigarrow M_{n_2} : (m)}$$

$$\begin{aligned}
\mathbf{Priv}_{n'} &= \text{release}(n')\\
m &= \frac{\text{decrypt}(\mathbf{Priv}_n, x_n^e) - \text{decrypt}(\mathbf{Priv}_{n'}, x_0^e)}{c_n}
\end{aligned}$$

Fig. 5. Final stage S_n of calculating mean value m. Includes sub-stages S_{n_0}, S_{n_1} and S_{n_2}, where S_{n_0} is an initial stage, S_{n_1} makes S_n reach the error milestone M_{n_1}, and S_{n_2} makes S_n reach the successful milestone M_{n_2} with the correct mean value m.

6. Conclusion

Intervention in community health research involves multiple actors in a variety of arenas. It is often directed towards towards the local community and children and their families. The consequence is that research in community health is complex. Collecting data means that the researchers involve many actors. It can be children, parents, teachers, peers, social workers, doctors, and nurses, to give some examples. The analyse process is often based on vertically partitioned datasets, which stresses ethical consideration because a selection of contributors is involved. Our approach has the potential to deal with this complexity in community health research. Data centric workflow modelling is a suitable approach to model complex analysis of data in community health research.

We have demonstrated how SNOOP can be used to perform privacy preserving distributed computation using SMC-algorithms, and we have given an example on how data-centric workflow modelling in D^{2}Worm can avoid computation on to small data sets. In future work we will better integrate these two approaches. The existing workflow management system in D^{2}Worm are not capable to enforce data privacy, and it is problematic that GSM do not provide the syntactical mechanisms to declare organisational boundaries and restricted data exchange across these. This can be achieved with SNOOP integration and more expressiveness in the specification language.

7. Acknowledgement

A special thanks to Martin Jergler at Department for Computer Science, Technische Universität Munchen (TUM) for discussions releated to D^{2}Worm and SNOOP.

References

1. Angiuli, O., Blitzstein, J., Waldo, J.. How to de-identify your data. *Communications of the ACM* 2015;**58**(12):48–55.
2. Starfield, B., Shi, L., Macinko, J.. Contribution of primary care to health systems and health. *Milbank Quarterly* 2005;**83**:457–502. doi:10.1111/j.1468-0009.2005.00409.x.
3. Friedberg, M.W., Hussey, P.S., Schneider, E.C.. Primary care: a critical review of the evidence on quality and costs of helath care. *Health Affairs* 2010;**29**(5):766–72.
4. Conti, G., Heckman, J.J.. The developmental approach to child and adult health. *Pediatrics* 2013;**131**(Supplement 2):133–141. doi:10.1542/peds.2013-0252d.
5. Heckman, J.J.. The economics of inequality: The value of early childhood education. *American Educator* 2011;**35**(1):31.
6. Campbell, F., Conti, G., Heckman, J.J., Moon, S.H., Pinto, R., Pungello, E., et al. Early childhood investments substantially boost adult health. *Science* 2014;**343**(6178):1478–1485. doi:10.1126/science.1248429.
7. Andersen, A., Hardersen, T., Schirmer, N.. Privacy for cloud storage. In: Reimer, H., Pohlmann, N., Schneider, W., editors. *ISSE 2014 Securing Electronic Business Processes; Highlights of the Information Security Solutions Europe 2014 Conference*. Brussels, Belgium: Springer-Verlag. ISBN 978-3-658-06707-6; 2014, .
8. Andersen, A., Karlsen, R.. Privacy preserving personalization in complex ecosystems. In: Linnhoff-Popien, C., Schneider, R., Zaddach, M., editors. *Digital Marketplaces Unleashed*. Springer-Verlag. ISBN 978-3-662-49274-1 / 978-3-662-49275-8; 2017, doi:10.1007/978-3-662-49275-8.
9. Dwork, C.. Differential privacy: a survey of results. In: *TAMC'08, Proceedings of the 5th international conference on Theory and applications of models of computation*. Springer-Verlag; 2008, p. 1–19.
10. Dinur, I., Nissim, K.. Revealing information while preserving privacy. In: *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS'03)*. San Diego, California: ACM. ISBN 1-58113-670-6; 2003, p. 202–210. doi:10.1145/773153.773173.
11. Goldwasser, S.. Multi party computations: past and present. In: *PODC'97, Proceedings of the sixteenth annual ACM symposium on Principles of distributed computing*. New York: ACM. ISBN 0-89791-952-1; 1997, p. 1–6. doi:10.1145/259380.259405.
12. Gentry, C.. Computing arbitrary functions of encrypted data. *Communications of the ACM* 2010;**53**(3):97–105. doi:10.1145/1666420.1666444.
13. Reijers, H.A., Russell, N., van der Geer, S., Krekels, G.A.. Workflow for healthcare: A methodology for realizing flexible medical treatment processes. In: *BPM 2009 Workshops*; vol. 43 of *Lecture Notes in Business Information Processing*. Springer-Verlag; 2010, p. 593–604.
14. Jergler, M., Sadoghi, M., Jacobsen, H.A.. D2WORM: A management infrastructure for distributed data-centric workflows. In: *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (SIGMOD '15)*. ACM. ISBN 978-1-4503-2758-9; 2015, p. 1427–1432. doi:10.1145/2723372.2735362.
15. Hull, R., Damaggio, E., Fournier, F., Gupta, M., Heath, F.T., Hobson, S., et al. Introducing the guard-stage-milestone approach for specifying business entity lifecycles. In: *Proceedings of International Workshop on Web Services and Formal Methods (WS-FM)*; vol. 6551 of *Lecture Notes in Computer Science*. Springer-Verlag; 2010, .
16. Andersen, A., Yigzaw, K.Y., Karlsen, R.. Privacy preserving health data processing. In: *Healthcom'14, 16th International Conference on E-health Networking, Application & Services*. Natal, Brazil: IEEE; 2014, .
17. Andersen, A.. SNOOP: Privacy preserving middleware for secure multi-party computations. In: Costa, F.M., Andersen, A., editors. *Proceedings of the 13th Workshop on Adaptive and Reflective Middleware (ARM 2014)*. Bordeaux, France: ACM. ISBN 978-1-4503-3232-3; 2014, .