Reusable Data in Public Health Databases-Problems encountered in Danish Children's Database

Anna Marie HØSTGAARD^{a,b,1}, Louise PAPE-HAUGAARD^{b,c} ^a Department of Development and Planning, ^b Virtual Centre of Health Informatics (V-CHI), Aalborg University ^c Department of Health Science and Technology, Aalborg University

Abstract. Denmark have unique health informatics databases e.g. "The Children's Database", which since 2009 holds data on all Danish children from birth until 17 years of age. In the current set-up a number of potential sources of errors exist - both technical and human-which means that the data is flawed. This gives rise to erroneous statistics and makes the data unsuitable for research purposes. In order to make the data usable, it is necessary to develop new methods for validating the data generation process at the municipal/regional/national level. In the present ongoing research project, two research areas are combined: Public Health Informatics and Computer Science, and both ethnographic as well as system engineering research methods are used. The project is expected to generate new generic methods and knowledge about electronic data collection and transmission in different social contexts and by different social groups and thus to be of international importance, since this is sparsely documented in the Public Health Informatics perspective. This paper presents the preliminary results, which indicate that health information technology used ought to be subject for redesign, where a thorough insight into the work practices should be point of departure.

Keywords. Public Health Informatics, System Engineering Methods, Sociotechnical Methods, Participatory Approaches, Public Health Databases, Data validation

Introduction

Public Health Informatics (PHI) is a new research area related to the Medical Informatics (MI) area. However, the focus in PHI is on promoting the health of populations as opposed to the health of individuals in MI [1]. PHI is defined as "*the systematic application of information and computer science and technology to public health practice, research, and learning*" [2] (p.239). The objective is to improve the public health status by means of information and communication technologies (ICT) within e.g. health and social services [3]. While research into the development of ICT within the MI area (e.g. technologies leveled at the hospital sector) started already in the 1970'ties [4, 5], research in ICT within the PHI area - e.g. technologies for monitoring of health/disease, home monitoring of chronically ill patients and improved information/communication between health authorities and citizen/patients as well as reuse

¹ Corresponding Author: Vestre Havnepromenade 5, 9000 Aalborg, Denmark, annamarie@plan.aau.dk

of Electronic Health Record data for PHI purposes - has only really started to gain momentum over the last decade [2, 3, 6-9].

Denmark has unique health informatics databases, e.g. the PHI database: "The Children's Database". The objective of this database is to monitor the health of all Danish children in order to detect infants at risk as early as possibly. It was established in 2009 by the Danish National Board of Health (NBH) and represents a national PHI database holding health related information on a *complete* population, which is unique, both internationally and nationally. However, spot checks performed by the NBH show that the data are flawed. Therefore, they are not used for either primary (e.g. monitoring) or secondary purposes (e.g. research).

The objective of the present on-going study is to develop *generic* methods to optimize data transmission and validate data in PHI databases in order to enhance patient treatment and public information, as this area is sparsely documented in the Public Health perspective. The study takes its point of departure in "The Children's Database", because this case enables a unique opportunity to study, analyze and evaluate methods and procedures best suited for exploiting the potentials for both primary and secondary use of the data. The case is characterized by data collection and data entry performed by various user groups in various local, social contexts and into different local IT-systems. This means that the knowledge and the methods developed can be applied in similar ICT cases/solutions within health and social services, e.g. for improving communication between the social/health sector and citizens/patients in the elderly, chronic disease and rehabilitation area. The objective of this paper is to present the preliminary results of the study: 1) The identified architecture used for accessing "The Children's Database" and 2) Evaluation of the impact of the social context where data are collected and entered.

1. Methods

The study is designed in three steps (Figure 1).



Figure 1. Methodologic approach. Ovals are symbols for steps, rhomboids are symbols for applied methods

1.1. Methods within the three steps and status on the study

1) Document analyses were conducted in order to identify the architectural design in The Children's Database: Reports and analysis from the NBH and the Danish data network (MedCom). These were supported by informal interviews with municipal and national stakeholders: The project leaders and IT-professionals within the three municipalities, a representative for the Danish data network (MedCom) and a representative for the NBH.

2) A preliminary statistical analysis was conducted in order to identify error types in stored data. The analysis was based on the data to be used in the upcoming *registry study* of data in the central, national database (The Children's Database) and the three decentralized, local IT-databases in Aalborg, Gentofte and Hedensted municipalities, respectively. Data about CPR (central personal registration number), height and weight for all children from the three municipalities, aged 0-17 years, during the period of 1. April 2009 to 1. June 2011, were studied (21.000 children). Permission to access data on the personal level was obtained from the Data Protection Agency, The National Board of Health and the three municipalities.

3) In a preliminary ethnographic study, the work practices associated with generating data in different social contexts (3 days at home visits and 2 days at school clinic examinations - primary care practitioners have not yet been observed) b) by different user groups (different municipal staff nurses handle home visits and school clinic examinations respectively) into one of the three local municipal IT-systems (different systems in the three municipalities) were studied using observation and interview techniques. Focus was (and will be in the ongoing ethnographic study) on the interactions between the user and the technology, in particular on the socio-technical problems associated with the development of PHI technologies. The data generating process was studied from the time the relevant information's were observed and documented in the decentralized, municipal, local IT-systems until they were finally transferred to the national, central PHI database (The Children's Database) in order to a) identify the types of errors and b) where in the data generating process the errors occur.

2. Results and Findings

2.1. Identified architecture and potential sources of errors

"The Children's Database" holds specific information on *all* Danish children and adolescents aged 0-17 years. For the current pilot-version of the database, only data about height, weight, passive smoking and breastfeeding are collected. Data from the 98 Danish municipalities are supposed to be entered into the national database in three different ways: 1) Up to one year of age, data collected from home visits by municipal staff nurses are entered into decentralized, municipal, local IT-systems (4 different systems in Denmark). 2) Between one and approx. five years, data are entered into primary care electronic health record systems during visits to the primary care physicians (11 different systems in Denmark). 3) In primary school, data are entered into decentralized, municipal, local IT-systems on visits to the municipal staff nurses/physicians at the schools (4 different systems in Denmark). Data from the decentralized, municipal, local IT-systems and the primary care electronic health record systems in each municipality are transferred to the national, central database (The Children's Database) once a day. The identified architecture used for accessing the national central PHI database is illustrated in Figure 2.

A statistical analysis of 21.000 children's data stored in the national central database (The Children's Database) show several potential sources of data errors. The types of errors can be divided into 'true doublets' and 'false doublets'. The 'true doublets' are identified as identical data on a child - e.g. height and weight - registered on the same date multiple times, also known as redundant data. The 'false doublets' are identified as different data on a child registered on the same date. These doublets are the most critical, because they influence the data quality in secondary use.



Figure 2. The architecture of the national central database: The Children's Database. It holds data on height, weight, passive smoking and breastfeeding as well as date time. The pointer for the stored data is the Danish CPR (central personal registration number) – a unique ID.

These doublets might be caused by a number of different mistakes occurring in different stages in the life cycle of data: when collected, when entered, when processed, when transmitted, when stored. In addition, the doublets might be caused by poor da-tabase-design. However, because of the identified architecture, the natural place to start exploring the origins to the identified data-errors is where data is first generated and entered. This is elaborated below.

2.2. The ethnographic study of potential sources of errors

Observations and interviews of five municipal staff nurses performing their daily work tasks in homes and at school clinics in one of the municipalities show that the present work practices between the technology, the nurse, the child and the family can easily cause misreporting, and hence data errors. At home visits, the nurses are supposed to enter data on their laptops immediately after having performed each of the different examinations (e.g. height and weight). However, as the children examined are between 0-1 years of age, most of the examinations take place while playing on the floor with the child. All nurses observed so far, find it inappropriate to disrupt the contact with the child and the parents while entering data into the PHI system. Therefore they have developed their own way of workarounds to make data-entry fit with their work practice: Most of them note down the data on a piece of paper and enter them into the system, when they have finished the examinations –while they are still in the home. Others note down the data for *all* children examined during a day and enter them into the system, when they are back in the office in the afternoon or the next morning. At school clinics, most nurses also find it disruptive to perform data-entry while examina-

ing and talking with the children and their parents. Therefore, they also note down the data on a piece of paper and enter data later on, when the children have left the clinic.

The nurses have the possibility of assessing their latest data entries by looking at curves of height and weight in relation to age generated by the system. However, they often fail to do this because of lack of time, or because they forget about it. The system does not provide any automatic warning in case of data entries inconsistent with former data on e.g. height or weight. Interviews with nurses and an IT-professional reveal that the local IT-system was developed by IT-professionals with users as consultants, rather than real participants.

3. Discussion

Our preliminary results and findings support previous studies showing that implementation of new health information technology causes a mutual impact on both the technology and the organization/the individuals, who are going to use it, because of mutual interdependency and interrelations [10,11]. Our observations and interviews show that the PHI system examined in the preliminary study is designed to be fed with data in one continuously workflow, where the nurses perform data entry, while they examine the children. However, due to the importance of physical contact and no disruptions between nurses and the observed children, this is not possibly, and therefore the nurses develop workarounds. Unfortunately, some of these induce data entry errors. In other words, our results indicate that the technology used ought to be subject for redesign, where a thorough insight into the work practices should be point of departure.

References

- O'Carrol P. Introduction to Public Health Informatics. In: O'Carroll PW, Yasnoff WA, Ward ME, Ripp L, Martin E, ed. Public Health Informatics & Information Systems. 2nd ed. New York: Springer; 2003.
- [2] Friede A, Blum HL, McDonald M. Public Health Informatics: How Information-Age Technology Can Strengthen Public Health. Annual Reviews Public Health 1995; 16:239-252.
- [3] Yasnoff WA, O'Carroll PW, Koo D, Linkins RW, Kilbourne EM. Public Health Informatics: Improving & Transforming Public Health in the Information Age. Journal of Public Health Management & Practice 2000; 6(6): 67-75.
- [4] Greenes RA, Shortliffe EH. Medical Informatics: An emerging academic discipline and institutional priority. The Journal of American Medical Informatics 1990; 263:1114-1120.
- [5] Høstgaard AM. Change is the spice of life? Process analysis case study of a region in Denmark. Ph.D. thesis. Aalborg: Aalborg Universitet; 2009.
- [6] Röcker C, Maeder A. User Centered Design of Smart Healthcare Applications. Electronic Journal of Health Informatics 2011; 6(2): e11.
- [7] Dinesen B, Seemann J, Gustafsson J. Development of a program for tele-rehabilitation of COPDpatients across sectors: coinnovation in a network. International Journ. of Integrated Care 2011; 11:e012. Epub 2011.
- [8] Safran C, Bloomrosen M. Hammond E, Labkoff S, Markel-Fox S, Tang P, etmer D. Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper. Journal of the American Medical Informatics Association 2007; 14(1): 1-9.
- [9] Kukafka R, Ancker J, Chan C, Chelico J, Khan S, Mortoti S, Natarajan K, Presley K, Stephens K. Redesigning electronic health record systems to support public health. Journal of Biomedical Informatics 2007; 40 (4): 398-409.
- [10] Wears RL, Berg M. Computer Technology and Clinical Work. Still Waiting for Godot. Journal of American Medical Association 2005; 293(10): 1261-1263.
- [11] Berg M, Aarts J, Van der Lei. ICT in health care: Sociotechnical approaches. Methods in Information in Medicine 2003; 42(4): 297-301.