Feasibility Study on Service-based Data Acquisition for Human Signal Molecule Profiling Database

Xinyan Zhao and Tao Dong,* *Member, IEEE*

Abstract— **A service-based model of human signal-molecule-profiling database (HSMPD) was proposed to prompt data acquisition of medical information from blood tests in traditional healthcare systems and to make the data be naturally converted into Health IT (HIT) services through the translational bioinformatics (TBI) community. A self-motivated mechanism to minimize the project investment was designed. The low-cost tool, called 'SMP chip', will be employed in blood tests inside hospitals, which results could be decoded and stored in HSMPD automatically. Successively, HIT services derived from HSMPD will provide a stable income to support the data acquisition processes. To explore the feasibility of the model, a small-scale survey study was performed in a common hospital. The positive results indicated that the cooperation of IT, engineering, healthcare systems and TBI community could provide a self-motivated solution to build the HSMPD platform.**

I. Introduction

Biological information which was stored in databases is of great importance to clarify the complex relations between proteins, genes, diseases and environmental factors . Studies which are on biobanks and medical databases supply chances for a wide range of applications and interests in translational bioinformatics (TBI). There are many projects in different countries that invest in health IT research by constructing health information databases and biobanks [1]. Researchers on medical science, biology, health care systems and bioinformatics will be beneficial from those types of biobanks and biomedical databases.

Health information technology (HIT), as one of the backbones in future medical systems, is based on qualified biomedical databases and large capacity [2]. Nowadays, the results of physical examination could only be analyzed by individual doctors; however, references of auxiliary diagnoses will tend to be provided by HIT with its additional services by comparing the diagnoses with previous clinical cases in the database. The service will substantially boost the reliability and efficiency of the medical diagnosis without increasing costs. Telehealth and personalized medicine may be achieved with the development of service-based HIT system. However, the HIT system has to depend on the construction of large biomedical databases. [3, 4]

Varied biomedical databases have already been constructed globally, such as DrugBank, Human Genome Database (HGD), Recon 2, ClinicalTrials, and Human Metabolome Database. [5-9] However, the existing databases mainly concentrate on laboratory studies rather than clinical studies. The maintenance and completion of these databases mainly relay on research finances from the government, which imply that there are still not any good business models for supporting these HIT systems. Therefore, present biomedical databases are still away from clinical auxiliary diagnoses. Few clinical databases have been established by the support of public departments, for example, ClinicalTrials currently has 144,133 studies with exact locations in 185 countries and 50 states. However, a large part of those clinical trial studies got non-ideal results. High quality clinical databases are still highly needed in TBI.

On the other hand, mass clinical data are generated in the worldwide healthcare system every day. These valuable data are collected from real clinical researches, instead of simulation tests on models of animals. TBI studies will be greatly improved when clinical data are translated to electronic health records (EHRs) which are located in a public database.[10] Researchers did many pilot studies and had made innumerable prototypes in clinical knowledge management. However, most of these studies have not been used in clinical practice, probably because of the high expenses of data acquisition as well as the traditional habits of doctors and nurses, to whom HIT innovations are not urgent requirements; HIT seems not to be essential for patients as well. Only bioinformatics researchers are the beneficiaries of EHRs at this stage. Two main strategies were proposed on the basis of the requirements.

The first strategy is a compromise between the current healthcare systems and HIT, i.e., traditional medical records will be changed into standard EHRs without the modification of the existing working habits and system. The major challenge is the cost and efficiency of EHRs' conversion. The second strategy aims to redesign medical tools to gather standard EHRs automatically during the ordinary procedure of traditional medical. The second

*Manuscript received on Aug. 3th, 2013. This article is supported by Norsk regional kvalifiseringsstøtte fra Oslofjorfondet (Et cellebasert digitalt mikrofluidisk system, Prosjekt nr: 220635) and Norwegian long term support from NorFab (Living-Cell-based LOC project). The Research Council of Norway is acknowledged for the support to the Norwegian Micro- and Nano-Fabrication Facility, NorFab (197411/V30). The travel grant for the international conference is partially offered by Norwegian Research Council through the Norwegian PhD Network on Nanotechnology for Microsystems (Nano-Network), contract no. 190086/S10.

T. Dong (IEEE member: 90889664) is with the department of Micro and Nano Systems Technology (IMST, TekMar), Vestfold University College, Norway. (PhD T. Dong is the corresponding author, E-mail: Tao.Dong@hive.no; Phone: +47 3303 7731, Fax: +47 3303 1103).

X. Zhao (IEEE member: 92708937) is also with the department of Micro and Nano Systems Technology (IMST, TekMar), Vestfold University College, Norway. (E-mail: Xinyan.Zhao@hive.no).

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

strategy only needs the change on machines, avoiding the conversion of medical records at the source. The second strategy was selected in the construction of human signal molecule profiling database (HSMPD) here.

The HSMPD is designed for collecting serological concentrations of hormones and cytokines i.e. signal molecule profiling (SMP). 'SMP chip', a standard cheap microfluidic microarray had been developed for the processing of data acquisition. As for data acquisition often acts as the bottleneck of HIT projects, requiring gigantic funds of scientific projects, a service-based model for the data acquisition of HSMPD is proposed. The user survey was also studied to evaluate the feasibility of data acquisition model.

II. Materials and Methods

A. The service-based model for HSMPD

HSMPD would be promising for TBI researchers to promote public HIT services. The key of a possible solution to install the HSMPD without involving an enormous scientific project lies in the cooperation of the traditional healthcare system, IT, instrumentation engineering and the TBI community, shown in Fig 1.

The first step is to develop a HSMPD-oriented tool that can standardize the high-throughput quantification hundreds of cytokines and hormones. To reduce the difficulty of data collection, SMP chips will be coupled automatically with the online HSMPD system. Moreover, this tool is made as low-cost as possible to be widely accepted by medical systems.

When patients go to the health centres or hospitals, SMP chips could perform routine blood tests for cytokines or hormones because of the competitive price of the chips and high efficiency. Before relevant analysis reports returns, SMP information will be decoded through a web-based program; the SMP data will be stored in the web-based HSMPD automatically. Obtaining commercial information of HSMPD at an acceptable cost, TBI researchers can generate a stable income to support data acquisition.

The traditional medical system can partially cover the cost of data acquisition. Normally, patients have to pay for their physical examinations in hospitals, thereby covering partial costs of SMP chips. Even though they are much cheaper than the normal commercial microarrays, the SMP chips are more expensive than traditional diagnostic kits available now. Under the present circumstances, there will be a deficit in the SMP data acquisition process. In the service-based model, the users in the TBI community will remedy the deficit. Personal users or patients of HSMPD could contribute the anonymous SMP data to the database; doctors will also contribute medical comments related with the SMP data. Thus, the valuable database will be provided as a paid service for universities in the TBI community and/or IT companies.

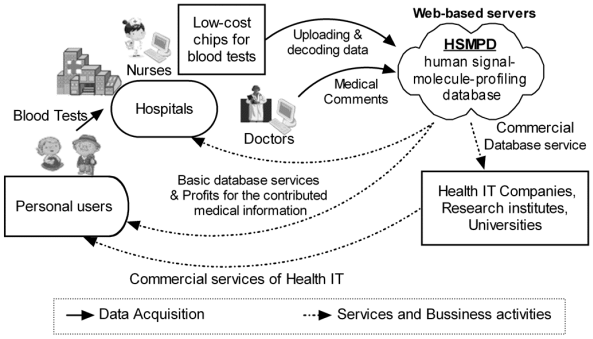

Fig. 1. The service-based model of HSMPD. The HSMPD system will be a bridge to connect traditional healthcare systems with the TBI community. The HIT services on TBI community can provide a stable income to support the data acquisition process of HSMPD, which forms a self-motivated mechanism.

The incomes of commercial data services could cover the deficit in the data acquisition process. Furthermore, the information donors in HSMPD, e.g. patients and doctors, might gain some rewords and compensates for their contribution. It is believed that some valuable diagnosis criterion of SMP would be discovered and developed into multipurpose HIT services for morning a variety of human diseases; HIT companies will perform enough studies of data mining in the HSMPD database. When such a commercial platform could be formed, both patients and IT companies will get the benefits from the HIT services. Thereby, the construction of HSMPD could be accompanied by the spontaneity in economics.

B. Survey study on data acquisition

This clinical investigation and survey study on the data acquisition process was assisted by the medial partner. A total of 200 pieces of SMP chips provided as blood-testing tools to measure 16 types of signal molecule targets, including TSH, TNF-α, cortisol, FSH, LH, HCG, EPO, GnRH, IFNγ, IGF-1, EGF, VEGF, interleukin (IL)-1 beta, IL-2, IL-6 and IL-8. The SMP chips were made before the investigation. As described previously [11, 12], the SMP chips were measured inside a microplate reader and operated manually in the laboratory of the medial partner.

The data acquisition process was performed in a passive way, shown in Fig 2. The respondents will go to the blood-taking room for blood sampling when doctors advise them to undergo blood tests. The researchers were permitted to wait in the room and give the suggestions of respondents to accept free additional SMP tests accompanied by their original blood tests. The SMP tests only need a small volume of blood sample; the remained sample left from an ordinary blood test is usually enough for a SMP test. As long as the respondents accepted free SMP tests, their remained blood samples would be measured by the SMP chips, thus no extra sampling of blood was performed in this study. We orally interviewed about 500 respondents regarding their attitude towards the HSMPD model. The investigators for the survey study were divided into two groups.

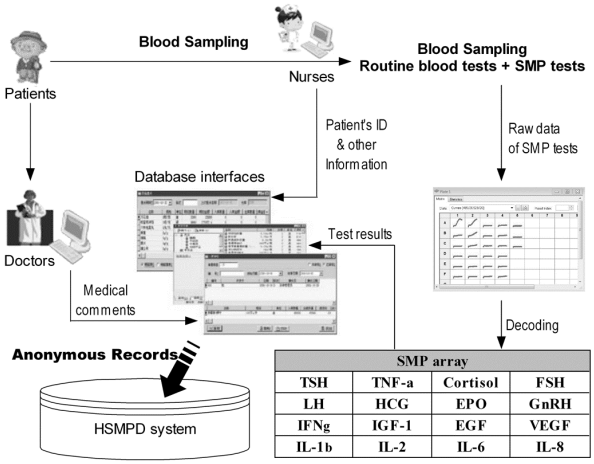

Fig. 2. Typical process of data acquisition in the HSMPD model. A flow chart above is to demonstrate the data acquisition process for HSMPD. The concentrations of signal molecule targets were measure on the SMP chip and stored into HSMPD system, along with the medical comments from doctors.

One group interviewed the people who would get their blood tested in the blood-taking room inside the hospital (Ziyang City, China); the other group randomly interviewed the pedestrians in city streets far away from hospitals. The respondents for this investigation were randomly selected and unfiltered; the time period of each group was evenly distributed in two weeks and consistent. We made further inquiries about the ethical issues about the data acquisition, as well as the physical conditions of interviewees in the blood-taking room. The interviewees were fully informed before the survey studies were performed and no private information of the respondents was divulged outward.

III. Results and Discussion

The results of survey about respondents were presented in Fig. 3. The survey conducted far away from hospitals (Fig. 3, Inset A) presented that over 90% of the general sample population refused to or did not want to participate in the HSMPD system, even though the SMP tests were free. However, the result gained in the blood-taking room was of significantly difference. Fig. 3, inset B showed that over 50% of the respondents were willing to accept the free SMP tests without additional blood sampling. Besides, if the additional SMP test was not free, no one was willing to participate.

The above information indicated that only participants who had planned to undergo blood tests could possibly accept free SMP tests. As for the arrangement of the survey study may not be strict enough, the deduction still needs to be verified by follow-up scientific studies. Nevertheless, the conclusion of this investigation is also logical and is likely to reflect group psychology. Preliminary results of the survey showed that it is impossible to ask the respondents to pay additional costs in the data acquisition for HSMPD. Actually, it is difficult to persuade a common respondent to accept a free SMP test.

Further inquiries about the background of interviewees were made in the blood-taking room. The topic of conversation referred to the possible challenges in the HSMPD model, for example, the privacy issues. Fig. 3, inset C described the reasons why 143 respondents refused to participate in the project. More than half of the respondents were unwilling to be interviewed and were distrustful of investigators. The reasons for the rejection might be divided into two parts: firstly, the respondents considered the SMP tests useless; secondly, this HSMPD project was distrusted.

The investigators generated more information by talking with the 189 amicable respondents who were willing to take SMP tests. Fig. 3, inset D showed the purposes and reasons of taking the blood tests. Among the interviewees, 86.77% needed medical examination under the doctor's advices and had uncomfortable symptoms. Moreover, 58.2% of the respondents did not know the cause of their diseases indicating that they subconsciously hoped to obtain additional information from the SMP tests. The other 13.23% claimed to experience routine physical examinations and took the SMP tests in passing. No person was found who claimed to take the SMP tests only aiming at scientific research. The results indicated that the respondents were unwilling to participate in scientific projects which are not associated with their health. SMP test might only be accepted when a patient spontaneously wants to take a blood test. Fig. 3, inset E showed the information of the points of view regarding the privacy issues in the SMP tests.

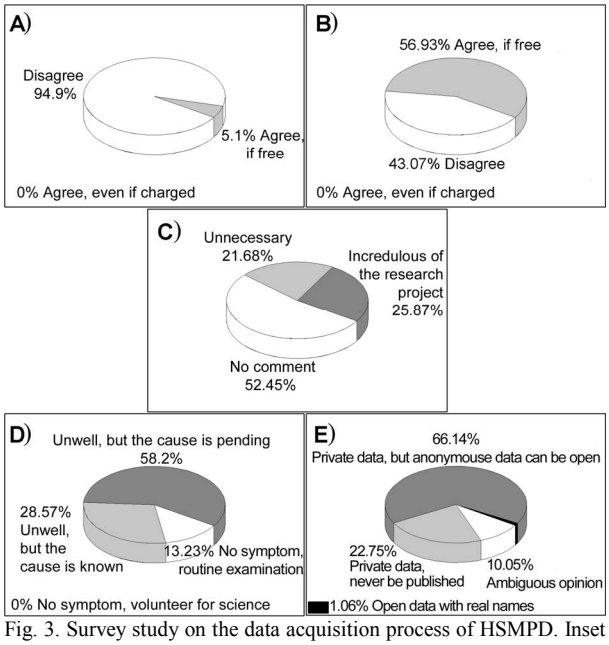

Fig. 3. Survey study on the data acquisition process of HSMPD. Inset A showed the survey result on the attitudes towards HSMPD among the people far away from hospitals. Inset B demonstrated the different survey result on the attitude of the public in the blood-taking room. Inset C showed the distribution of reasons given by 143 interviewees who refused to participate in the HSMPD project. Inset D revealed the reasons and purposes of the 189 participants in the HSMPD project. Inset E showed the points of view of the participants regarding the privacy issues of the HSMPD.

The respondents argued private information should not be involved in the database, except for physiological for diagnosis. According to the foundations of the above comments, users' information should be divided into two parts. The first part can be accessed with the mediation of a trusted intermediary, whereas the public part is allowed to contain medical comments and SMP values from the doctors. A string of identification numbers will replace the name of patients, thereby making it possible for TBI researchers retaining clues for in-depth studies and distinguishing individual records from different persons. With the help of trusted intermediary, TBI researchers could indirectly contact with doctors or important patients. If doctors or the patients are willing to participate in the corresponding study, the researchers can achieve their detailed clinical data which may be acceptable for both sides. Moreover, 22.75% of the respondents who refused to share their data should also be respected. Allowing users to completely conceal their data in HSMPD, they have to be fully responsible for their costs of blood tests because no data contribution leads to no compensate from HIT services. TBI researchers will be allowed to circularly use anonymous medical records by the HSMPD system; however, detailed personal information of patients might be required in advanced HIT services.

Although HSMPD has the protection design for the participants' privacy, methodological medical methods approach to be avoidless. Thus, social controversy and tremendous ethical may be generated between participants and scientists of the HSMPD. The logical result is to seek a trusted intermediary in the HSMPD model which represents the participants, e.g., the government. The trusted intermediary can invite participants who also have the choices to refuse such invitations to attend some specific studies.

The passive method of data acquisition can only fulfil that need when the selection of test objects that are often unable to meet the needs of experimental studies is not allowed by the passive mode of data collection in the service-based HSMPD system because experimental objects need to be chosen by the experimental research. The medical experiments often includes making assumptions, designing experiments between the experimental group and the control group, and conducting clinical trials to verify the hypotheses. However, a feasible solution to construct the HSMPD system without astronomical amounts of investment is provided by the service-based model of HSMPD. Moreover, the passive method of data collection may be helpful for medical experimental research in the future. Being able to result in more reliable experiments, the distribution of any single disease will be large enough to verify the hypotheses through statistical analyses when sufficient data of random users are accumulated in the HSMPD.

IV. Conclusion

From the preliminary survey study of the HSMPD model, we confirmed that the method of data acquisition for HSMPD could be accepted as an additional medical service by the participants in traditional healthcare systems. The cooperation of IT, engineering, healthcare systems and TBI community will provide a feasible solution for constructing the HSMPD system in a self-motivated model. Besides, this study also indicated that the service-based methods of data acquisition might be a universal and effective solution to accelerate the development of other medical information databases, which could prompt the outbreak of TBI.

Acknowledgment

The research activities were supported by Ziyang Maternal &Child Health Hospital (Yiyang, China), Nanjing University of Science and Technology (Nanjing, China), Hubei University (Wuhan, China) and Xiamen University (Xiamen, China). In particular, the authors would like to acknowledge our colleagues Z. Yang, F. Karlsen, J. Soriano, Doctor H. Huang and Doctor X. Liu et al. for providing medical support, technical contributions and helpful suggestions.

References

[1] D. F. Sittig, A. Wright, L. Simonaitis, J. D. Carpenter, G. O. Allen, et al., "The state of the art in clinical knowledge management: an inventory of tools and techniques," Int J Med Inform, vol. 79, pp. 44-57, Jan 2010.

[2] S. Colantonio, M. Esposito, M. Martinelli, G. De Pietro, and O. Salvetti, "A Knowledge Editing Service for Multisource Data Management in Remote Health Monitoring," IEEE Trans Inf Technol Biomed, Aug 27 2012.

[3] S. Lista, F. Faltraco, and H. Hampel, "Biological and methodical challenges of blood-based proteomics in the field of neurological research," Prog Neurobiol, vol. 101-102, pp. 18-34, Feb-Mar 2013.

[4] A. G. Ekeland, A. Bowes, and S. Flottorp, "Methodologies for assessing telemedicine: a systematic review of reviews," Int J Med Inform, vol. 81, pp. 1-11, Jan 2012.

[5] I. Thiele, N. Swainston, R. M. Fleming, A. Hoppe, S. Sahoo, et al., "A community-driven global reconstruction of human metabolism," Nat Biotechnol, Mar 3 2013.

[6] M. D. Sorani, W. A. Ortmann, E. P. Bierwagen, and T. W. Behrens, "Clinical and biological data integration for biomarker discovery," Drug Discov Today, vol. 15, pp. 741-8, Sep 2010.

[7] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, et al., "DrugBank 3.0: a comprehensive resource for 'omics' research on drugs," Nucleic Acids Res, vol. 39, pp. D1035-41, Jan 2011.

[8] D. S. Wishart, T. Jewison, A. C. Guo, M. Wilson, C. Knox, et al., "HMDB 3.0--The Human Metabolome Database in 2013," Nucleic Acids Res, vol. 41, pp. D801-7, Jan 2013.

[9] W. Bodmer, "Human Genome Project," in Brenner's Encyclopedia of Genetics (Second Edition), M. Editors-in-Chief: Stanley and H. Kelly, Eds., ed San Diego: Academic Press, 2013, pp. 552-554.

[10] A. Geissbuhler, C. Safran, I. Buchan, R. Bellazzi, S. Labkoff, et al., "Trustworthy reuse of health data: a transnational perspective," Int J Med Inform, vol. 82, pp. 1-9, Jan 2013.

[11] X. Zhao, T. Dong, Z. Yang, N. Pires, and N. Hoivik, "Compatible immuno-NASBA LOC device for quantitative detection of waterborne pathogens: design and validation," Lab on a Chip, vol. 12, pp. 602-12, Feb 7 2012.

[12] X. Zhao and T. Dong, "Multifunctional sample preparation kit and on-chip quantitative nucleic Acid sequence-based amplification tests for microbial detection," Anal Chem, vol. 84, pp. 8541-8, Oct 16 2012.