Leading article

Accelerating the development of an information ecosystem in health care, by stimulating the growth of safe intermediate processing of health information (IPHI)

Harshana Liyanage BSc MBCS
Research Fellow, Department of Computing, University of Surrey, Guildford, UK

Siaw-Teng Liaw PhD FRACGP FACHI FACMI
Professor of General Practice, University of New South Wales, Randwick, NSW, Australia

Simon de Lusignan BSc MBBS MSc MD(Res) FHEA FBCS CITP FRCGP
Professor of Primary Care & Clinical Informatics, Editor *Informatics in Primary Care*, Department of Clinical Informatics, University of Surrey, Guildford, UK

ABSTRACT

Health care, in common with many other industries, is generating large amounts of routine data, data that are challenging to process, analyse or curate, so-called 'big data'. A challenge for health informatics is to make sense of these data. Part of the answer will come from the development of ontologies that support the use of heterogeneous data sources and the development of intermediate processors of health information (IPHI). IPHI will sit between the generators of health data and information, often the providers of health care, and the managers, commissioners, policy makers, researchers, and the pharmaceutical and other healthcare industries. They will create a health ecosystem by processing data in a way that stimulates improved data quality and potentially healthcare delivery by providers of health care, and by providing greater insights to legitimate users of data. Exemplars are provided of how a health ecosystem might be encouraged and developed to promote patient safety and more efficient health care. These are in the areas of how to integrate data around the unsafe use of alcohol and to explore vaccine safety. A challenge for IPHI is how to ensure that their processing of data is valid, safe and maintains privacy. Development of the healthcare ecosystem and IPHI should be actively encouraged internationally. Governments, regulators and providers of health care should facilitate access to health data and the use of national and international comparisons to monitor standards. However, most importantly, they should pilot new methods of improving quality and safety through the intermediate processing of health data.

Keywords: data aggregation, distributed systems, ecosystems

Introduction

Health care, in common with many other industries, is generating large amounts of routine enterprise data, which can be mined and even combined with comments, tweets and blogs. This mass of data is termed 'big data'. A challenge for informatics is to make sense of these data, which can sit in numerous disparate systems; and due to their sheer volume are hard to analyse, process and curate en masse.[1]

Making sense of these data offers opportunities for the surveillance of disease[2] and addressing complex

public health issues, as well as for running complex healthcare providers.[3] Ontologies that support the use of heterogeneous data sources will provide part of the answer;[4,5] ontological concepts can be used as a means of improving the quality of complex research data.[6] In addition, there is a place for advanced data mining and data modelling methods, including 'Mashups' of data,[7] enhanced use of metadata and semantic enhancement (Box 1).[8] However, the best way to stimulate the development of these tools is through the development of safe intermediate processors of health information (IPHI) working within a health ecosystem (Figure 1).

IPHI will sit between the generators of health data and information, often the providers of health care, and the users of this information. The users of information are health service managers, commissioners, policy makers, researchers and the pharmaceutical and other healthcare industries. They will create a health ecosystem, by processing data in a way that stimulates improved data quality and potentially healthcare delivery by providers of health care and by providing legitimate users of data with greater insights. Figures 2 and 3 contrast traditional flows of data within health systems, which are often slow and unfriendly to interpret, with how a new IPHI provider might work. The IPHI may also have data feeds from patients' comments and social media; their outputs might also go to the press and public.

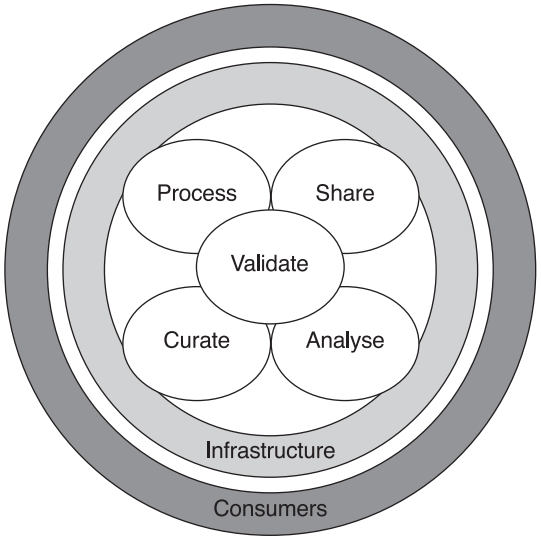

Figure 1 Information ecosystem

Meaningful Mashups

The last decade introduced a new wave of information sources in the form of World Wide Web (WWW)-based Mashups (a term invented for combining music tracks that has been taken over as a term for combining multiple data sources). Mashups are Web applications that combine multiple sources of information to generate a secondary source which gives a new perspective of the data involved.[9] Mashups were the highlight of the Web 2.0 era of the WWW.[10] They have evolved from simple data mixing to platforms that synthesise complex information structures. These 'meaningful' Mashups have technological similarities with the IPHIs discussed here. They have been piloted for looking at genes, and clinical and geographical data,[11] and also been developed using ontologically rich processes.

Box 1 What is an information ecosystem?

An information ecosystem is a complex environment in which data and information providers, users and processors interact in a mutually interdependent and transformational process.
The information ecosystem is made more dynamic by the diversity and effectiveness of the information processors:
data mining, combining multiple information sources to produce new information (Meaningful Mashups);
multiple methods of data presentation and visualisation, some will be brief and superficial;
interactive processes involving linked data;
'sherpas', methods to guide users to the information they seek; and
interactive databases
Within the health environment these tools are what we refer to as the *intermediate processors of health information* (IPHI), unlike other sectors they need to ensure privacy—assuring professionals, patients and the public that they are adequate provision for information governance and security in place.

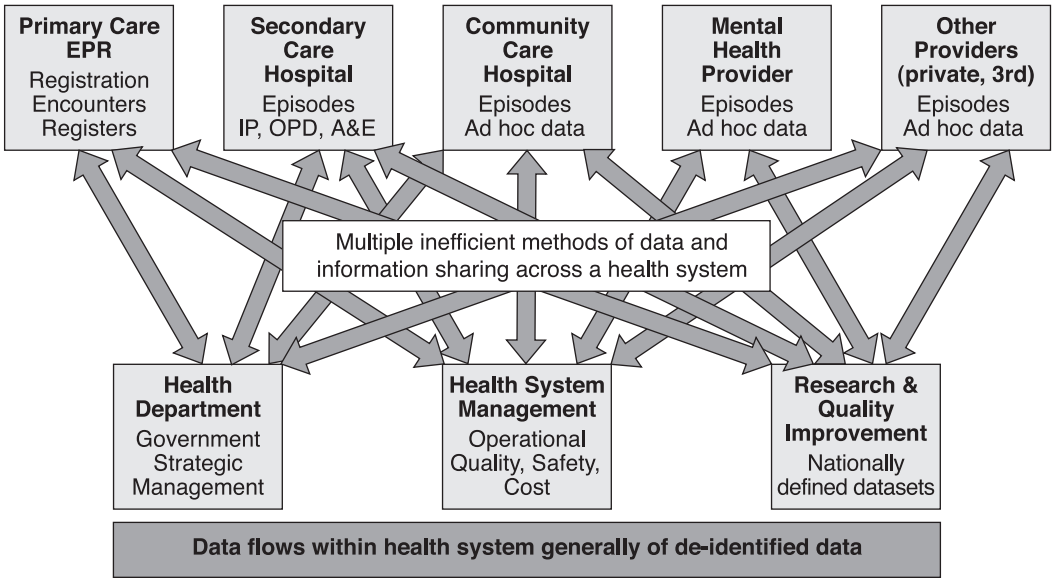

Figure 2 Traditional flows of data in a health system

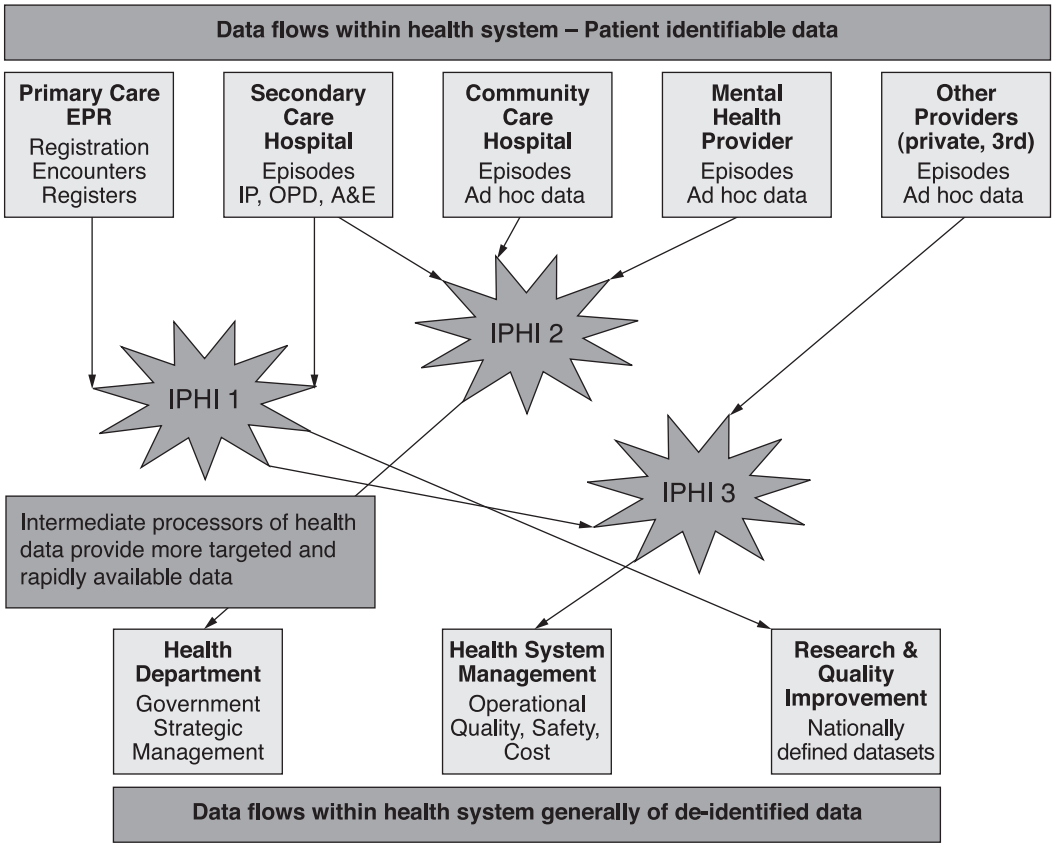

Figure 3 Role of intermediate processors of health information (IPHI) in a health system

Exemplar roles of an IPHI: unsafe use of alcohol and vaccine coverage, benefits and risk

Exemplars are provided of how a health ecosystem might be encouraged and developed to promote patient safety and more efficient health care. These are in the areas of how to integrate data around the unsafe use of alcohol and to explore vaccine coverage, benefits and safety.

Vaccine effectiveness monitoring

The 2009 pandemic influenza (H1N1) experience demonstrated the presence of a major gap in vaccine monitoring and benefit–risk assessment across Europe. The key issue appears to be not the absence of data for monitoring such public health emergencies, but the coordinated aggregation of such data within a realistic time frame. In theory, open standards are an idealistic approach for the harmonious collation of data across the health enterprise. However, the perception about incorporating open data standards has changed in the recent years due to numerous information systems projects failing to effectively adopt such standards in usable implementations.

Vaccine monitoring and benefit–risk assessment would be successful only if the data generated from thousands of data sources can be orchestrated to form a 'unified information flow' within the healthcare ecosystem. Existing efforts to monitor vaccine outcomes utilise only a fraction of the 'big data' available from the healthcare enterprise. An information architecture that could leverage wider coverage of the available health data would have to be flexible. We would also have better results in these initiatives if our approach shifted from being 'data collection centric' to 'data processing centric'. Recent advances in distributed computing have made the latter approach more feasible than ever.

The challenge for generating a 'unified information flow' for vaccine monitoring would also require overcoming information governance requirements, which are usually specific to the locality of the data sources. Therefore, an improved information flow would also need to be complemented by standardised governance procedures that would allow flexibility while not affecting data integrity. One of the key success factors in big data implementations so far have been the abundant availability of open data. It will be interesting to see if 'big data' would be equally successful within data ecosystems with restrictive governance policies.

Monitoring the unsafe use of alcohol

Alcohol use among under-aged youth is a growing problem and a burden on emergency departments.[12] Despite having certain legislative frameworks in place, there seems to be a clear increase in alcohol-related violence and injuries.[13] Easy access to alcohol is mainly due to the ability to obtain fake licences online.[14] Addressing such issues is complicated because the data involved are available at multiple granularities. Nevertheless, sharing data on alcohol-related injuries with local partners to monitor local trends and take preventive action, such as targeted policing and licensing enforcement, has been successful in the past.[15] We need to examine novel methods of generating the information required specifically for this purpose. This data may originate from crime records, ambulance logs,[16] emergency admissions and other related sources. An IPHI produced for this purpose may serve as an information source for off-licence renewals or alcohol control for the beverage industry.

Alcohol consumption is a major risk factor for the global burden of disease. Monitoring alcohol consumption patterns on a global scale provides critical insight for enforcing control through healthcare policies. The World Health Organization (WHO) has developed a comprehensive information system that includes data on more than 200 alcohol-related indicators, named the Global Information System on Alcohol and Health (GISAH).[17] This system uses the Statistical Data and Metadata Exchange–Health Domain (SDMX-HD) data-exchange format to facilitate the exchange of indicator definitions and data from their data sources.[18] GISAH and its hierarchy of data sources across various national health services form a specialised ecosystem based on WHO proprietary standards. In addition, this demonstrates that implementing IPHIs in a controlled environment can be achieved on a large scale.

Safe and private processing of health and other linked data

A challenge for IPHI is how to ensure that their processing of data is valid, safe and maintains privacy. Information governance (IG) plays a key role in ensuring controlled access at the source of data. The dynamics of IG can potentially be more complicated in a setting in which multiple IPHIs cascade information across the healthcare ecosystem. Considering the nature of the data involved, it may be more effective to achieve privacy at the data level. Recent advances in cryptographic techniques such as homo-

morphic encryption allow data to be processed while being encrypted.[19]

Building isolated information ecosystems that guarantee the highest levels of security can be expensive. Also, present-day security attacks are highly sophisticated and isolation is a weaker form of security compared with other types. Therefore, the ideal case would be to use security mechanisms at every level of the technology stack and working with data in which the identifiers, if not all the data elements, are 'hashed' to reduce the risks of identification by the people who work with the datasets.

Discussion

Development of the healthcare ecosystem and its associated IPHI should be actively encouraged internationally. Such developments could help tackle complex health issues such as how to reduce the risks associated with alcohol and how to monitor vaccine effectiveness. Governments, regulators and healthcare providers should facilitate access to health data and the use of national and international comparisons to monitor standards. However, most importantly, they should pilot new methods of improving quality and safety through the intermediate processing of health data.

Not everything in a data and information ecosystem is good, it will be hard to test the validity and quality of the outputs of these relatively 'black box' processes. The creation of processed data will include inaccurate and misleading data, so-called information pollution.

> Expansive nature of information ecosystems would tend introduce pollution, which is the gap between information-rich and information-poor entities in the ecosystem.[20]

Our primary source of information, the Internet, presents us with a data smog, making identification of credible information sources a challenge.[21]

A health ecosystem will maximise the use of data, and create new knowledge and insights. However, within a health ecosystem, information processing and integration should be regulated to ensure that data integrity and privacy are maintained between the sources and destinations of the orchestrated information flows. There may a need to tolerate imperfect processing, but not breaches of privacy. IPHIs could be the key to achieving the vision of making better use of health data.

REFERENCES

1 de Lusignan S, Liaw ST, Krause P *et al.* Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. Contribution of the IMIA Primary Health Care Informatics Working Group. *Yearbook of Medical Informatics* 2011; 6(1):112–20.

2 Thacker SB and Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. *American Journal of Epidemiology* 1994;140(5):383–97.

3 Bain CA and Standing C. A technology ecosystem perspective on hospital management information systems: lessons from the health literature. *International Journal of Electronic Healthcare* 2009;5(2):193–210.

4 Liaw ST, Rahimi A, Ray P *et al.* Towards an ontology for data quality in integrated chronic disease management: a realist review of the literature. *International Journal of Medical Informatics* 2013;82(1):10–24. doi: 10.1016/j.ijmedinf.2012.10.001

5 Sahoo SS, Bodenreider O, Rutter JL, Skinner KJ and Sheth AP. An ontology-driven semantic mashup of gene and biological pathway information: application to the domain of nicotine dependence. *Journal of Biomedical Informatics* 2008 Oct;41(5):752–65. doi:10.1016/j.jbi.2008.02.006

6 de Lusignan S, Liaw ST, Michalakidis G and Jones S. Defining datasets and creating data dictionaries for quality improvement and research in chronic disease using routinely collected data: an ontology-driven approach. *Informatics in Primary Care* 2011;19(3):127–34.

7 Kamel Boulos MN, Sanfilippo AP, Corley CD and Wheeler S. Social Web mining and exploitation for serious applications: technosocial predictive analytics and related technologies for public health, environmental and national security surveillance. *Computer Methods and Programs in Biomedicine* 2010 Oct;100(1):16–23. doi: 10.1016/j.cmpb.2010.02.007

8 Shotton D, Portwin K, Klyne G and Miles A. Adventures in semantic publishing: exemplar semantic enhancements of a research article. *PLoS Computational Biology* 2009;5(4):e1000361. doi: 10.1371/journal.pcbi.1000361

9 Brownstein JS, Freifeld CC and Madoff LC. Digital disease detection—harnessing the Web for public health surveillance. *New England Journal of Medicine* 2009; 360(21), 2153–7.

10 Giustini D. Web 3.0 and medicine. *BMJ* 2007;335(7633): 1273–4.

11 Boulos MN, Scotch M, Cheung KH and Burden D. Web GIS in practice VI: a demo playlist of geo-mashups for public health neogeographers. *International Journal of Health Geographics* 2008;7:38. doi: 10.1186/1476–072X-7–38

12 Surrey Primary Care Trust. *Health Needs Assessment On Alcohol in Surrey.* Surrey Primary Care Trust, 2008.

13 Newton A, Sarker SJ, Pahal GS, van den Bergh E and Young C. Impact of the new UK licensing law on emergency hospital attendances: a cohort study. *Emergency Medicine Journal* 2007;24(8):532–4.

14 Morleo M, Cook PA, Bellis MA and Smallthwaite L. Use of fake identification to purchase alcohol amongst 15–16 year olds: a cross-sectional survey examining alcohol access, consumption and harm. *Substance Abuse Treatment, Prevtion and Policy* 2010;5:12.

15 Quigg Z, Hughes K and Bellis MA. Data sharing for prevention: a case study in the development of a comprehensive emergency department injury surveillance system and its use in preventing violence and alcohol-related harms. *Injury Prevention* 2012;18(5):315–20.
16 Martin N, Newbury-Birch D, Duckett J, Mason H, Shen J, Shevills C and Kaner E. A retrospective analysis of the nature, extent and cost of alcohol-related emergency calls to the ambulance service in an English region. *Alcohol and Alcoholism* 2012;47(2):191–7
17 World Health Organization. *Global Information System on Alcohol and Health*. World Health Organization: Geneva, 2012.
18 Androvitsaneas C, Sundgren B and Thygesen L. *Towards an SDMX user guide: exchange of statistical data and metadata between different systems, national and international*. Meeting of the OECD Expert Group on Statistical Data and Metadata Exchange, Geneva, 2006:6–7.
19 Naehrig M, Lauter K and Vaikuntanathan V. *Can homomorphic encryption be practical?* Proceedings of the 3rd ACM workshop on cloud computing security, 2011, pp. 113–24.
20 Capurro R. Towards an information ecology. Information quality: definitions and dimensions. In Wormell I (ed.) *Information quality. Definitions and dimensions*. Proceedings of the NORDINFO International seminar "Information and Quality", Royal School of Librarianship, 23–25 August 1989, Copenhagen. Taylor Graham: London, 1989, pp. 122–39. www.capurro.de/nordinf.htm
21 Nelson DE. Reducing information pollution in the Internet age. *Preventing Chronic Disease* 2007;4(1):A03.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Professor of Primary Care & Clinical Informatics
Clinical Informatics
University of Surrey
Guildford
GU2 7XH
UK
Email: s.lusignan@surrey.ac.uk