

# Combining Health Data Uses to Ignite Health System Learning

J. Ainsworth<sup>1,2</sup>; I. Buchan<sup>1,2</sup>

<sup>1</sup>Health eResearch Centre, Farr Institute for Health Informatics Research, University of Manchester, Manchester, UK;

<sup>2</sup>Centre for Health Informatics, Institute of Population Health, University of Manchester, Manchester, UK

## Keywords

Health data reuse, secondary uses, meaningful use, learning health systems, adaptive health systems, intelligence pipelines, health-care evidence, population health

## Summary

**Objectives:** In this paper we aim to characterise the critical mass of linked data, methods and expertise required for health systems to adapt to the needs of the populations they serve – more recently known as learning health systems. The objectives are to: 1) identify opportunities to combine separate uses of common data sources in order to reduce duplication of data processing and improve information quality; 2) identify challenges in scaling-up the reuse of health data sufficiently to support health system learning.

**Methods:** The challenges and opportunities were identified through a series of e-health stakeholder consultations and workshops in Northern England from 2011 to 2014. From 2013 the concepts presented here have been refined through feedback to collaborators, including patient/citizen representatives, in a regional health informatics research network ([www.herc.ac.uk](http://www.herc.ac.uk)).

**Results:** Health systems typically have separate information pipelines for: 1) commissioning services; 2) auditing service performance; 3) managing finances; 4) monitoring public health; and 5) research. These pipe-

lines share common data sources but usually duplicate data extraction, aggregation, cleaning/preparation and analytics. Suboptimal analyses may be performed due to a lack of expertise, which may exist elsewhere in the health system but is fully committed to a different pipeline. Contextual knowledge that is essential for proper data analysis and interpretation may be needed in one pipeline but accessible only in another. The lack of capable health and care intelligence systems for populations can be attributed to a legacy of three flawed assumptions: 1) universality: the generalizability of evidence across populations; 2) time-invariance: the stability of evidence over time; and 3) reducibility: the reduction of evidence into specialised sub-systems that may be recombined.

**Conclusions:** We conceptualize a population health and care intelligence system capable of supporting health system learning and we put forward a set of maturity tests of progress toward such a system. A factor common to each test is data-action latency; a mature system spawns timely actions proportionate to the information that can be derived from the data, and in doing so creates meaningful measurement about system learning. We illustrate, using future scenarios, some major opportunities to improve health systems by exchanging conventional intelligence pipelines for networked critical masses of data, methods and expertise that minimise data-action latency and ignite system-learning.

## 1. Introduction

The imperative for health systems to deliver better care, for more people, from fewer resources is stronger than ever. The most recent worldwide economic crisis shrank the resources available for health-care [1, 2]. The growth in demand for care-services, however, continues unabated, propelled by the twin engines of need and expectation: the needs of ageing populations living longer with naturally more chronic diseases and the persistence of unnatural risks such as diabetes through obesity; and the expectation of more care from the growth in healthcare technologies and historical investments in services [3, 4]. The promise of “personalised medicine” in particular raises demands for ever more specialized and costly diagnostics, therapeutics, devices and procedures [5, 6]. As patients, we rightly expect our health systems to be safe and effective. As consumers, we rightly expect those services to be accessible and acceptable. And as citizens, we rightly expect public resources for health-care to be organized to achieve the maximum gain for the maximum number.

So, does the data flowing through health systems work to optimise care for patients, consumers and citizens? Here we address this question in general and illustrate it using the English National Health Service (NHS) – the NHS is one of the largest and longest running universal, population-based healthcare systems, where the use of personal health data for the public good is mandated by law [7].

We use the term “reuse” to refer to any uses of individual health data outside the information system or subsystem in which the data were collected, for example total serum cholesterol levels used to inform planning for the provision of cholesterol lowering drugs (statins) in the health sys-

## Correspondence to:

John Ainsworth  
Centre for Health Informatics  
University of Manchester  
Manchester, M13 9PL  
UK  
E-mail: [john.ainsworth@manchester.ac.uk](mailto:john.ainsworth@manchester.ac.uk)

Methods Inf Med 2015; 54: 479–487  
<http://dx.doi.org/10.3414/ME15-01-0064>  
received: May 5, 2015  
accepted: June 9, 2015  
epub ahead of print: September 17, 2015

tem. Consider the payer/commissioner question ‘should 10% of the budget for prescribing statins in the primary prevention of cardiovascular disease (CVD) be diverted to smoking-cessation initiatives?’ The evidence needed to support this decision requires local data on: the prevalence and trends in CVD risk factors; incidence and trends in CVD; uptake of CVD screening services; uptake of CVD interventions; and costs of interventions and disease states. The local data are synthesised with global estimates from trials of how much the risks of various adverse outcomes such as early death or heart attack can be reduced by interventions such as statins or smoking cessation. National health policies on such topics cannot be interpreted for local populations in purely qualitative terms because there are so many different possible conformations of risk, disease and service factors. Yet the data, methods and expertise required to contextualise information like this are seldom available to local health systems. However, the investment decision needs to be made and so the options that are appraised are usually under-informed. The outcome of this investment could be used to inform future decisions, except in general it is not. Health systems frequently miss the opportunity to learn. In the US this has led to the Learning Health Systems movement [8, 9], which is similar to the NHS’ uses of data at local population levels, which we have extended previously to research in population-based “e-Labs” [10, 11].

In addition, the global evidence base for healthcare is patchy. Roughly a third of the main outcome of the average patient to the average treatment can be predicted from current knowledge-bases [12]. A typical ‘low resolution health system’ assumes that most patients respond to the same treatment in the same way. Indeed, most drug trials have excluded people with multiple diseases on the assumption that disease-drug responses operate in isolation. This assumption is increasingly recognised as flawed as real-world healthcare deals with many interacting disease and treatment factors [13]. The notion of diseases in isolation is still pervasive and reflected in the functional and organisational structures of our health systems.

## 2. Objectives

In this article we will show by way of example that current health data reuse is too little, too late, and at too high a cost of information management [14]. By contrast we examine how innovations in data reuse have transformed social and economic life over the past two decades, exploring how these paradigms may translate to the design and development of health systems and what are the prerequisite conditions for system-learning. We argue that for health systems to operate efficiently they must adapt to the needs of the populations they serve, and that they can only do this if pipelines of evidence are established that allow bidirectional exchange between research and practice. We describe how these pipelines can be implemented and used in real health systems, drawing on existing research work. Finally, we posit tests of health system learning and suggest a maturity model to encourage progress toward this challenging goal.

## 3. Methods

The conceptual synthesis presented here is based on multiple literature reviews and stakeholder consultations in shaping a regional health informatics research strategy over ten years. From 2004–12, the Northwest Institute of Bio-health Informatics (for Northwest England from a base at The University of Manchester) ran monthly research meetings focused on better digital phenotyping with healthcare data, resulting in “e-lab” and “research object” approaches to using health data more systematically [10, 15]. In 2013 this work expanded to a national colloquium as The Health e-Research Centre ([www.herc.ac.uk](http://www.herc.ac.uk)) of the UK’s Farr Institute for Health Informatics Research ([www.farrinstitute.org](http://www.farrinstitute.org)). The heuristic development of the health data reuse concepts presented herein have incorporated input from patient and public representatives, and can be seen at [www.herc.ac.uk/peppi](http://www.herc.ac.uk/peppi).

## 4. Results

### 4.1 Legacy Assumptions that Impede System Learning

We identified three assumptions of health systems and healthcare delivery that work against the adaptation and optimisation of health systems necessary for them to meet the needs of the communities they serve. These assumptions are out-dated in the face of modern data, methods and technologies, yet they pervade:

#### 4.1.1 Time Invariance: Assuming that Evidence is Eternal

The Translational Medicine movement has delivered less than expected, partly due to a flawed assumption that once produced, evidence can be translated into practice ad infinitum [16]. We illustrate this flaw with the example of predictive modelling, which is used in healthcare to provide evidence to support a wide range of decisions, from individual patient risk scoring to planning long-term investment in the provision of services. These models are produced and validated at a particular point in time. The model developers are typically rewarded with a publication and don’t revisit the research for a long time. For example, the EuroSCORE model, which is used to predict the risk of death among patients undergoing coronary artery surgery or stenting, was developed in the late 90s, then it drifted so much in calibration that it now predicts more than double the observed mortality [17, 18]. This highlights a general problem that such models are treated as static when there is a need for dynamic alternatives that can be continuously recalibrated. The evidence production line is broken, not only its translation into practice.

#### 4.1.2 Isolation: Assuming that Specialisation Is Scalable

The organisation of health systems into specialities for the treatment of disease reflects the isolation assumption. Typically a patient with cardiovascular disease (CVD) may see a cardiologist, diabetologist, renal physician or stroke specialist, each working in separate organisational units of a hospital, served by separate clinical guidelines

from separate areas of clinical research. These specialists deal with different facets of CVD. Generalist clinicians may try to knit the patient's management together across specialty areas, but the evidence pipelines are usually separate. We are not arguing for healthcare provider organisational structures to change, it simply would not be possible for an individual physician to be an expert in all these areas. However we do contend that this assumption of isolation means that evidence is not easily combined to guide the management of co-morbidities. The coproduction of health between patient and physicians can enable holistic patient care resulting in better outcomes for patients and optimised service provision [19].

#### 4.1.3 Homogeneity: Assuming that 'One Size of Evidence Fits All'

The evidence base that underpins current health systems focuses on average treatment effects in the major diagnostic categories. Healthcare planning and care quality management therefore assume that 'one size of evidence fits all'. Considerable variation in care outcomes, however, may be due to unrecognised subgroups (or endotypes) being wrongly treated as 'the average patient'. Many common chronic diseases such as asthma are actually a collection of more specific diseases, which may have different responses to treatments. [20] Basic biomedical research, for example genomics, can unpick some of this stratification in the quest for "personalised medicine". Other factors, however, may have greater influence on outcomes: for example, the social support available to a patient on discharge from hospital. The same applies at the population level, where a national policy that is easily implemented in one community may be impractical to implement in another. So, there is a need for health information systems that can 'phenotype' patients and communities more deeply, enabling healthcare resources to be titrated more specifically to needs.

## 4.2 Barriers to the Data Uses Needed for Health Systems to Learn

The assumptions identified in the previous section are not necessary when healthcare

data are reused to characterise the population that a health system serves. However, there exist barriers to healthcare data reuse that must be overcome in order to realise: dynamic, scalable and contextualised evidence. These barriers will not necessarily be the same in all health systems, communities or countries. Fragmented and disconnected data, fear of disclosure and the shortage of skilled human resources are commonplace across many health systems and are examined in this section.

### 4.2.1 Fragmentation and Disconnection of Data

We class current secondary (not for direct care) uses of clinical data as fragmented and/or detached:

*Fragmented* repositories contain a subset of care records, for example a disease-specific register, such as a cancer registry. Fragmented repositories may have been created by extracting information from care records and/or by explicit data entry. The data contained in the repository is a partial view of the patient's care and so its utility is limited.

*Detached* repositories are created from electronic health records, but they are managed and governed by an agency not directly involved in the care of the patient/citizen. The removal of the data from the community creates a number of problems:

- i. The data are under a different ethical and governance framework, therefore their uses are beyond the day-to-day control or sight of the contributing community.
- ii. The metadata, typically contained in the tacit knowledge of people in the health system, are lost, for example the knowledge of the clinical biochemist that the laboratory assay for creatinine changed at a certain time point, after which a different formula for estimating kidney function needs to be used [21].
- iii. Structured data detached from unstructured data have lower value, for example the research value of clinical codes can be increased by allied mining of clinical narrative or rubrics [22, 23].
- iv. Once a detached data repository is created, there is no way for a patient to dis-

sent (withdraw their consent and have their data removed).

- v. There is no feedback loop between the analysis of the data and the provision of health services in the communities from which the data are derived.
- vi. Anonymisation removes the capacity to link individual records across multiple data sources.

### 4.2.2 Fear of Disclosure of Person Identity

The prevailing attitude in healthcare is to minimise disclosure risk [24]. This has a direct and significant impact on health data reuse – it stifles it. The problem is that disclosure risk is considered separately from the risk management of information supply and quality. A holistic risk assessment of the healthcare system's operations is required. The technical and operational procedures for mitigating disclosure risk and well researched and widely available and so are available for organisations to apply [25]. Assessing disclosure risk in isolation results in under-information that in turn reduces quality and efficiency. It is essential to assess disclosure, quality and efficiency together.

### 4.2.3 Informatics Workforce: Capacity and Capability Gaps

Even when high quality data are made available for reuse health systems face a further problem. Where are the people to prepare, analyse and interpret these complex data? The nascent health informatics profession is under capacity and there are few education and training programmes that can deliver skilled professionals able to prepare low quality data, perform an analysis and interpret the results [26–28]. Effective reuse of healthcare data requires an understanding of how the data were created, to enable effective interpretation. This includes understanding the organisational processes in which the data were captured, for example was the patient's weight recorded because they seem overweight or as part of a systematic review? Was the clinical code assigned by a clinician during the clinical encounter or by a coding clerk? Have records been searched for missing,

remunerated diagnoses by inferring missing diagnoses from medication lists? Increasing the denominator size will not compensate for unexplained heterogeneity in datasets when trying to make robust inferences with biostatistical and epidemiological methods [29, 30]. Health systems need a workforce that can combine informatics, statistical and (clinical) epidemiological understanding on demand.

### 4.3 Technologies for Innovation in Health Data Reuse

Over the past three decades the information revolution has fundamentally changed economies and society. This revolution has been enabled by information technology. The Internet provides a global data communications infrastructure that enables us to exchange information with almost anyone without restriction. The World Wide Web provides the means to connect and search for information. More recently, social media have removed the boundaries between producers and consumers of information. Industry and commerce have exploited this infrastructure in activities that are now core to most economies [31]. This rapid growth has been fuelled by innovative ways of using the vast quantities of connected data at the level of the individual that has been an emergent feature of this information revolution. This innovation has been characterised by four paradigms: the Network Effect [32]; Crowd Sourcing [33]; the Long Tail [34]; and (Big) Data Analytics [35].

- i. The Network Effect is most visible in social networking sites whereby the value of the products and services increases in proportion to the number of users. On a social networking site where data on individuals preferences and behaviour are routinely provided by users either directly or indirectly, this can be utilised by the collectors of the data to target typically the marketing of products and services. For example, the matching of one consumer with another to provide recommendations of products based on past behaviour.
- ii. The Long Tail in retail refers to the paradigm of “selling less of more” [36]; what this means in practice is serving a

wider more diverse consumer base with the products they want. The analogous situation in healthcare would be treating less of more. What this means in practice is being able to identify individual needs and being able to provide timely, targeted care. It can also be seen as another expression of Rose’s primacy of sick populations over sick individuals [37].

- iii. Crowd Sourcing is where motivated individuals contribute on-line to a call for participation in tasks, such as raising capital, data collection [33], and problem solving.
- iv. Data Analytics, often referred to as “Big Data”, refers to the techniques of discovering hidden structure in data that can then be used to model individual behaviours and make predictions based on those models. The techniques include cluster analysis, graph databases and machine learning. It is widely used in social networking [38], recommendation systems [39], and search engine advertising [40].

Other industries/sectors have gone through a period of rapid innovation centred on using the information available to them to identify and meet the needs of their consumers/customers. Healthcare needs to learn these lessons. The three paradigms above could help healthcare to scale-up, titrating its scarce resources to the needs of better characterised populations. PatientsLikeMe ([www.patientslikeme.com](http://www.patientslikeme.com)) provides a compelling example, leveraging each of the three paradigms, and even stimulating new research to address weaknesses in the medical evidence base [41].

The current technology momentum is centred on ubiquitous computing and pervasive connectivity, now widely available via smartphones and an explosion of personal monitoring and tracking devices – the impacts are likely to be profound [42]. The Quantified Self movement reflects the self-motivation to track and monitor personal wellbeing, which is moving from a fitness niche into the mainstream [43, 44]. We are on the cusp of a data deluge, where individual data from bio-medical sensors, GPS, and accelerometers can be combined to with population and environmental data

to provide a rich longitudinal picture of our health.

Self-quantification is laying cultural and technical foundations for self-experimentation, which can translate to routinely personalised healthcare employing ‘n-of-1’ trials [45]. Health systems need to prepare to move beyond the prescription pad, to delivering packages of care that are tuned to patient experiences and outcomes reported between clinical encounters. Are we about to encounter the health data paradox – struggling to make sense of limited clinical data to optimize health systems when self-tracking, personal health monitoring and ubiquitous computing and connectivity are about to provide a deluge of high resolution data?

Health systems need to embrace the concept of coproduction [19], where data generated by patients are used to optimize care. This requires the interaction of patients, clinicians and machine algorithms to extract the signal from the data and to transform it into knowledge for shared decision-making, as shown in ► Figure 1.

Thus health information systems are set to move from separate requirements for machine-clinician and machine-patient interaction to the triangle of digital health illustrated in ► Figure 1. Communication within this triangle will need the union of records and models that represent an individual’s health to be conveyed with context-aware ‘personalities’, which is more avatar than filing cabinet (the stale paradigm of medical records) [46, 47].

For health systems, the interaction of individuals over shared health information is critical. For example citizens influencing each other to be more physically active and reducing the burden of type 2 diabetes. And in the clinical context, patients, carers and practitioners being armed with early warning of conditions deteriorating in community settings, able to take actions that avoid hospital admissions. So future health information systems must be able to support networked conversations over ‘health avatars’. The evolving concept of social machines holds promise that this will be tractable in the next decade [48].

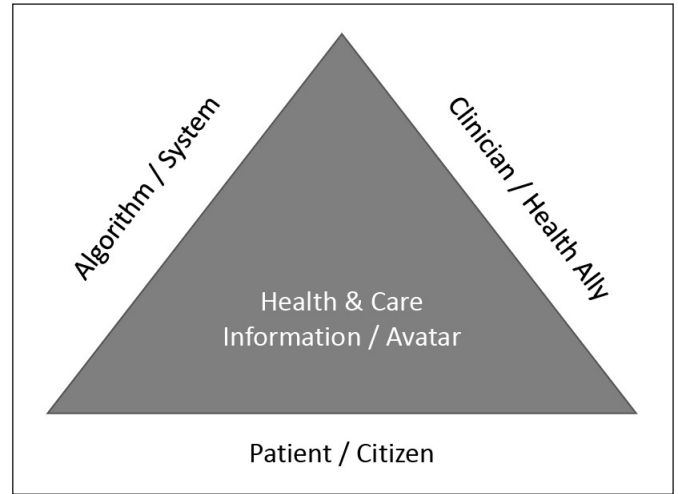
## 4.4 Architectures for Health System Learning

### 4.4.1 Bidirectional Evidence Pipelines

The assumptions of time invariance, isolation, and homogeneity identified above, can be eliminated by ensuring that timely evidence is available at the point of need, by creating new pipelines of evidence that connect the generators of data with the consumers of the data. We propose a conceptual framework for pipelines of evidence as illustrate in ► Figure 2. Each evidence pipeline essentially represents a bond of trust between the generator of the data and their consumer. A pipeline has a custodian who is the representative of the patient/citizen and has oversight of the establishment, and operation of the pipeline. A pipeline is licensed to operate for a defined period of time, and at the end of the licence period the custodian will have the power to revoke or renew the licence. A large healthcare organisation such as a hospital may manage a data node that has all of these roles, whereas a small information intermediary might have just one. One organisation might act on behalf of others in respect of a role, for example transforming data to maintain a mortality risk model for per cutaneous coronary intervention, or consuming data to lead on a supra district audit of the same intervention.

Each pipeline may have zero or more transformers, the role of which is to process the raw data to produce information or knowledge. The consumer may use the resulting evidence for the purpose defined by the custodian. The pipeline of evidence conceptual framework ensures that the in-

**Fig. 1**  
The health & care co-production triangle



terests of all parties are represented, the reasons for and the usage of the pipeline is well defined and that governance rests with the representative of the patient. Wherever possible the pipeline should incorporate a feedback loop delivering information and/or knowledge derived from the data. This feedback loop enables the generator to adapt and optimise.

To illustrate this with the EuroSCORE example given above: the data generators are cardiac surgeons and cardiologists completing audit forms; the custodian is the National Institute for Cardiac Outcomes Research in the UK; the transformers are researchers modelling the risk of patients dying during, or within 30 days of, procedures to unblock coronary arteries; and the consumers are providers and payers monitoring the quality of services between provider organisations and individual practitioners. Currently, there is no systematic feedback, which means that poor quality data continue to be collected,

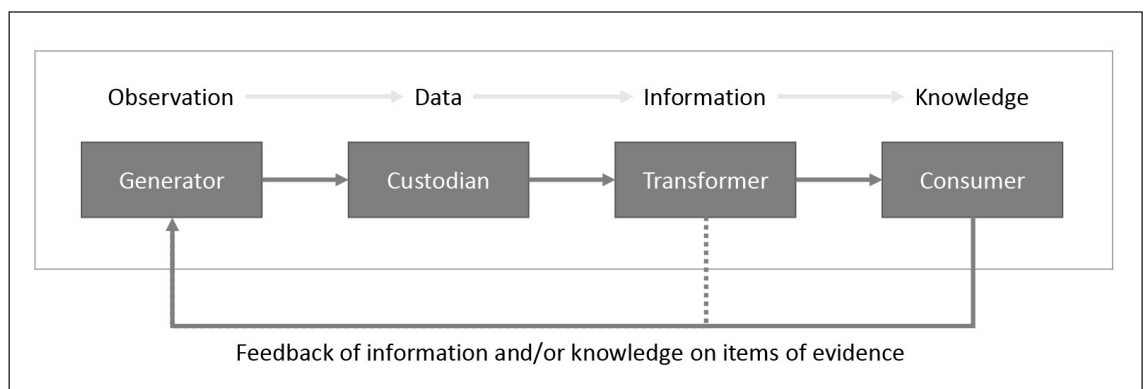
including a binary flag for renal function instead of a more informative creatinine value. Systematic feedback to the practitioners whose performance is being measured may improve data quality, and feedback to researchers on the calibration drift of their published model may trigger timely recalibration.

Under this framework health systems would transition from the current unidirectional model of ‘evidence into practice’ to a bidirectional model of evidence pipelines with feedback.

### 4.4.2 Distributed Sense Making

In order to eliminate the barriers to health data reuse described above we developed the population health e-Lab model [10, 49]. Under this model, communities may integrate pipelines of evidence from primary care, secondary care, specialist care sources such as disease registries, and administrative sources such deaths and

**Fig. 2**  
Conceptual framework for evidence item pipelines



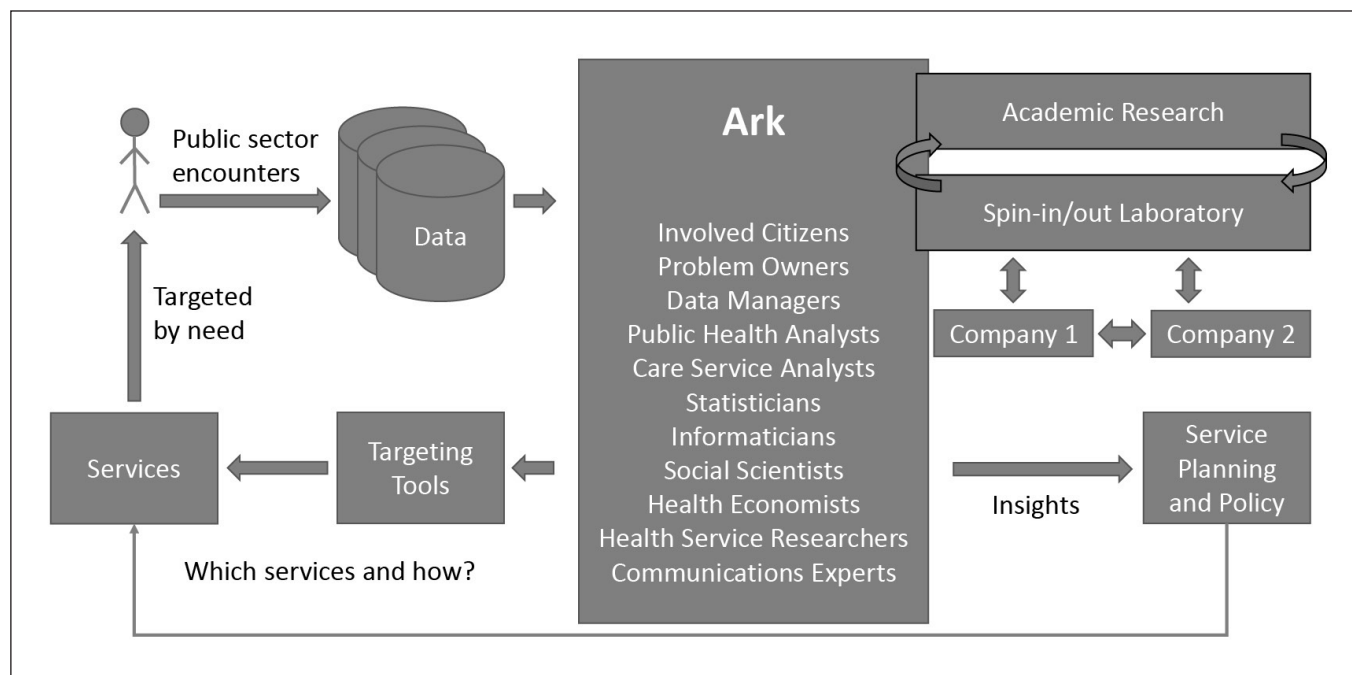


Fig. 3 Connected Health City model for health system learning

demographic data – this would be natural for the formation of an integrated health intelligence system for that community. Record linkage, across multiple pipelines, is performed for the purposes of: informing individual patient care; managing the quality of local care-services; commissioning/planning care services; public health intelligence; and research. This linkage takes place within the secure network of the local healthcare agencies before removing data that would unnecessarily identify individuals to analysts. Deep analysis of the data can then be performed within the same information governance environment used to protect identifiable records. Thus the risk of deductive disclosure of individual identities is managed.

In the UK, laws permit large-scale linkage of health data under an opt-out consent model for multiple purposes, including research. [7] This has been difficult to apply across large populations such as England's 53m but it is happening at the level of natural health systems (populations of around 1–6m and sub-districts thereof) [50]. Such systems map to geographically-defined populations with cohesive cultural identities such as small countries, metropolitan districts and civic regions. Civic-centred e-Labs for handling large-scale health data

afford tangible public oversight of data uses. So research with opt-out consented data has a different ethical context to more remote uses of the data, with greater clarity over autonomy, beneficence, non-maleficence and justice in determining data uses [51].

There are two important situations where different health systems may need to borrow analytic strength from each other without sharing potentially disclosive data: 1) where an expert analyst is external to custodian of the data; and 2) where research findings need to be replicated across heterogeneous populations/settings. Here we have developed the Research Object [15] (or Work Object [10]) model whereby computable representations of a research protocol can travel between e-Labs gathering results while leaving potentially disclosive data within the custodian's firewall.

Coordinated research across a global network of population health e-Labs could effectively eliminate the problem of non-reproducible research that plagues the medical literature [29, 52].

#### 4.4.3 Connected Decision-making

So when does a health system have enough connectedness of data, methods and ex-

pertise to start learning in computable ways? In addition to bidirectional evidence pipelines and distributed sense-making we consider the connection between evidence and person-accountable decision-making to be the other missing link for igniting health system learning. The evidence base on which most healthcare decisions are taken is currently not recorded. This applies to population/system level planning decisions, clinical and self-care decisions alike.

Recognising this problem, the North of England (population 15m) is building four pilot "Connected Health Cities" that will use care pathway optimisation as the organising principle for seeding learning health systems. In this model, a critical mass of data, methods/tools and expertise is brought together in one physical and digital location under strict healthcare information governance. We call the location an Ark to reflect the dual metaphor of a covenant of trust with the public over the uses of their personal data, and protection from the data flood that might otherwise impeded system learning. The information/action flows for targeting care along pathways is illustrated in ► Figure 3.

The Ark is an extension of the e-Lab model to defined populations and comput-

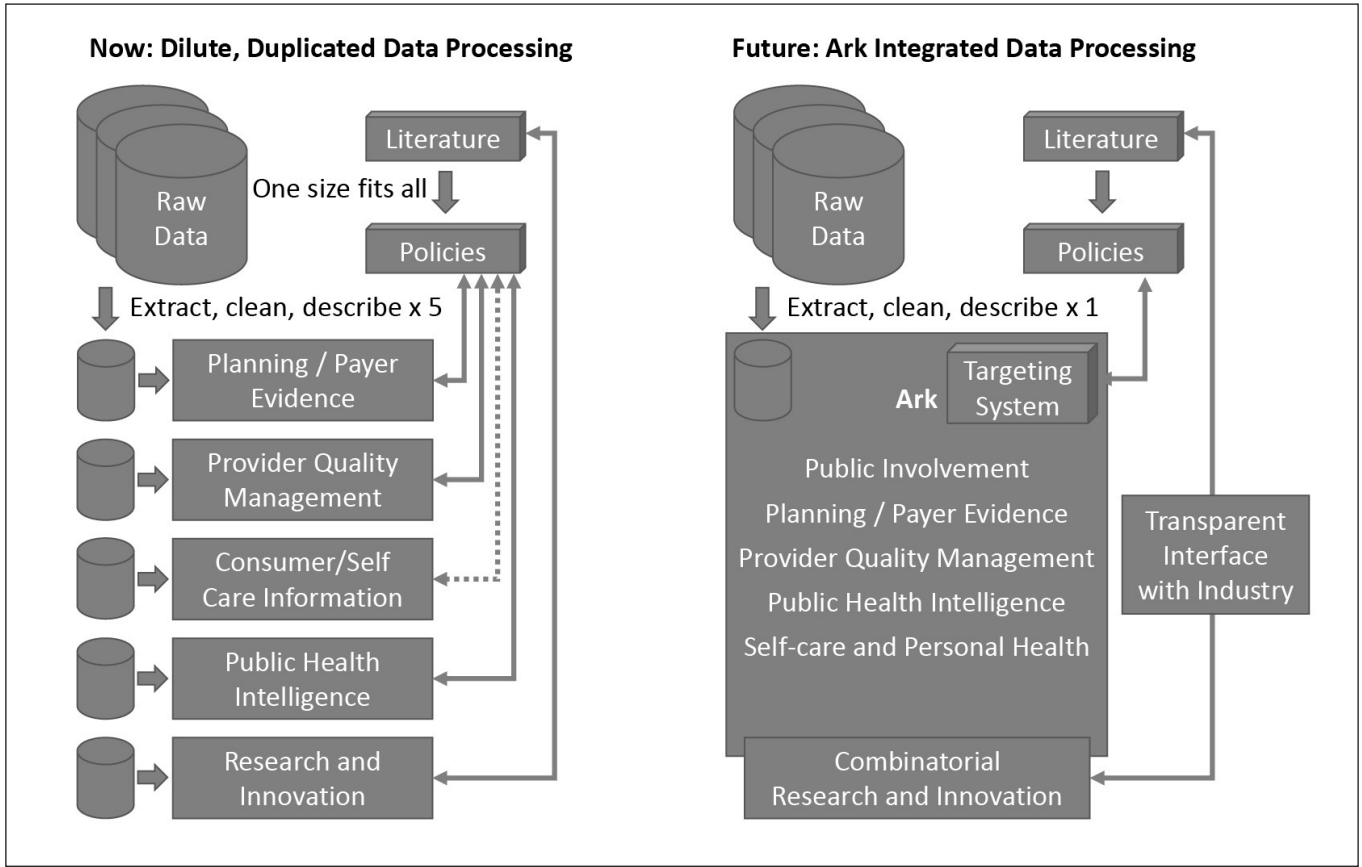


Fig. 4 Reducing data-processing duplication by connecting up a health system to learn

able information/action loops. It works not only for the population served but also offers external researchers and enterprises ways to interact with the Ark's large-scale linked data in ethical, efficient ways. For example, where a company needs clinical prediction algorithms to be developed/-calibrated they can 'spin-in' to the Ark using the object model described above and be validated in a trusted-third-party manner, employing local contextual knowledge that would not exist in the company. Similarly, a network of Arks using similar operating procedures can provide replication facilities for research or algorithm development across heterogeneous populations/settings.

The creation of the Ark critical mass should also save the health system money by reducing the current duplication of data processing where disparate teams of analysts run separate data warehouses and overlapping analysis pipelines as shown in ► Figure 4.

## 5. Discussion

The transition from data-sparse to data-intensive healthcare cannot be achieved through technologies and system management alone, as there is a major culture change involved. We note a train of relevant culture change in healthcare from the introduction of integrated care pathways in the 90s, [53] through health systems integration in the 00s to current aspirations of personalised and precision medicine [54]. These are not fashions they are phases in moving from the expected information core being the clinician to it being the health system [47].

Clinician performance is already measured using quality standards derived from the clinical practice guidelines that underpin care pathways. In most parts of the world these metrics emerged in the 80s and 90s from clinical audit/governance/quality management initiatives. The ideal was the "close the audit loop", which is in effect a

'learning health sub-system'. The challenge of quality management is greater in primary care than more specialist areas: it covers more clinical processes & outcomes so is more data-hungry. The UK was able to establish a Quality and Outcomes Framework on the back of its established capture of structured clinical data by general practitioners [55]. Work has undertaken to integrate this 'quality conversation' between primary and secondary care by analysing linked care records to follow patient journeys along care pathways between provider organisations, comparing observed with guideline-suggested care, as illustrated in ► Figure 5 [56, 57]. It is our belief that cross-provider pathway optimisation is a potential ignition point for system-wide learning, particularly when coupled with service planning/commissioning.

We posit data-action latency (DAL) as the key metric of health system learning: this is the time lag between data being

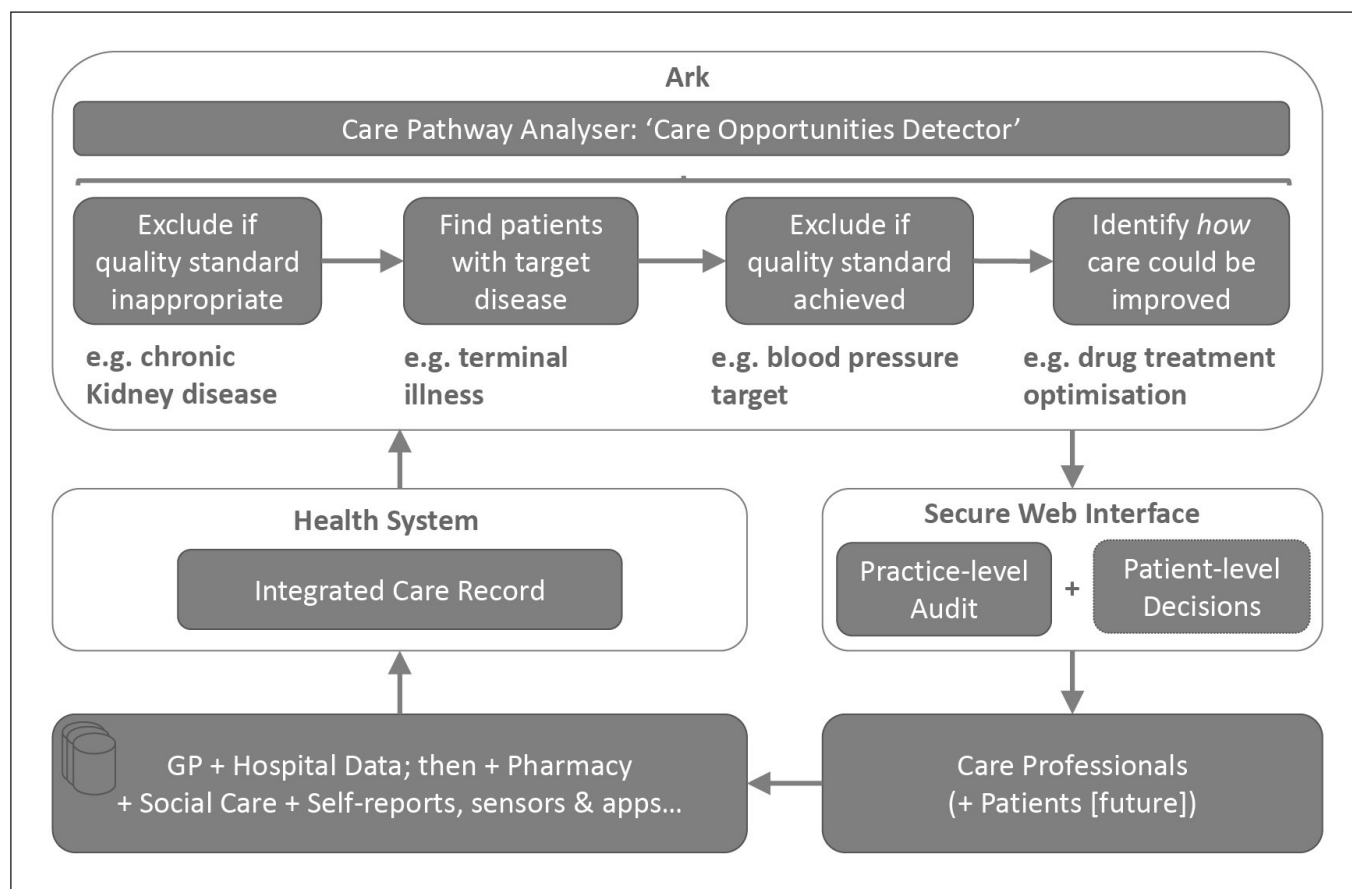


Fig. 5 Care pathway optimisation, a catalyst for health system learning

available that an action should be taken and the action being taken. Minimising DAL requires concerted data capture, aggregation and analysis followed swiftly by interpretation of results, assignment of responsibility for any actions, and recording of actions. Bidirectional evidence pipelines will enable the measurement of DAL.

## 6. Conclusions

Current health data reuse is too little, too late, at too high a cost of information management. Pressures on health services from obesity, diabetes, and aging populations require more to be done with less. There is a pressing need for health systems that: can adapt to the needs of the individuals and population they serve; and can optimise the usage of resources to provide the maximum health gain to the maximum number. We have described three legacy assumptions (structural, organisational and tech-

nical) of health systems and argued the case that they prevent health systems from delivering optimal care, but can be fixed by building pipelines of evidence that enable health systems to adapt to their communities, through the integration of multi-source healthcare data. Our analysis of existing examples of data re-use, reveals some fundamental issues that need to be resolved. As health informaticians we have much to learn from the socioeconomic information revolution and the way innovation has been driven by the Long Tail, Crowd Sourcing, the Network Effect and (Big) Data Analytics. Similarly we must fully grasp the opportunity that ubiquitous computing and pervasive connectivity will provide to understand individual patients in unprecedented detail. We will all increasingly be producers and consumers of health information and need to establish pipelines of evidence between patients, algorithms and clinicians to enable the co-production of care.

## Acknowledgments

This work was supported by funding from the MRC Health eResearch Centre grant MR/K006665/1 and NIHR Greater Manchester Primary Care Patient Safety Translational Research Centre grant.

## References

1. Lafond S, Charlesworth A, Roberts A. Hospital finances and productivity: in a critical condition? London: 2015.
2. Neville S. Financial crisis adds to health providers' woes – FT.com. *Financ Times* 2015.
3. Fineberg H V. A Successful and Sustainable Health System – How to Get There from Here. *N Engl. J Med* 2012; 366 (11): 1020–1027.
4. Wanless D, Appleby J, Harrison A, Patel D. *Our Future Health Secured?: A review of NHS funding and performance*. London: 2007.
5. Rigby M. Personal Health, Person-centred Health and Personalised Medicine – Concepts, Consumers, Confusion and Challenges in the Informatics World. *Yearb Med Inform* 2012; 7 (1): 7–15.
6. Haycox A, Pirmohamed M, McLeod C, Houten R, Richards S. *Through a glass darkly: economics and*



- personalised medicine. *Pharmacoeconomics* 2014; 32 (11): 1055–1061.
7. The Stationary Office. Health and Social Care Act 2012. UK: 2012.
  8. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, Gunter C, Musen M, Platt R, Stead W, Sullivan K, Van Houweling D. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. *J Am Med Informatics Assoc* 2014; 22 (1): 43–50.
  9. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. *Sci Transl Med* 2010; 2: 57cm 29.
  10. Ainsworth JD, Buchan IE. e-Labs and Work Objects: Towards Digital Health Economies. In: Mehmood R, Cerqueira E, Piesiewicz R, Chlamtac I, editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009 pp 205–216.
  11. Powell J, Buchan I. Electronic health records should support clinical research. *J Med Internet Res* 2005; 7 (1): e4.
  12. Valderas JM, Starfield B, Sibbald B, Salisbury C, Roland M. Defining comorbidity: implications for understanding health and health services. *Ann Fam Med* 2009; 7 (4): 357–363.
  13. Fortin M, Dionne J, Pinho G, Gignac J, Almirall J, Lapointe L. Randomized Controlled Trials: Do They Have External Validity for Patients with Multiple Comorbidities? *Ann Fam Med* 2006; 4: 104–108.
  14. Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, Leese A, Richardson C, Mantas J, Murray P, De Moor G. Trustworthy reuse of health data: A transnational perspective. *Int. J Med Inform* 2013; 82: 1–9.
  15. Bechhofer S, Buchan I, De Roure D, Missier P, Ainsworth J, Bhagat J, Couch P, Cruickshank D, Delderfield M, Dunlop I, Gamble M, Michaelides D, Owen S, Newman D, Sufi S, Goble C. Why linked data is not enough for scientists. In: *Future Generation Computer Systems*. 2013 pp 599–611.
  16. Abernethy AP, Wheeler JL. True translational research: bridging the three phases of translation through data and behavior. *Transl Behav Med* 2011; 1 (1): 26–30.
  17. Grant SW, Hickey GL, Dimarakis I, Cooper G, Jenkins DP, Uppal R, Buchan I, Bridgewater B. Performance of the EuroSCORE models in emergency cardiac surgery. *Circ Cardiovasc Qual Outcomes* 2013; 6 (2): 178–185.
  18. Hickey GL, Grant SW, Murphy GJ, Bhabra M, Pagano D, McAllister K, Buchan I, Bridgewater B. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. *Eur J Cardiothorac Surg* 2013; 43 (6): 1146–1152.
  19. Nesta. People Powered Health: Co-Production Catalogue 2012.
  20. Lötvall J, Akdis CA, Bacharier LB, Bjerner L, Casale TB, Custovic A, Lemanske RF, Wardlaw AJ, Wenzel SE, Greenberger PA. Asthma endotypes: A new approach to classification of disease entities within the asthma syndrome. *J Allergy Clin Immunol* 2011; 127: 355–360.
  21. New J, Aung T, Baker P, Yongsheng G, Pylpoczuk R, Houghton J, Rudenski A, New R, Hegarty J, Gibson J, O'Donoghue D, Buchan I. The high prevalence of unrecognized anaemia in patients with diabetes and chronic kidney disease: a population-based study. *Diabet Med* 2008; 25 (5): 564–569.
  22. Lependu P, Iyer S V, Fairon C, Shah NH. Annotation Analysis for Testing Drug Safety Signals using Unstructured Clinical Notes. *J Biomed Semantics* 2012; 3 (Suppl 1): S5.
  23. Overhage JM, Ryan PB, Schuemie MJ, Stang PE. Desideratum for evidence based epidemiology. *Drug Saf* 2013; 36 (Suppl 1): S5–14.
  24. Information Governance Taskforce. Privacy Impact Assessment Risk Stratification. Leeds: 2014.
  25. Khokhar RH, Chen R, Fung BCM, Lui SM. Quantifying the costs and benefits of privacy-preserving health data publishing. *J Biomed Inform* 2014; 50: 107–121.
  26. Kuhn KA, Knoll A, Mewes H-W, Schwaiger M, Bode A, Broy M, Daniel H, Feussner H, Gradinger R, Hauner H, Höfler H, Holzmann B, Horsch A, Kemper A, Krcmar H, Kochs EF, Lange R, Leidl R, Mansmann U, Mayr EW, Meitinger T, Molls M, Navab N, Nüsslin F, Peschel C, Reiser M, Ring J, Rummeny EJ, Schlichter J, Schmid R, Wichmann HE, Ziegler S. Informatics and medicine – from molecules to populations. *Methods Inf Med* 2008; 47 (4): 283–295.
  27. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh AHB. Big data: The next frontier for innovation, competition, and productivity. *McKinsey Glob Inst* 2011; (May): 156.
  28. BBSRC AND MRC Review of Vulnerable Skills and Capabilities. London: 2014.
  29. Thompson S, Kaptoge S, White I, Wood A, Perry P, Danesh J. Statistical methods for the time-to-event analysis of individual participant data from multiple epidemiological studies. *Int J Epidemiol* 2010; 39 (5): 1345–1359.
  30. Ioannidis JPA. Why most published research findings are false. *PLoS Med*. 2005; 2 (8): e124.
  31. Pelissié Du Rausas M, Manyika J, Hazan E, Bughin J, Chui M, Said R. Internet matters: The Net's sweeping impact on growth, jobs, and prosperity. 2011.
  32. Sundararajan A. Network Effects (Arun Sundararajan, part of the Industrial Organization of Information Technology Industries web site). [date unknown]; <http://oz.stern.nyu.edu/io/network.html>. Accessed 2015–4–11.
  33. Brabham DC. Crowdsourcing as a Model for Problem Solving: An Introduction and Cases. *Converg Int J Res into New Media Technol* 2008; 14 (1): 75–90.
  34. Brynjolfsson E, Hu Y (Jeffrey), Simester D. Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Manage Sci* 2011; 57 (8): 1373–1386.
  35. Lavalle S, Lesser E, Shockley R, Hopkins MS, Kruschwitz N. Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Manag Rev* 2011; 52 (2): 21–32.
  36. Anderson C. The long tail: why the future of business is selling less of more. New York: Hyperion; 2006.
  37. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 2001; 30 (3): 427–432.
  38. Tusoo Ashish, Zheng Shao, Suresh A, Dhruva Borthakur, Namit Jain, Joydeep Sen Sarma, Raghotham Murthy, Hao Liu. Data warehousing and analytics infrastructure at facebook. In: *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*. ACM, 2010. pp 1013–1020.
  39. Sumbaly, Roshan, Kreps J, Sam Shah S. The big data ecosystem at linkedin. In: *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*. ACM, 2013. pp 1125–1134.
  40. Dean J, Ghemawat S. MapReduce: Simplified Data Processing on Large Clusters. *Proc. 6th Symp Oper Syst Des Implement* 2004. pp 137–149.
  41. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nat Biotechnol* 2011; 29 (5): 411–414.
  42. PwC. Socio-economic impact of mHealth; An assessment report for the European Union. 2013.
  43. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 2012; 14 (2): e46.
  44. Wolf G. The quantified self. TED@Cannes 2010.
  45. Guyatt G, Sackett D, Taylor DW, Chong J, Roberts R, Pugsley S. Determining optimal therapy – randomized trials in individual patients. *N Engl J Med* 1986; 314: 889–892.
  46. Buchan I. Informatics for Healthcare Systems. In: Walshe, Kieran, Smith, Judith, editors. *Healthcare Management*. McGraw-Hill International; 2011. pp 321–336.
  47. Buchan I, Winn J, Bishop C. A unified modeling approach to data-intensive healthcare. *Fourth Paradigm Data-Intensive Sci Discov* Washington, DC. Microsoft Res 2009. pp 91–98.
  48. Hendler J, Berners-Lee T. From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. *Artif Intell* 2010; 174 (2): 156–161.
  49. Ainsworth J, Cunningham J, Buchan I. ELab: Bringing together people, data and methods to enhance knowledge discovery in healthcare settings. In: *Studies in Health Technology and Informatics* 2012. pp 39–48.
  50. McCartney M. Care.data: why are Scotland and Wales doing it differently? *BMJ* 2014; 348: g 1702.
  51. Beauchamp T, Childress J. *Principles of Biomedical Ethics*. 7th ed. Oxford: Oxford University Press; 2012.
  52. Open Science Collaboration. An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspect Psychol Sci* 2012; 7 (6): 657–660.
  53. Campbell H, Hotchkiss R, Bradshaw N, Porteous M. Integrated care pathways. *BMJ* 1998; 316 (7125): 133–137.
  54. Mirnezami R, Nicholson J, Darzi A. Preparing for precision medicine. *N Engl J Med* 2012; 366 (6): 489–491.
  55. Kontopantelis E, Buchan I, Reeves D, Checkland K, Doran T. Relationship between quality of care and choice of clinical computing system: retrospective analysis of family practice performance under the UK's quality and outcomes framework. *BMJ Open* 2013; 3 (8): e003190.
  56. Ainsworth J, Buchan I. COCPIT: a tool for integrated care pathway variance analysis. *Stud Health Technol Inform* 2012; 180: 995–999.
  57. Brown B, Williams R, Ainsworth J, Buchan I. Missed opportunities mapping: computable healthcare quality improvement. *Stud Health Technol Inform* 2013; 192: 387–391.