Human-Data Interaction in Healthcare: crossing the chasms between the uses of health information

Angela Locoro[1] and Federico Cabitza[1]

[1] Università di Milano-Bicocca, Viale Sarca, 336, 20126 Milano, Italy
angela.locoro@disco.unimib.it, cabitza@disco.unimib.it

Abstract. In this paper, we focus on a new strand of research: Human-Data Interaction (HDI). This regards both how humans create and use data, by means of interactive systems that can both assist and constrain them, as well as to passively collect and proactively generate data. Healthcare provides a challenging arena to test the potential of HDI to provide a new, user-centered and ethnomethodologically informed perspective on how data work should be supported and assessed, especially in the light of the fact that data are becoming increasingly big and that many tools are now available for the lay people, including doctors and nurses, to interact with health-related data.

1 Introduction

Twentyfive years ago, medical informatics was defined as "dealing with the storage, retrievel and optimal use of biomedical data" (Shortliffe et al. 1990). At that time, little emphasis was put on the practices of data production, that is on how medical practice, and single stories of illness, care and recovery are represented, accounted and "factified" in some objective manner. However, these practices, which include policies, rules, habits, conventions, tools and techniques, have always been intertwined with and affected by the available ITs, as well as by the expectations of the stakeholders on how to make sense and use of health-related data. Different perspectives on these expectations, and on what valuable data are, lead to manifest chasms between primary use and other uses of health information, as often discussed in the CSCW literature (Fitzpatrick and Ellingsen, 2013). To try to cross these chasms, we need to create the suitable language to describe them and give some operational definitions.

2 Perspectives on Human-Data Interaction

We distinguish between primary data, which come from a broad range of sources and are produced both within a caring process and for its unfolding; and derivative data, which are produced from the primary data (Abdelhak, Grostick, & Hanken,

2012). We also distinguish between two kinds of derivative data: secondary data and tertiary data. The analogy from the agriculture domain is intended: primary data are like the produce of the land, which farmers grow for themselves as well as the external market. Secondary data are the product of a transformation of these primary data, like the one going on in food industry where vegetables are cleansed and chopped. Tertiary data are further and possibly (not necessarily) transformed from secondary data to make them consumable and valuable, that is conveyed to a broader population of consumers in terms of information services, like fresh-cut vegetable products can be seen as the service to have vegetables already ready-to-eat.

This tripartition reflects the different uses in which data are produced and consumed. The "primary use" of health information is "to use it to directly support patient care", both by aiding medical decision-making and by ensuring continuity of care by all providers, that is both interpretation of medical signs (represented by data) for decision making, and coordination among the actors involved "around" the patient (Berg, 1999). Secondary use regards both other uses of the same data collected for the primary use within the administrative domain, and the generation of derivative data for other aims than care, like billing and reimbursement, performance and care quality evaluation, resource planning and management, service design and public policy making. Tertiary use regards the heterogeneous uses through which end consumers put tertiary data to the test of (their) life, and hence the publication and dissemination of valuable indications for the citizens and the taxpayers about the available healthcare services, so as to enable the benchmarking and comparison of care facilities.

The distinction between primary, secondary and tertiary data is quite different from others discussed in the Information Quality (IQ) literature: for example, the one inspired by the manufacturing domain (Shankaranarayan et al., 2003) which distinguishes between raw data, component data items, that is semi-processed information and information products, which are composed out of these items. Primary data are not necessarily raw (Gitelman, 2013), because they are meaningful to and usable by the primary consumers that generated them. On the other hand, both secondary and tertiary data are unfinished information products: the former ones are resources for specialist work (both clerical and managerial) within specific organizational boundaries and processes, while tertiary data result from the enactment of information services conceived for the non specialist and the external consumers, that is for the public. Thus, while the distinction introduced by Shankaranarayan and his colleagues focuses on a incremental definition of the information product, the distinction we propose focuses on the different roles that produce and consume different information products that are definitive in relation to their context of use.

The tripartition that we propose allows to interpret the phenomenon of "low" quality of primary data (and hence of any derivative information) within the still unresolved chasm between primary use and secondary uses of data in healthcare (Fitzpatrick 2000; Wears and Berg 2005; Berg and Goorman 1999), and between secondary use and tertiary use, which has not yet been deeply investigated. The primary and secondary chasm goes beyond the usual tension between clinical vs. administrative purposes, while the secondary and tertiary chasm goes beyond the mere making secondary data more open and accessible, because also the end

consumers' readiness to access, comprehend and exploit them, as well as their unanticipatable purposes, are to be considered.

3 A Research Agenda

We trace back the challenge to bridge these chasms in terms of the problem of reusing primary data, which is "always entangled with the context of its production" (Berg and Goorman, 1999), in different contexts (either secondary or tertiary ones). As Berg and Goorman note, reuse is possible only if data are made "transportable", that is sufficiently disentangled from their context of production, and this can occur only if specialized additional work (that is processes) is performed on data. The point is then how to conceive and perform these additional processes.

Our approach grounds on the following points. We would argue that quality of information should not be assessed irrespective of the distinction between primary / secondary / tertiary uses, that is, by adopting the same metrics and methods in a context-independent manner: on the contrary, data used in care processes should be evaluated on the basis of the efficacy they enable appropriate and timely action (fit to use), also on the basis of work conventions and tacit knowledge that are difficult to bring back to the usual dimensions of accuracy, completeness and consistency (Gregory et al., 1995). Likewise information quality in services for the general public should be evaluated in terms of the extent to which information is informative and can be appropriated by its consumers and appreciated in their lives (i.e., according to its social value).

We believe that a socio-technical and ecological perspective that integrates and enhances the results of two main research streams is needed:

1) mutual awareness between data producers and consumers, whose objective is to increase commitment and awareness of consequences in the producers of the primary data; and, at the same time, raise awareness in the secondary consumers of the contextual and social nature of primary data and hence of their limitations (Cabitza and Simone, 2012a) . This can be achieved, for example, by endowing the interfaces by which data are collected and presented with specific affordances that adapt to the context according to specific business rules in order to convey the so called "awareness promoting information" (Cabitza and Simone, 2012b); this can be done also by means of simple visual clues (Cabitza et al., 2013) like text highlighting or side messages, which do not impose any behavior to the data producers but help their interpretation.

2) support for understandability of data analytics and visualization, whose objective is to support the transformation of secondary data that are openly available online for the general purpose users (i.e., typically the taxpayers), that is on how to make tertiary data out of them. The transformation of data into information services does not necessarily require a massive processing of data but rather the application of state-of-the-art human interaction techniques to develop interactive infographics and highly tailorable dashboards that enable user-friendly online analytical processing and hence the transformation, even by end users

(Lieberman, Paternò, Klann, & Wulf, 2006), of secondary data into socially valuable information.

Focusing on this third stream, that is on tertiary data, would call researchers to take Human-Data Interaction (HDI) seriously. We use this expression to refer to a new research field that explicitly and programmatically applies the tenets and approaches of Human-Computer Interaction to the design and development of the interaction between the people and the data through which they make informed decisions (the readers should mind that human-data interaction is not human data interaction, that is it is not solely focused on the interaction with human data, or with one' own personal data (Crabtree and Mortier, 2015), but it includes also these dimensions recently considered within the CSCW field).

HDI covers three phases: design, development and evaluation of the systems by which to extract information and support knowledge in data-intensive application domains. In particular, HDI regards the user-centered elicitation of better requirements of configuration, adaptation and appropriation of big data analytics cockpits and dashboards to optimize usability and the user experience, i.e., efficiency, effectiveness and satisfaction; the application of End-User Development techniques and tools to allow end-users tweak the tools by which data are extracted and visualized; and user-centered methodology for the assessment and continuous improvement of the quality of the interaction of the humans with their data of interest, so as to reduce both information overloading and information funneling/complacency (Parasuraman and Manzey, 2010), and improve awareness. This also includes the exploration of new techniques toward better interactive visualization environments and above all better data-telling, that is the capability to build and share stories that can explain data and facilitate correct interpretations (e.g., in medical domain by adopting a natural frequency approach - Hoffrage et al., 2002) and to allow, instead of curbing, the social exchange within a community of data-users of multiperspective, sound and viable interpretations around the data that are supplied by the computational systems to inform decision making and knowlegeable action.

References

1. Abdelhak, M., Grostick, S., & Hanken, M. A. (Eds.). (2012). Health information: management of a strategic resource (4th ed). Elsevier.
2. Berg, M., & Goorman, E. (1999). The Contextual Nature of Medical Information. IJMI, 56, 51–60.
3. Berg, M., & Toussaint, P. (2002). The mantra of modeling and the forgotten powers of paper: A sociotechnical view on the development of process-oriented ICT in health care. JMI, 69(2-3), 223–34.
4. Cabitza, F., & Simone, C. (2012a). "Whatever works": Making Sense of Information Quality. Phenomenology, Organizational Politics, and IT Design, 79, IGI Press.
5. Cabitza, F., & Simone, C. (2012b). Affording Mechanisms: An Integrated View of Coordination and Knowledge Management. CSCW, 21(2), 227–260.
6. Cabitza, F., Simone, C., & Colombo, G. (2013). ``Worth a thousand fields''. Arguing for a visual turn in computer-supported general practice. In eHealth 2013, IADIS Press.

7. Crabtree, A., & Mortier, R. (2015). Human data interaction: Historical lessons from social studies and CSCW. In ECSCW 2015. Springer.
8. Fitzpatrick, G. (2000). Understanding the Paper Health Record in Practice: Implications for EHRs. In HIC'2000.
9. Fitzpatrick, G., & Ellingsen, G. (2013). A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. CSCW, 22(4-6), 609-665.
10. Gitelman, L. (2013). Raw data is an oxymoron. MIT Press.
11. Gregory, J., Mattison, J. E., & Linde, C. (1995). Naming notes: transitions from free text to structured entry. MIM, 34(1-2), 57–67.
12. Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. Cognition, 84(3), 343-352.
13. Lieberman, H., Paternò, F., Klann, M., & Wulf, V. (2006). End-User Development: An Emerging Paradigm. In End User Development, 9, 1–8.
14. Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors: Hum. Factors, 52(3), 381–410.
15. Roberts, A., Gaizauskas, R., Hepple, M., & Guo, Y. (2008). Mining clinical relationships from patient narratives. BMC Bioinformatics, 9(11), S3.
16. Sauleau, E. A., Paumier, J.-P., & Buemi, A. (2005). Medical record linkage in health information systems by approximate string matching and clustering. BMC Medical Informatics and Decision Making, 5(1), 32.
17. Shankaranarayan, G., Ziad, M., & Wang, R. Y. (2003). Managing data quality in dynamic decision environments: An information product approach. JDM, 14(4), 14–32.
18. Shortliffe, E.H., Perreault, L.E., Wiederhold, G., and Fagan, L.M. (eds.) (1990). Medical Informatics: Computer Applications in Health Care and Biomedicine. Addison-Wesley.
19. Smith, P. C., Araya-Guerra, R., Bublitz, C., Parnes, B., Dickinson, L. M., Van Vorst, R., Pace, W. D. (2005). Missing clinical information during primary care visits. Jama, 293(5), 565–571.
20. Swinglehurst, D., Greenhalgh, T., & Roberts, C. (2012). Computer templates in chronic disease management: ethnographic case study in general practice. BMJ Open, 2(6)
21. Timmermans, S., & Berg, M. (2003). The gold standard: the challenge of evidence-based medicine and standardization in health care. Temple University Press.
22. Wears, R. L., & Berg, M. (2005). Computer Technology and Clinical Work: Still Waiting for Godot. JAMA, 293(10), 1261–3.