

# A Map is worth a Thousand Data: Requirements in Tertiary Human-Data Interaction to Foster Participation

Angela Locoro

Dipartimento di Informatica, Sistemistica e Comunicazione, Università degli Studi di Milano  
Bicocca, Viale Sarca 336, 20126 Milano

**Abstract.** This paper aims to shed light on an emerging class of phenomena that are related to the abundance of data, which either come from personal life records or from open data portals, and to the strategies to tame this abundance to enable the human making of sense and decision. In particular, a new category can be introduced for these kind of data, which are neither primary (that is being deeply engrained into a work practice), nor secondary (i.e., processed by and for specialists, like the members of clerical, managerial or research staff) but, in a way, “tertiary”, as they are consumed (when not directly produced) by “final customers”. Hence, new ways of engaging users to either enable or facilitate the direct comprehension, and the ad-hoc manipulation and tailoring of data to unpredictable and unstructured tasks should be devised, also by means of more active and inter-active visualization techniques, in order to reduce the information overload and to let users shape their data landscapes in a virtuous cycle that may also return benefit to the same (primary) production and uses of data. In so doing, both data can partake in the end-users lives and these latter partake in improving the quality of data in face of the current (open) big data tide.

**Keywords:** tertiary data, human-data interaction, data visualization, mapping data.

## 1 Introducing tertiary human-data interaction

Information is one of the most abundant resources that are available in our society. This is thanks to the multiple sources of data that we can get access to and the multiple means by which we can interpret them. However, when data come in big quantities, so that also the phenomenon of information overload can be detected, then real life interaction with those data to reach their informative gist can become a daunting task. It is frequent the case in which data are just too many, loosely structured, highly redundant and complex [2]. In order to generate order and meaning out of them, people usually have to also rely on visual aids, annotations, and context, to mention only a few of the well-known affordances for data interpretation [3],[4].

In this paper, we introduce the expression “Human-Data Interaction” (HDI) denoted as “the human manipulation, analysis, and sense making of large, unstructured, and

complex datasets” [5]. In so doing, we can address the “different nature and dynamics of HDI” with respect to the traditional and bi-directional stances of Human-Computer Interaction (HCI), and we can draw a distinction between data elaborated for specific tasks and with specific goals, and the massive amount of information published “out there”, without knowing or worrying about how they will be accessed, interpreted and exploited by disparate people to widely different aims. A new class of problems emerges within this complex ecosystem of personal and open data, which “could greatly benefit from novel and intuitive data aggregation, summarization, analytics and visualization techniques” [5].

As a first step of our work, we would like to extend the concept of HDI by enriching it through a tripartite distinction. We distinguish between *primary data* (and *primary human-data interaction*), which come from a broad range of sources and are produced both *within* a practice and *for* its orderly unfolding; the main characteristics of primary data is to be entangled with the practices of a community (of practice) and therefore not easily detachable from their context of creation, manipulation and use. We call *derivative data* [1], what is produced from the primary data for other aims than those related to the original practices of data production and use. Within this very broad class of derivative data, we distinguish between *secondary data* and *tertiary data*. Accordingly, a distinction is made between activities pertaining to *secondary human-data interaction*, on one hand, and to *tertiary human-data interaction*, on the other hand. The former ones encompass the processes by which clerical, managerial and research staff make sense of primary data, often by also processing them to make them more suitable to their professional-oriented (and even sometime biased) interpretation, by representing data in suitable formats, i.e., typically tabular data at various level of aggregation with respect to primary data. Tertiary human-data interaction encompasses activities that mostly pertain to the use of data from social categories of people such as policy-makers, final users, tax-payers and citizens, broadly speaking. Most of these data users do not share the same purposes and, most notably, the same capabilities of other categories of professionals in dealing with “their data”.

An analogy from the agriculture domain may be drawn for the sake of clarity to illustrate our intended tripartition of data and interaction-with-data: primary data are like the produce of the land, which farmers grow for themselves as well as the external market, in terms of “raw” material. Secondary data are the product of a transformation of these primary raw data, like the one going on in food industry where vegetables are cleansed, chopped, and, after a precise categorization and selection, properly wrapped. Tertiary data are further and possibly (though not necessarily) transformed from secondary data to make them easily 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 tangible service to have vegetables already ready-to-eat or ready-to-cook. This latter distinction is particularly useful in order to frame the interaction-with-tertiary-data problem with a problem of active users participation to the process of managing such data according to a collective and perhaps data-community orientation.

## 2 Motivations of our study and research questions

In our previous research, we analysed several open data initiatives, which are characterized by a great number of datasets that are made available online by local governments and institutions and accessible by everyone. A rapid glance at their formats of presentation and contents unveils how most of these outputs are produced according to a secondary human-data interaction perspective: more precisely, the predominant publication style is strongly accounting-oriented, and little or none is left to the potential social value of data for citizens' lives, individual decision making, and social well being at large, which remain still untapped [14,15], or worse yet, hampered as in the healthcare domain [17].

Some initiatives towards a “tertiary” (i.e., layman-oriented) use of data are, for example, those of the Open Indicator Consortium<sup>1</sup> (OIC); this is a learning community of organizations and individuals that are experimenting more powerful ways of presenting data on the Web to achieve greater social impact. In particular, these ways are related to the so called “mapping data”, i.e., presenting them on richly rendered and geo-referenced maps, or to create customized eBooks, where users are supposed to have “some data to use” and tweak among different rendering modalities (e.g., a chart, a map or a table) so to choose which one fits their needs best in any context of use. OIC seeks to advance data visualization for community improvement by tools such as, for example, WEAVE<sup>2</sup>. This is a Web-based visualization platform focusing on the ‘functional art’ [11] of mapping data, that is enriching pictures of local areas with information of various kinds, e.g., crimes, public transports, cultural points of interests, geo-located dietary and healthy activities, and so forth.

We believe that the topic of tertiary data visualization, being part of the wider data visualization field but also exhibiting still under-researched peculiarities that regard user engagement, participation and sense-making, deserves more efforts also to investigate the most effective ways to enable data appropriation and its consequences on daily actions. In particular, in this paper we focus on the dimensions of end user participation in the process of data visualization for a particular use that is both *situated* and *casual*: here we refer to both the idea that data usage is always aimed at some purpose, mediated by the interpretation of someone embedded in a particular social and cultural context and situation [12], but also to the idea that such a use is “casual” because carried out by “casual users” [2] using “casual infovis” [8]. In so doing, we highlight the need to visualize data in tertiary HDI differently from how they could be presented for other, often professional, purposes. Tertiary HDI entails a multiple, sporadic, and short term but socially relevant interaction, which can be improved by an ergonomic study of tools of data visualization<sup>3</sup>, on the one hand, and by the direct participation of users in the tailoring of their data [16] to get greater situated relevance, on the other hand. This process may, in its turn, also benefit the primary data

<sup>1</sup> <http://openindicators.org/>

<sup>2</sup> <http://oicweave.org/>

<sup>3</sup> <http://www.forbes.com/sites/teradata/2014/12/17/the-ergonomics-of-human-data-interaction/>

producers, as it hints at the most valuable information they should invest on for improved publication and maintainance.

Thus, our main research question is: what are the challenging requirements that underlie data visualization for tertiary HDI from secondary tabular data formats (like in HTML tables, CSV or similar files)? How active users engagement with data visualization tools may contribute to tertiary data interaction and sensemaking?

### **3 Requirements to make visualizations we (may) live by**

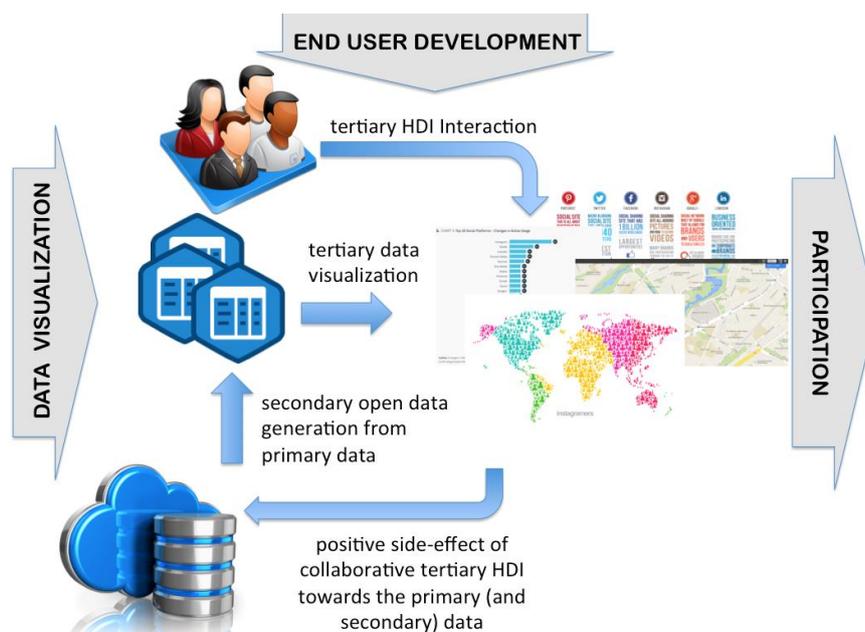
The problem of mapping tabular data, their dimensions, and their relations with visual forms (e.g., maps, charts, icons, geometric structures, and so forth) belongs to our visual perception capabilities (see [6] and [9]), to the possibility of letting end user develop their tools for the spatial organization of data and, above all, to the problem of making users partake in the active management of tertiary data visualization in their everyday practices, in order to enable the customization of their local sphere of life influence, decisions and actions.

Some studies put an emphasis on the character of casuality and inexperience of (tertiary data) users and on the meta-cognitive dimensions that can be triggered by visual tools besides the analytical insight. In particular, tools for ambient, social and artistic infovis have been analyzed in [8] and some of the traits that this not-analytical tools let emerge of the human-data interaction are related, for instance, to colors, shapes and dimensions of objects that change to reflect data flows (the “ambient infovis”); to the visualization of social networks to make people more aware of community situations for social inclusion (the “social infovis”); and to the challenging of beliefs and preconceptions through the promotion of the awareness of the recording, computational, and processing mechanisms of the computer infrastructure that are usually hidden to users (the provocative and reflective insights of “artistic infovis”).

As expressed in Figure 1, lay users should cope with tools that enable to interact with data at a different level of granularity: for instance, they should have the possibility to go back and forth from an analytical view of the data to an iconic representation, so that the requirements of accuracy and transparency can be met. In so doing, they could choose the dimensions to emphasize in data visualization, among the many dimensions that a dataset may express, so that the requirement of informativeness is met); and also the more natural forms and “aesthetic structures” into which to enframe data, so that they can manipulate them and make sense of them, so that the requirements of usability (or accessibility), easiness and clarity are met as well.

To recap, then, the main challenge we detect is to empower the lay consumer in tailoring the multiple dimensions of the visualization space so that it can fit her need and enable insight. This can be achieved, for example, by providing users with 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 [13], of secondary data into socially valuable information.

Fig. 1. The participation challenge in tertiary human-data interaction.



#### 4 Conclusions

By definition, technologies (may) empower our perceptual and cognitive capabilities; since these are especially affected by our visual apparatus, the importance of technologies devoted to enhance data presentation and consumption cannot be overstated. Technology “may enhance our peripheral reach by bringing more details into the periphery. This is encalming when the enhanced peripheral reach increases our knowledge and so our ability to act without increasing information overload” [7]. The new challenges and opportunities that novel approaches to the data visualization for tertiary human-data interaction are of great interest, especially for the potential of these technology to both deliver data in ways that make them more valuable in situated and casual use (in the line of the old saying “knowledge is power”), and to improve the participation of lay people to the great machine of data use (in terms of better interpretation) and production (in terms of quality improvement). This short contribution is aimed at referring interested researchers to the new paradigms of human-data interaction and the techniques of tertiary data visualization that could allow the current open and big data availability positively affect human action, instead of jeopardizing its efficacy in redundancy and overload.

## Acknowledgement

This research has been supported in part by FP7/2013-2015 COMSODE (under contract number FP7-ICT-611358).

## 5 References

1. Luciano Floridi. Two approaches to the philosophy of information. *Minds Mach.* 13, 4 459–469, (2003)
2. Elmqvist, N. Embodied Human-Data Interaction. In Proceedings of CHI 2001, (2001).
3. Ardito, C., Bottoni, P., Costabile, M. F., Desolda, G., Matera, M., Piccinno, A., & Picozzi, M. Enabling end users to create, annotate and share personal information spaces. In *End-User Development* (pp. 40-55). Springer (2013)
4. Ardito, C., Costabile, M. F., Desolda, G., Lanzilotti, R., Matera, M., & Picozzi, M. Visual composition of data sources by end users. In *AVI 2014* (pp. 257-260). ACM (2014)
5. Haddadi, H., Mortier, R., McAuley, D., & Crowcroft, J. Human-data interaction. *University of Cambridge* (2013)
6. Few, S. Data visualization for human perception. *The Encyclopedia of Human-Computer Interaction, 2nd Ed* (2013)
7. Weiser, M., & Brown, J. S. The coming age of calm technology. In *Beyond calculation* (pp. 75-85). Springer New York (1997)
8. Pousman, Z., Stasko, J. T., & Mateas, M. Casual information visualization: Depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1145-1152 (2007)
9. S. K. Card, J. Mackinlay, and B. Shneiderman. Readings in Information Visualization, Using Visualization to Think, chapter 1, pages 1–34. Morgan Kaufmann (1999)
10. Ralph Lengler and Martin J. Eppler. Towards a periodic table of visualization methods of management. In GVE '07, ACTA Press, Anaheim, CA, USA, 83-88 (2007)
11. Cairo, A. The Functional Art: An introduction to information graphics and visualization. New Riders (2012)
12. Suchman, L. Human-machine reconfigurations: Plans and situated actions. Cambridge University Press (2007)
13. Lieberman, H., Paternò, F., Klann, M., & Wulf, V. End-user development: An emerging paradigm (pp. 1-8). Springer Netherlands (2006)
14. Batini, C., Locoro, A., Cabitza, F. Conceptual Schemas as a Means to Compare and Measure Open Data Social Value. Procs of SEBD (2015)
15. Cabitza, F., Locoro, A., Batini, C. A User Study to Assess the Situated Social Value of Open Data in Healthcare, Procs of HCIST (2015)
16. Cabitza, F. and Simone, C. Building Socially Embedded Technologies: Implications about Design. In Wulf, V., Schmidt, K. And Randall, D. (Eds.) Designing Socially Embedded Technologies in the Real-World. Springer (2015)
17. Cabitza, F., Simone, C., & Colombo, G. ``Worth a thousand fields''. Arguing for a visual turn in computer-supported general practice. In eHealth 2013: Proceedings of the IADIS International Conference on eHealth, pp. 95–102. IADIS Press (2013)