The value of not being lost in our digital world

Elena N. Naumova[1]

Hardly a day passes without a reminder that data–big data, large unsurmountable volumes of data–are being generated. And we expect them to be processed, examined, and digested. Data are making their own footprints in history, threatening us with information overload, attention fatigue, and generating a solid mistrust of almighty algorithms. Yet, the excitement over what has already been achieved by science and industry and for our daily lives, plus what else can be achieved with additional data sources and streams, ought to fend off negativity, data doomsday predictions. ToxicDocs' creators conclude that rediscovered information gems "will remain a resource free and open to all, anywhere in the world," supporting their excitement [1].

Invented 600 years ago, the printing press boosted the literacy of the masses and broke the literate elite's monopoly on education. I see a parallel in the efforts of today's historians and librarians to convert printed text into a digital form and bring us valuable information–to boost data literacy. Without careful digitizing and archiving, the assembled material is likely to be lost. In the digital age, if it can't be searched, it is neither available nor accessible. It is simply lost for the broad audience. ToxicDocs adds fuel to the information revolution. It provokes the interest of those who are looking for answers and those who are interested in how to get answers in innovative and efficient ways.

As data are acquired, they have to be retrieved, verified, compiled, explored, summarized, visualized, and interpreted. Then we can reach the ultimate goal of converting data into actionable knowledge. Each of these steps requires data literacy and critical conceptual thinking. The renowned Dutch computer scientist, Edsger Dijkstra, described the most desired ability to be mastered in the digital age: "to be

✉ Elena N. Naumova
Elena.Naumova@tufts.edu

[1] Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA

able to think in terms of conceptual hierarchies that are much deeper than a single mind ever needed to face before."

Big Data, in many instances, demand the ability to work effectively in interdisciplinary teams. Until now, team-working skills have not been encouraged by our educational system nor by government grant making. But academic institutions and publishers have opportunities to promote careers that reward team science. Data scientists, together with researchers, may be able to offer answers to new and complex questions. By developing metrics within funding mechanisms that ensure effective review of transdisciplinary science, processes that produce meaningful results from secondary data analysis may emerge.

Efforts, such as those started by the ToxicDocs project, are likely to bring together many enthusiasts. As data sets become ever larger and more complex, users will need to master skills for linking, managing, mapping, and communicating their findings. In the digital age, key discoveries will be made by enabling data harmonization, building data linkages within and across data repositories, forming solution-oriented ontologies and vocabularies, enabling domain-specific capacity for integration of data from basic, environmental, social, and population sciences, and by developing new standards for reporting research results based on secondary data analysis.

To make sense out of compiled data, consumers' data literacy has to keep pace with data processing speed. Effective and efficient delivery of data, transformed into knowledge, will lead to effective and efficient solutions to big problems and to relevant policies. All recipients of ToxicDocs' free information: "investigative journalists, toxicologists, policymakers, historians of public health like ourselves, environmental justice advocates, and the general public" must now embrace lifelong learning. Can we teach data sciences so that the densest materials become digestible, meaningful, and thought provoking?

Of course, there must be some protective caution. While building a data repository, the task of digitizing, arranging, and archiving is overwhelming. Yet the life of the data repository after its creation is no less challenging. Only this year, scientists and environmental advocates found themselves frantically trying to protect climate data. Were governments going to protect "inconvenient" data stored on public domains, data containing valuable yet disturbing information? Already data were dismissed, tweeted to be 'fake.' As reported by the *New York Times* on 20 January 2017: "Scientists fear the online deletions [of climate change data] will extend far beyond changes to introductory websites and into the realm of government data" [2].

Will anyone take charge to maintain ToxicDocs.org and similar sites that will surely grow exponentially? As these valuable preprocessed data repositories grow, substantial resources will be needed. Are we ready to embrace the challenge that 'with big data come big responsibilities'? Are we ready to recognize that valuable data are treasures, and have to be treated as such? The amount of data on all aspects of our lives will grow in response to technologies and concerted efforts to protect and disseminate knowledge.

These questions will not go away. Rather, they are likely to bring new ones, making this early thinking by the creators of ToxicDocs.org even more valuable.

References

1. Rosner D, Markowitz G, Chowkwanyun M. Guest Editorial: ToxicDocs (http://www.ToxicDocs.org): From history buried in stacks of paper to open, searchable archives online. J Public Health Policy [special section] "ToxicDocs: Opening a new era of evidence for policies to protect public health" (Guest Eds. Rosner D, Markowitz G, Chowkwanyun M). 2018; 39(1); doi:10.1057/s41271-017-0106-8.
2. Davenport C. With Trump in charge, climate change references purged from website. New York Times, 20 January 2017.

Elena N. Naumova, Ph.D., is Professor of Mathematics in the Friedman School of Nutrition Science and Policy and Director of the Initiative for the Forecasting and Modeling of Infectious Diseases (InForMID) at Tufts University, Boston, MA, USA. She is also the Statistical Editor of the *Journal of Public Health Policy*.