

# Digital Health and a New Drug Discovery

We are facing an industrial revolution in the era of Big Data and artificial intelligence (AI) which affects all fields, including biomedicine. This was possible thanks to the capability to store a great amount of data and to the growing speed of computers which are able to perform more operations in a shorter time. This industrial revolution started first in society and in the relationship between people, driven by commonly used social media, and has finally reached medicine, where it is changing our approach to understanding and fighting diseases. Data sets grow rapidly in part because they are increasingly gathered by cheap and numerous information-sensing internet devices such as mobile phones. These technologies are entering in medicine too, speeding up the collection and increasing the amount of data. For instance, epidemic surveillance can now rely on large but low-quality data from social media to complement traditional public health approaches. Algorithms for automatic image analysis are rapidly outperforming trained humans in the classification of certain types of cancers or cognitive disorders including Alzheimer's, schizophrenia, etc.

I started to work on these topics (AI and Big Data) about 15 years ago when very few people in biomedical research were trying to explore quantitative biology or quantitative imaging and apply machine or deep learning to biological data. Publishing these results, which were highly interdisciplinary, was not easy and it was also difficult to find the right journals that could appreciate a multidisciplinary point of view.

Like every revolution, in the beginning it is often ignored by most, but then suddenly becomes popular, contributing to a disruptive change. In fact, innovation does not happen gradually, and this is also true for learning and for other important steps in human life.

The main achievement of digitalisation in biomedicine is the possibility to extract useful information from all these data. It is well known that AI works very well on image recognition and indeed this has been its first application in biology. However, nowadays the mainstream is the possibility to connect genomic data to transcriptomes, proteomes and images in a multiplex network (Figure 1). This allows the identification of key aspects of a specific subject and the proposal of a personalised treatment. In this connection, the most pressing issue is to develop computational tools that are able to integrate and analyse data coming from different sources and, at the same time and more importantly, to ask the right questions to the data. In Figure 2 is shown a typical workflow where the data that are usually unstructured should become structured and then integrated, eliminating possible batch effects. Thus, they can be analysed using machine learning, through a complex network analysis or unsupervised learning to cluster the data and identify subcategories. The system is circular since the results and their meaning can help improve the analysis of other data. Combining data into large sets allowed us to better explore hidden but relevant information that would not appear when analysing small data sets because of the inevitable background noise. Big data analysis in biology is, however, still difficult and one of the main sources of difficulty is related to the fact that data shared within the scientific community are not uniform in their format. The potential of using these approaches, not only in the discovery of new drugs but also

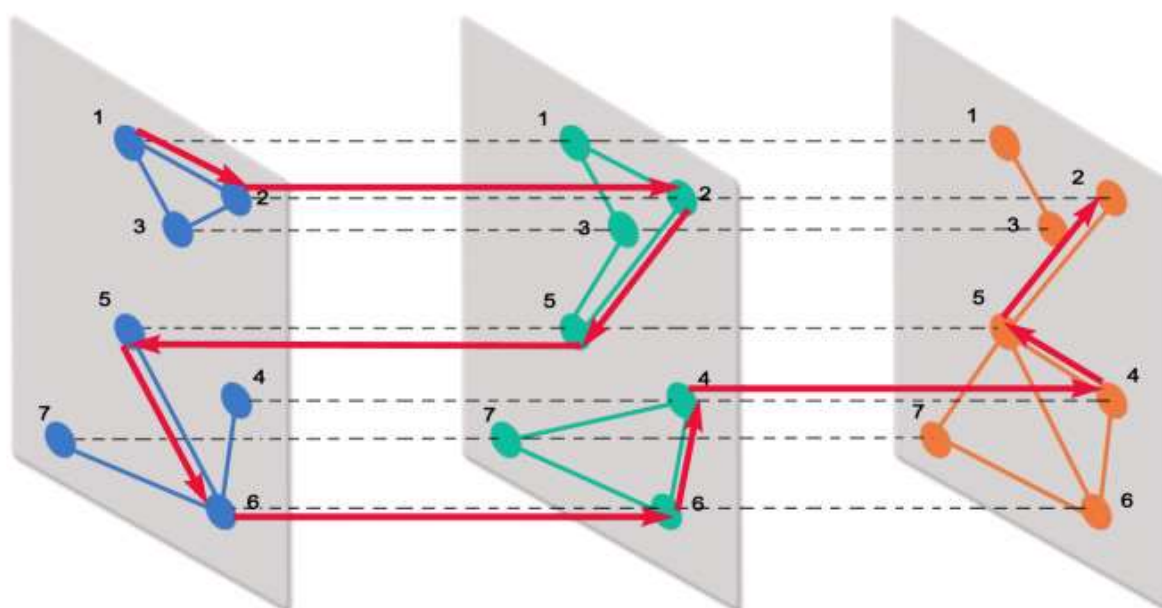


Figure 1. An illustration of a multilayer network (CC BY-SA 4.0 Manlio de Domenico).

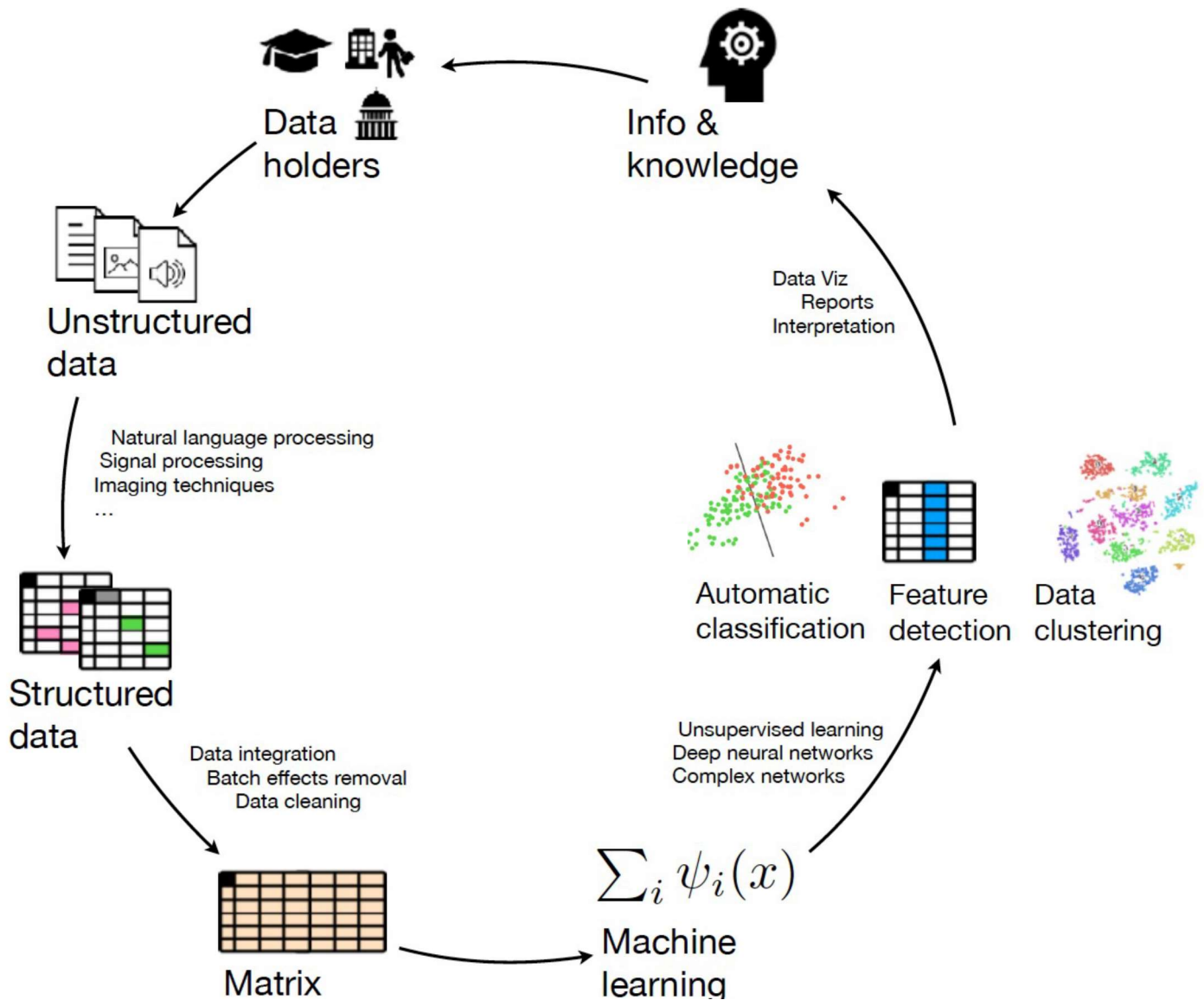


Figure 2. The cycle of data. Image by Francesc Font-Clos.

during the treatment of a patient, would be very powerful. I provide here a simple example: if we had a platform that was able to segment the patients between responding and non-responding to a specific drug during treatment, it would be possible to perform a personalised medicine and the physician could modify the treatment accordingly. Pharmaceutical companies are also changing their approach and start to appreciate the potentially interesting information hidden in already available data. Real-world data (RWD), in fact, include all the data going beyond what is usually collected in Phase III clinical trials and also any outcome that is not purely interventional. Healthcare decision-makers are now starting to devise policies based on integrated evidence coming from a multitude of sources. This is important because it can provide information that goes beyond what has been obtained in the trial, such as the way a particular drug works in populations not covered by the trial. Finally, RWD includes information about the actual treatment patients received, also including co-morbidity, so that RWD can be used to study the effect of multiple interventions.

### Caterina AM La Porta



Caterina AM La Porta is an expert of Digital Health, Professor of General Pathology at the University of Milan, co-founder of the Center for Complexity & Biosystems of the University of Milan ([www.complexitybiosystems.it](http://www.complexitybiosystems.it)), group leader of the OncoLab ([www.oncolab.unimi.it](http://www.oncolab.unimi.it)) and CEO of ComplexData ([www.complexdata.it](http://www.complexdata.it)), an Innovative startup of University of Milan. She developed and brought to the market ARIADNE ([www.ariadneweb.it](http://www.ariadneweb.it)) an innovative platform that use AI to calculate the score of aggressiveness of triple negative breast cancer. ARIADNE has filed an international patent and got the CE mark form April 2021.

Email: [caterina.laporta@unimi.it](mailto:caterina.laporta@unimi.it)