Substantia
An International Journal of the History of Chemistry

Firenze University Press
www.fupress.com/substantia

OPEN ACCESS

Citation: L. Campanella, L. Teodori (2019) Where does chemistry go? From Mendeelev table of elements to the big data era. *Substantia* 3(2) Suppl. 6: 9-11. doi: 10.13128/Substantia-741

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Editorial

Where Does Chemistry Go? From Mendeelev Table of Elements to the Big Data Era

Luigi Campanella[1], Laura Teodori[2,*]

[1] *Department of Chemistry "Sapienza" University of Rome, P.zzale Aldo Moro, 5, 00185, Rome, Italy*
[2] *Laboratory of Diagnostics and Metrology, FSN-TEFIS-DIM, ENEA-Frascati, Via Enrico Fermi, 44, 00044 Rome, Italy*
*E-mail: laura.teodori@enea.it

Who is each of us if not a combination of experiences, information, readings, imaginations? Every life is an encyclopedia, a library, an inventory of objects, a sample of styles, where everything can be continually re-mixed and rearranged in all possible ways " (from Italo Calvino, American Lessons, Six Memos for the Next Millennium, 1988)

One hundred and fifty years ago the Russian chemist Dmitri Ivanovich Mendeleev published the first "Periodic System of the Elements" originated to display the periodic trends of the chemical elements known at that time and possibly to predict unknown elements supposed to fill the empty spaces, by predicting their properties. His prevision turned out to be essentially correct. He had about sixty elements in his periodic table of 1869. Other naturally occurring elements were discovered or isolated in the following years, and various further elements have also been produced synthetically. In his honor element 101, discovered in 1905, was named "**mendelevium**". The modern periodic table, of 118 elements now, constitutes an important framework for exploring chemical reactions; it provides the basis for the discovery or the synthesis of further new elements and for the development of new theoretical models. Although other chemists at the time of Mendeleev attempted to organize the known chemical elements in a system, the extraordinary and visionary intuition of Mendeleev was to use the trends in his periodic table to predict the properties of missing elements. The philosophy behind the Mendeleev conceptions about systemizing the extant knowledge of chemistry and possibility to predict the missing information, thanks to the network support, can be considered a pioneering approach of the new science called "Systems Chemistry" and the harbinger of the modern "Predictive Chemistry". Indeed, systems chemistry is defined as "the science which study the networks of interacting molecules, to create new functions from an ensemble of molecular components at different hierarchical levels with emergent properties" [1]. As in any systems science, systems chemistry too benefits of the massive outburst of big data.

Big Data indicate data sets large and/or complex enough, that traditional processing and analysis are not sufficient. Now as then, in the Mendeleev's age, the need for rationalizing and systemizing data is compelling. Indeed, in the case of big data, one must deal with a large amount of data with the need of dimension reduction, as in the process of zipping them, to compress large quantity of data into smaller equivalent sets. Statistical/computational intelligence tools such as principal component analysis, fuzzy logic, neuro-computing, evolutionary computations etc. are developed to reduce the size of big data sets and extract valuable information. In this regard, we see the today-approach towards data-driven chemistry as an evolution of the Mendeleev philosophy, rather than a revolution. Dmitri Mendeleev was actually the first to envision the possibility to systemize chemical knowledges in a frame were much space would be available to the unknown elements which would fit within a "systemic" view of the system, and he was, therefore a real pioneer of the modern predictive data science able to extract knowledge or insights from large data sets. The figure of Dmitri Mendeleev has inspired much fascination and his story about the idea that he said to have had it envisioned in a dream is amazing: he dreamed all the elements falling into the right place. However, we think that his philosophical thoughts had not influenced and not reported enough by the historians of science. As confirmation of this idea is the fact that Mendeleev never got the Nobel Prize although candidate several times: in 1901, 1905 and 1906, but he lost because, according to the committee, his work was already too old and well known: paradoxically, the Mendeleev's table was victim of its own success. In 1906 the Nobel award went instead to Henry Moisson for the discovery of fluorine, an element that was right were the table predicted to be.

The following year Mendeleev died, and so his table of the elements could not boast a Nobel. However, we think that with the advent of Systems Chemistry, Mendeleev's philosophy of logic systematization and prediction of missing elements is taking a rematch. Being the focus of systems chemistry research on the overall network of interacting molecules and on their emergent properties, the way in which specific interactions between the components propagate through the system may predict these emergent properties. The term "systems chemistry" was first used in 2005 by Von Kierowski[2]. He stated that: "combining kinetic, structural, and computational studies on complex dynamic feedback systems may lead to the field of systems chemistry". The approach is exemplified by the analysis of a simple organic self-replicating system that has the potential to express both homochiral autocatalysis and heterochiral cross-catalysis. Von Kiedrowski claimed that this new approach could pave the way to a new field he named "systems chemistry", that is to say, the design of prespecified dynamic behavior. Later on, this proposal moved away from its reductionist approach to the study of multiple variables simultaneously [3,4,5]. Several topics related to systems chemistry bring also philosopher and existential questions such as: what made possible on the prebiotic Earth the "transmutation" of a complex mixture of molecules into living chemical systems?; why the biochemical building blocks of life were selected and how some of these biomolecules developed to have specific chirality? The latter poses fundamental questions about the origin of chiral asymmetry in biological molecules which still remains without answer[6,7]. Systems chemistry attempts to address these issues by creating synthetic systems models with properties that could reflect aspects of prebiotic biogenesis. Another topic at the core of systems chemistry is the quest for de novo life.

However, systems chemistry encompasses much more than these issues and put forward a plethora of new opportunities for the discovery of dynamic figures in all areas in chemistry. In 2005 in Venice during a conference an early consensus definition of systems chemistry was established as below[8]:

- A conjunction of supramolecular and prebiotic chemistry with theoretical biology and complex systems research addressing problems relating to the origins and synthesis of life.
- The bottom-up pendant of systems biology towards synthetic biology.
- Searching for a deeper understanding of structural and dynamic prerequisites leading to chemical self-replication and self-reproduction.
- The quest for the coupling of autocatalytic systems, the integration of metabolic, genetic, and membrane-forming subsystems into protocellular entities.
- The quest for the roots of Darwinian evolvability in chemical systems.
- The quest for chiral symmetry breaking and asymmetric autocatalysis in such systems.

Since then, systems chemistry has had a big boost due to the advent of data science tools.

Data science is defined as a multi-disciplinary science that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data[9]. It has been presented as the fourth pillar of science (being theory, experimentation and simulation the other three). With the advent

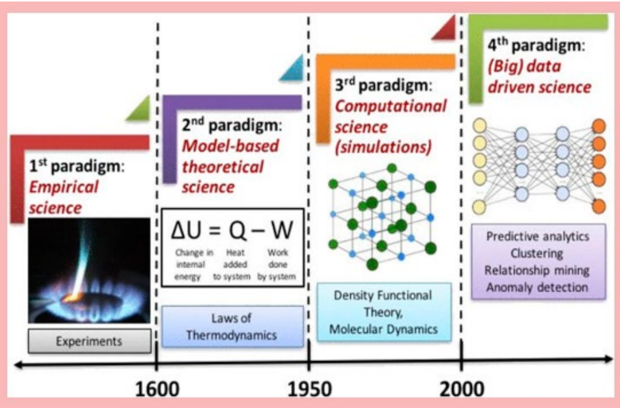

Figure. The four paradigms of science: empirical, theoretical, computational, and data-driven. Image from Agrawal and Choudhary [10].

of "omics" in life sciences (genomics, proteomics, transcriptomics, metabolomics etc.) and the advent of modern high-throughput techniques of analytical chemistry and molecular biology we are able to produce a huge amount of data. Thus, the way we undertake research is presently changed and the data drive science is considered the fourth paradigm (see Figure). The increasing rate of data generation in all scientific disciplines is providing incredible opportunities for data-driven research, transforming our current processes. The exploitation of so-called 'big data' will enable us to undertake research projects never possible before but also stimulate us to re-evaluate our previous data.

The 2002 was identified as a turning point in data and a landmark year when digital took over from analog. Indeed, it was observed that in 2009, more data worldwide were produced than all the preceding years put together. The advent of the big data age changed irreversibly the paradigm of science. Thousand year ago, science was empirical, based on, or confirmed by observation rather than theory or logic speculations. A few hundred years ago science was based on theoretical models. A few decades ago, when computer modeling simulation was introduced to understand complex phenomena, the paradigm of science changed again. Today we are witnessing the coming of the fourth paradigm of science which unifies theory, experiments, simulation, computation, creating big data sets and entering the era of "Data Science" or "Systems Sciences", originating the fourth paradigm of science which is data-driven discovery. The possibility of collecting big data has surpassed, by far, the present capability of analyzing them. At this purpose more and more dedicated, open-source "high-performance computing platforms" are being developed. Open-access data repositories, where multiple databases or files or experimental results are loaded by scientists, are the backbone of these platforms and stimulate a collaborative attitude among scientists.

Unfortunately, data science approach represents still a rather unexplored field among the community of chemical scientists, thus, limiting many opportunities for advancing chemical sciences. Conversely, many advances are being put in place in the systems biology area and learning from biological complexity can be a way of stimulating new chemistry. Biological systems display an incredibly large amount of amazing capabilities that can be a rich source of models for new areas of chemistry to design nonbiological systems. It is a big challenge for the chemistry of the 21st century, perhaps it is the challenge.

REFERENCES

1. J.W. Sadownik, Otto, S., Systems Chemistry, Encyclopedia of Astrobiology, Springer-Verlag Berlin Heidelberg, 2014.
2. M. Kindermann, I. Stahl, M. Reimold, W.M. Pankau, G. von Kiedrowski, *Angewandte Chemie*, **2005**, 44, 6750-5.
3. R.F. Ludlow, S. Otto, *Chemical Society reviews*, **2008**, 37, 101-8.
4. J.R. Nitschke, *Nature*, **2009**, 462, 736-8.
5. J.J. Peyralans, S. Otto, *Current opinion in chemical biology*, **2009**, 13, 705-13.
6. A. Ricardo, J.W. Szostak, *Scientific American*, **2009**, 301, 54-61.
7. K. Ruiz-Mirazo, C. Briones, A. de la Escosura, *Chemical reviews*, **2014**, 114, 285-366.
8. J. Stankiewicz, Henning Eckardt, L., *Angew. Chem. Int. Ed.*, **2006**, 45, 342–344.
9. V. Dhar, *Communications of the ACM*, **2013**, 56, 64–73.
10. A. Agrawal, Choudhary, A., *APL Materials* **2016**, 4.