**The** **Data Cycle**

Every decision-making process is based on a data cycle culminating in a decision being made. The cycle can be very short and based on a few data items, such as when we decide whether it is safe to cross the street. In such a simple case we first identify the problem (or the mission), collect visual data on cars passing by, estimate the width of the street and our walking speed, integrate this data, operate an algorithm based on our past experience (i.e., machine learning), analyse the results, make a decision, store feedback for future similar activities.

Obviously, most of the decisions taken by organizational bodies and by teams are much more complicated. However, the stages of the Data Cycle (DC) are nearly the same for each degree of complexity, in each sector, and for each discipline. Figure 1 portrays the Data Cycle.

Feedback

Visualization

Processing and analysis

Data mining

Data integration

Data collection and storing

Identifying pertinent data sources

Problem definition

Learning and decision-making

Figure 1: The Data Cycle

We will briefly describe each stage of the DC and list out potential tools that can support each stage:

1. **Problem definition**: An initial definition of the problem, or the mission, or the purpose, for which data is required. ***Potential tools:*** *formulation methods, quantitative models, qualitative approaches, mathematical tools, and the like.*
2. **Identifying pertinent data sources**: Understanding what data are pertinent, and where they can be located. ***Potential tools****: browsers, indices, search engines, international organizations, statistics bureaus, and the like.*
3. **Data collection and storing**: retrieval of data from various sources and store them in an accessible location. ***Potential tools***: *data transfer technology – communications, clouds, database management software, and the like*.
4. **Data integration**: This (very important) stage should allow the user to incorporate data from varied sources whose data definition and format were not initially compatible, nor are they synchronized. ***Potential tools***: *conversion programs, indices, metadata tools, and the like.*
5. **Data mining:** Selection of relevant data out of the Big Data. ***Potential tools***:*filters, data retrieval techniques, identification tools, AI tools, heuristics,* *and the like.*
6. **Processing and analysis:** The data that were selected earlier are now screened, processed, and analysed. ***Potential tools***:*algorithms, AI tools, machine learning, data processing programs, heuristics, and the like.*
7. **Visualization:** Presentation of the results to the decision maker(s). ***Potential tools***: dashboard software, graphical tools, reporting systems, interactive systems, voice, and the like.
8. **Learning and decision-making:** The final stage that is the purpose of the data cycle. The results are displayed to the decision makers and decisions are taken. ***Potential tools:*** *decision support tools, what-if software, visualization tools.*
9. **Feedback for further cycles:** This stage is not always necessary. However, very often, the need to make a certain decision is repetitive, so the customer (the decision maker) can affect the usefulness and the effectiveness of the cycle by forwarding comments and changes. ***Potential tools:*** *reporting systems, interactive reactions, fine tuning tools, DEVOPS tools, agile design tools, and the like.*

**Data Sciences is an academic discipline that covers the theories, paradigms and tools the serve all the components of the Data Cycle.**