

R&D Outsourcing

da_now

How data became the raw material of the 21st century

In this article, the ordinal number "fourth" occurs three times - the fourth transformation in data representation, the fourth paradigm in science and the "Fourth Industrial Revolution". Where exactly it came from is not clear, but it is natural that all three are united by data that have become a critical raw material of the 21st century. It is no coincidence that data has been called the oil of the Fourth Industrial Revolution. The journalist Leonid Chernyak in the material prepared for TAdviser tells about fundamental changes in the relation of mankind to data.

The difference between data and information

Even in the middle of the 2000s of the 21st century, it was difficult to imagine such a thing. Data as a component of computing was out of the question. Since the advent of computers, that is, since the mid-forties of the 20th century, attention has been focused first on hardware, and later on software. As for the data, they were considered as something obvious, self-evident. As a result, a strange one-sidedness of IT has developed, which distinguishes them from other industries. Production can be imagined as consisting of two things: a complex of technologies and raw materials, which, passing through the technological chain, turns into the final product. In IT, the technological process of converting input data into resulting data remains, as it were, "behind the scenes".

The soul-searching, recognition of the value of data and data processing that began around 2010 took only a few years. Ironically, data is now often given too much attention. Part of the computer and near the computer community is clearly suffering from a painful condition called datamania (data-mania). One of its manifestations is the abuse of the term "Big Data".

Another misunderstanding related to IT is that the concepts of "data" and "information" have long been considered as synonyms, which of course contributed to the statistical theory of information, which would be more accurately called the theory of data transmission. The name "information theory" was proposed by John von Neumann to Claude Shannon, who was extremely modest in his claims. In this theory, bits and bytes serve as a measure of transmitted information, although by definition they refer to data represented in a binary system.

It is significant that for many years the author, using the opportunities of a journalist, at the first opportunity asked his interlocutors the same question: "What is the difference between data and information?" However, I have never (!) received a meaningful answer. The fact that the so-called information technologies deal with data, and not with information at all, almost no one thought. Disregard for the nature of data has led to the fact that over the decades until the 2010s, only engineering methods have been developed to ensure the transfer, storage and processing of data. All that needed to be known about them came down to binary or decimal units for measuring the amount of data, formats and forms of organization (arrays, bytes, blocks and files).

But the situation, chipping around the data, has changed dramatically. Its reflection was the popular slogan "It's the data, stupid", reflecting the growing role of data in modern science, business and other areas of human activity. The shift in focus to data is a consequence of the greatest cultural

transformation.

There are four fundamental transitions, each of which is characterized by an increase in the availability of content:

The invention of paper and the transition from clay and wax tablets, parchment and birch bark to a practical and inexpensive medium.

Invention of the printing press and transition from manual copying of manuscripts to machine-printed editions.

Transition from material, most often paper media, to digital ones; separation of content from physics. Transforming content into data that can be processed and analyzed automatically.

The main feature of the latter is that in the 21st century data were abstracted from the carrier. The necessary tools were created to work with them, which opened up unlimited possibilities for extracting information from data.

From data to knowledge, the DIKW model

In fairness, it should be noted that in the academic environment, the importance of data as a source of knowledge and their place in the knowledge accumulation system began to be thought about earlier than in business - approximately from the end of the eighties of the XX century. Then the classic four-link DIKW model was formed, which includes data, information, knowledge and deep knowledge (data, information, knowledge, wisdom).

Data is obtained from the outside world as a result of human activity or from various sensors and other devices.

Information is created by analyzing the relationships and relationships between pieces of data as a result of answering the questions: Who? What? Where? How many? When? Why?

Knowledge is the most difficult concept to define; it is obtained as a result of the synthesis of the received information and the human mind.

Deep understanding (wisdom?) serves as the basis for acceptance

decisions

The DIKW model has been the basis for research in what is known as Knowledge Management (KM) for several decades. It is generally accepted that KM studies the processes of creating, preserving, distributing and applying the basic elements of intellectual capital necessary for the operation of an organization, allowing the transformation of intellectual assets into means to increase productivity and efficiency.

By means of KM, it was not possible to obtain tangible results and go beyond general reasoning by creating appropriate tools. KM has been and remains an area of interest for a very limited community of scientists. The failure of KM is due to several reasons - the fact that the desire to manage knowledge was ahead of its time, and the fact that the need to work with knowledge has not yet formed. But most importantly, level D from the DIKW model was out of sight of KM.

However, it does not follow from the failure of KM that there is no such problem as automating the extraction of knowledge from data. As they say, "a holy place is never empty", and in the second decade of the 21st century, the place of KM was taken by a new direction, which received the not very successful name of Data Science. The role and place of Data Science in the knowledge accumulation system are shown in the figure below. For millennia, people have observed the world around them using various tools and recorded knowledge in an accessible form. Today, the process is divided into the accumulation of data and the analysis of this data. A vivid example is modern astronomy or geophysics, where observation with data accumulation and subsequent analysis of these data are independent tasks.

Link to article: [How data became the raw material of the 21st century](#)