This study explores the complex, scientific, theoretical, and pragmatic characteristics of the new format of analytical media content using digital computer and Internet databases (data journalism). The author, for the first time, identifies the significant transformation of the communication pattern of data journalism on the Internet, and defines data journalism as a new format of media communication. The paper presents the analysis of Russian data journalism projects and for the first time describes a national model of data journalism. Data journalism fundamentally
determines new opportunities for the strategic development of media content and media communication; functions of journalism as an element of open public governance and socially integrated communication systems; and for the development of media theory media as well.

**Key words:** human technology; media; media communication; the Internet; fact; data; data journalism; national model; open governance.

В профессиональной коммуникации использование цифровых компьютерных и интернет-баз данных определяет параметры нового формата представления аналитического медиаконтента и появление нового формата медиакоммуникации — дата-журналистики. В статье представлены методологические подходы исследования феномена дата-журналистики в рамках формирования современной науки о медиа, комплексно исследуются научно-теоретические и прагматические предпосылки его появления, характеристики. Автор впервые выявляет существенные трансформации свойств, модели, функций журналистики данных в Интернете, определяет дата-журналистику как формат медиакоммуникации. Анализ российских практик дата-журналистики впервые фиксирует свойства и особенности национальной модели данного формата профессиональной медиакоммуникации.

**Ключевые слова:** гуманитарная технология; масс-медиа; медиакоммуникация; интернет; факт; данные; дата-журналистика; национальная модель; открытое управление.
Introduction

The mass media is undergoing significant changes as an industry and a scientific field. In this decade, theoreticians and specialists of mass media are searching for optimal economic and creative models of media, because journalism is still under technological pressure.

Experts (UNESCO, 2011) have identified a number of trends that could “grow” in global practices of media this decade, such as new technologies, new screens and portals for information delivery, increasing participation of consumers based on creativity and self-expression, and the rise of using of gaming technologies in discourse as well.

Digitalization, convergence, and the Internet have constructed the features of a new model of communication between media professionals and their audiences. In 2014 more than 97% of information in the world became digital. In 2018 the half of global population, media professionals, and audiences will be connected by the Internet and communicate virtually and digitally. Among the essential characteristics of digital communication and communication on the Internet, are databases (open, linked data) of digital information.

According to Gartner (Gartner, 2011), “big data” is the second greatest challenge (after virtuality) not only of media, but also of contemporary civilization. In other words, big data is a concept that is attracting and increasing attention in many different quarters, as well as praising the business, economic, political, and social potentials. Many fields that previously had been separated from computer science and the Internet are now using big data. The growing interest in data occurs in business and politics (Mayer-Schönberger, Cukier, 2013).
Thus, mass media based on new interactive and creative technologies would become more and more technology mediated.

In this context, contemporary journalism becomes a kind of human technology (Gorokhov, Shilina, 2014). The newest fields in journalism using computational techniques and tools illustrate this computational journalism (Cohen, Hamilton, Turner, 2011), as well as increasingly contributing to data journalism.

Contemporary journalism research means focusing on the integrative (or convergent) nature of media communication and on technological and technical aspects which define journalism. Contemporary journalism studies are a multidisciplinary field on the borderline between humanities, social sciences, technology, computer science, and now data science.

**Background**

The newest practices of journalism (since 2005) are connected by using large amounts of various types of data (big, open, linked data).

What makes data journalism different from the rest of journalism? According to “The Data Journalism Handbook”, the difference is the new possibilities that open up when a journalist combines the traditional “nose for news” and ability to tell a compelling story with the sheer scale and range of digital information, which is now available.

According to Appelgren and Nygren, data journalism is a form of storytelling based on journalistic research in the form of collection, compilation, processing, analysis, and presentation of large datasets. This form of journalism is often presented as interactive graphics or simple interfaces for the user on web-pages or on mobile publishing networks. It is published in combination with news articles and print, audio, or video content. (There are many other definitions
of data journalism, but here we use this one, describing the main characteristics of this phenomenon.)

Data-oriented journalists (Lorenz, Kayser-Bril, McGhee, 2011) even stated that mass media should leave the old business model, based on selling the attention of their audience to advertisers, alone. To illustrate this they mentioned that mass media should adapt to new technologies and opportunities and become data hubs, which build infrastructures that turn them into centers of reliable data.

Data journalism practices are more than discussed. Most publications on this theme are non-academic articles in the press. Data journalists have described their experience in interviews and articles (Nippard, 2011; Rogers, 2011; Cohen, 2014), and submitted the best ones to “The Data Journalism Handbook” (Gray et al., 2012). Scientists mentioned that data-journalists and researchers are just making the very first steps in this field. There are only a few attempts to describe practices of data journalism and to analyze some features of this phenomenon. There were some case studies published about data journalism in Sweden (Appelgren, Nygren, 2013) and Russia (Begtin, 2013; Shilina, 2013), EU (EJC, 2010) and about trends in using Open data (Sirkkunen, Aitamurto, Lehtonen, 2011).

According to Appelgren and Nygren, data journalism is still “one example of an area in the borderlands of journalism – hard to define, and consisting of a combination of several previously separate trades and skills.” In fact, there is no academic definition of data journalism. So, this field strongly needs complex research.

**Methodology and methods**

In journalism studies as multidisciplinary field theories, methodology and methods from interconnected areas are used,
and this multidisciplinary approach, both humanitarian and technological, is useful when it comes to data-journalism. In our study, systematic and function methodology were used.

From technological point of view, the emergence of the data journalism concept is associated with massively increasing opportunities for assembling data via the Internet and growing statistical and technological power for processing data.

The term “data” describes large volumes of digital computer information, but in the humanities, from the communicative point of view, the database definition was not given. In this study we use a definition based on the interpretation of Article 1280 of the Civil Code of the Russian Federation and a number of definitions of communication. The database is presented in the form of an objective set of materials (statistics, articles, etc.), systematized in such a way that the data can be found and processed digitally. The term “database” is used today to refer to various types of databases, including the important social digital data that can be stored in the public domain, used freely by the mass audience.

From communication and social (and economic, political) points of view, data journalism raises a vast number of questions that need to be addressed.

In the present study, we sought to investigate and define data journalism as a type of media communication. The key research questions guiding this study are: what is data journalism as a type of media communication? How do technologies determine data journalism? Is data journalism technology mediated in a new and specific way? This paper presents an attempt to analyze the main origin of data journalism, a data journalism communication model, its basic elements and functions, and a technological cycle.

Also we will discuss the current development of data journalism in Russia, and how data journalism projects are created in leading
Russian newspapers, magazines, news agencies, and Internet news organizations. What are their aims, values, tools and results? Our research interest is focused on communicative model features of these data projects as well. This part of research is based on analysis of the first 10 data journalism projects (beta versions), gathered in Russia in 2013, and interviews with the key experts of those teams devoted to the main aspects of these cases.

**Media communication and media system: contemporary paradigm shift**

The classic model of journalism was hierarchical, author-oriented, and of one-to-many formats. In this model, the journalist always was a unique subject of communication, and the audience was an object (usually of passive mode). Messages in this model have been created by the journalist, and the content has been strictly determined by channels (print, audio, visual media). The basic function of journalism in this classic model was to present various types of information.

Since 1960s all basic elements of contemporary journalistic communication have been significantly changed by the influence of digitalization, convergence, and the Internet.

After presenting some methodological points in the field of communication on the Internet, which is important for this research, we argue that the most important feature of communication on the Internet is web hypertext, which determines the basic properties of technical, technological, and social human-centered levels of interaction in this virtual and physical space (convergence both physically and digitally), makes it phygital. Hypertext provides subject-to-subject interactivity and also provides hyperlinks and multimedia synergy of two opposite processing formats of
communication such as logical and associative. Other characteristics of communication are openness, non-limited access to information, and enhanced speed of presenting and distribution of information.

In fact, due to those characteristics any content can be characterized as media content on the Internet, and the former is available for consumption and changes by the audiences.

New dominant professional capabilities are related to the role of the author in the media communication model: a journalist becomes not only a “news-maker,” and a “sense-maker,” but also an analyst implying increasing demands for competence, social role, and determines the need for creative solutions based on ICT. A professional journalist might be an “emotion-maker,” while at the same time considers interactivity of audience, common readers and users, and so called citizen journalists (as elements of media communication models with new functions).

And another change within the area of journalism is journalists collaborating with computer scientists, programmers, web designers, or the audience (Shilina, 2013).

Techniques and methods of work (and text writing) have been changed, and they are still permanently changing. For example, now the audiences are interactive, and consumers need the integration of news with the social context, so they seek information not only about the event, but also the information about social problems and trends, which journalists should offer.

Now journalists use computers and ICT in every field. Cohen et al. (Cohen, Hamilton, Turner, 2011) even describes journalism as computational study. Broadly defined, it can involve changing how stories are discovered, presented, aggregated, monetized, and archived. Cohen stresses that computation can advance journalism through innovations in topic detection, analysis, personalization, aggregation, visualization, and sense-making.
Thus, in the new digital space all the elements of media communication model have been changed profoundly. In the contemporary media communication model a journalist is collaborating with programmers and web-designers; audiences “work” as consumers; content becomes integrated. The basic functions of journalism are to involve audiences in communication and to construct participation. Media communication as the basic element of a media system changes the system in the same way.

**Data journalism: genesis, features & functions**

Data journalism as a kind of professional communication and creation of media content, which based on the use of large amounts of digital and Internet data, appeared in the first decade of this century.

Foreign and Russian practices of data journalism are in the process of formation, but the examples of using various types of data to create a more effective journalistic text are dated back by some scientists to the beginning of the 21st century. In particular, Simon Rogers describes the first example of data journalism as the article in the British newspaper “The Guardian” (1821), where a list of schools of Manchester with its number of categorized pupils were analyzed and published. Rogers emphasizes that “data journalism has been around as long as there’s been data” (Rogers, 2011). Rogers mentions that data journalism may be trendy, but it’s not new.

The protoforms of today’s data journalism phenomenon is the latest journalistic practice of using computer equipment, digital and Internet technologies for analysis, and the presentation of journalist information in various media sources.
In the present study we sought to mark different types of collecting, using, and presenting facts influencing journalists of “computer” times. In the 1950s electronic computer information, the concept of usage, and the presentation of information, primarily in the U.S., has changed due to digitalization. The distinctive feature of the computer-assisted reporting (CAR) is the systematic use of computer and digital information, facts, statistics, studies, and the analysis of databases to enhance the quality of content and strengthen the evidence base, especially in the investigative journalism.

At this time another direction of journalism developed based on literary fiction and images that shaped differences in methodology of professional creativity. So-called new journalism (or new conceptualism) was based on the use of literary techniques (plot, dialogue, fiction, etc.).

In the early 1970s, alleged precision journalists have applied sociological research methods in order to work with information.

According to Meyer, the practice of precision journalism based on large data sets adapts to the scientific method, scientific objectivity, and scientific ideals of the process of mass communication (Meyer, 2002).

The main differences and disadvantages of such approaches were outlined by American journalists at once. According to Dennis and Rivers (1974), “at the moment we have the means like the census, public opinion polls, the detailed results of the election that allow us to achieve exactly that tell us a lot about people. (...) The problem is that when you turn the charts in the article, it makes people yawn. (...) To read the same materials written by “new journalists” is like to read fairy tales”.

Thus, assayed and interpreted facts were seen as opposing. But now on the Internet, due to hypertext (and multimedia), all these
features are integrated and used in journalism and data journalism as well.

For the first time the term “data journalism” was defined in 2006 by one of the first American data journalists Adrian Holovaty as outlined, structured, machine-readable data, which is used with traditional media texts (Holovaty, 2006).

At present, this definition of data journalism is not quite correct because data journalists use not only computer and Internet data and communication, but also many new different types of integrated content, which becomes increasingly diverse.

The specificity of data communication is the author’s use of digital data not as a tool for creating text, images, or infographic series for it, but as a major source for the formation of topics as a key condition for choosing the genre. Implementation of this online format means the use of united multimedia hypertext as a new type and method of media communication in every media model: the data and the author, the author and the content, the data and the audience, the audience and the content, the audience and the author, etc.

Patterns of data communication and data journalism, noted above, are different from the classical analytical journalistic texts. According to Tertychny (Tertychny, 2010), the types of classical analytical journalism are: report, correspondence, interview, survey, discussion, comment, reply, profile, sociological summary rating, monitoring, review, article, version, experiment, epistle, essay, and recommendation.

According to Rogers (Rogers, 2011), journalists have to analyze data as quickly as they can, and do it correctly; with the right tools of analyses anyone can do it, but only professionals can produce a good design and visualization. Rogers mentioned that the bigger task is to conceptualize the data like a journalist rather than an analyst.
So, answering main professional questions (What’s interesting about these numbers? What’s new?) is more important than programming. In addition data journalism (as 80% perspiration, 10% great idea, 10% output, by Rogers) is about telling the story in the best way possible.

As the leading data journalist of “The Financial Times”, John Burn-Murdouch, mentioned in the interview (28/11/2014) that “data is the foundation, not the facade, and the reader should not have to know it”.

According to Cohen (Cohen, 2014), “the first priority is reporting rather than data”.

Database and programming tools (SQL, R, various command, etc.) are open and freely used, so the main aims of a journalist working with data are methods of analysis, interpretation and presentation of content.

According to Burn-Murdoch, the goals and guidelines of data journalism are based on serving current and prospective subscribers and meeting their expectations.

The typical example of data journalism is the data blog of “The Guardian” (Guardian Data blog, launched in 2009 by Simon Rogers). Today at least three to five journalists write to the blog several times a day, which illustrates the increasing demand of such information.

According to our analyses of the texts of “The Guardian” (2010-2014) and “The Financial Times” (2013-2014), the British model of data journalism is a form of storytelling based on journalistic research. Information is often presented as text, diagrams, graphics, or simple interfaces for the user on web-pages or in mobile publishing channels to access. It is in the form of news articles with primarily printed content, and photo, audio, video, or multimedia content is rather rare. Texts are non-interactive as usual. And the main author
of the story is the journalist. Practices of BBC are similar, but they use greater varying types of content in data journalism materials. BBC develops this field of journalism, with the special Data News Lab. In British data journalism model, the main function is rather classical: to inform.

The practices of most popular American news organizations, for example “The New York Times” (2009-2014), demonstrate that data journalism is becoming increasingly popular as well. According to our analyses of data journalism texts of NYT, the model of American data journalism presents a form of storytelling based on journalistic research presented as text or a project. Texts are published as combination of verbal and visual content and are usually interactive. The main function in an American data journalism model is not only to inform, but to involve audiences in the communication.

Journalist Sarah Slobin (“The Wall Street Journal”) gives simple, understandable information to the reader, which is personally oriented and an effective example of data journalism, realized by “The Las Vegas Sun”. Journalists have published several articles about health care, prepared about the analysis of 2.9 million hospital records, invoices from public sources, which are allowed to disclose 3,600 injuries, infections, and surgical errors that could have been prevented. The interactive graph, which identifies the problems in all hospitals, gave the readers an opportunity to view information relatable to each of them, so they can choose the best hospital for themselves.

What are the main features of data communication and content? The main features are formed at all stages of the process. Journalists, web-designers, and programmers as subjects in this communication model find the threads of information, analyze it using software (MySQL, Python, R, etc.), visualize and present original data in the
story (or as an equal part of the plot). The authors have to be good analysts, writers, and users of technical tools for data visualization (e. g. Google Docs, IBM ManyEyes, Wordle, Excel, etc.). Thus, data journalist should possess various types of thinking: both logical and associative.

According to Cohen et al. (Cohen, Hamilton, Turner, 2011), in order for journalists and computationalists (programmers and web-designers as well) to work together to create a new generation of reporting methods, each needs an understanding of how the other views all basic parameters of the working process and primarily “data”. The key competencies in this field are: combining information from varied digital sources, making an information extraction, document exploration and redundancy, audio and video indexing, extracting data from forms and reports.

According to Burn-Murdoch (2014), the main condition and aim of the work is a collaborative effect from the involvement of reporters, correspondents, editors, as well as interactive teams.

Data journalism as a profession and as a field of research becomes increasingly integrated.

Data, both on the computer and Internet, is a source of information for journalists, an original tool for creating a new type of content and communication, and it’s a basis for all journalistic classical genres in all types of digital media.

Data journalism, due to the openness and accessibility of data on the Internet which may be shared between both a journalist and any user, has the ability to create and publish texts on the grounds of professional online media, as well as increases the media literacy of readers and the quality of open media content. It’s free of charge as well. According to Cohen, engaged citizens need computer scientists more than journalists and journalism of the public interest.
Additionally, the potential damage from the incomplete disclosure of information, such as the error in the analysis of large data sets, especially in the forward-looking research, is potentially boundless, which increases the responsibility of the journalistic analyst. Another problem, despite the obvious potential of vast and open data, is a growing amount of private personal data and its disclosure which may violate the rights of readers. There are also many technical problems; in particular most of the data generated now, is unstructured and informally organized so that it cannot support the validity of the analysis.

Data journalism in Russia: analyzing first cases

In Russia the media system has changed profoundly over recent decades through market influence, and the consequences have been similar to those in foreign media including “a restructuring of the media system, quantitative growth, increased diversity in media channels and media content, regionalization of the media markets, and the introduction of ICT into the media infrastructure and into the media system itself” (Vartanova, 2013). So, the trends of Russian media development could be similar to those in the world industry, particularly in media trends and data journalism.

The media sector of Russian Internet (Runet) is one of the most rapidly developing fields of professional information and communication. But there are digital divides, which include the lack of broadband access all over the Russian territories, the low technical culture of society, and rather high prices for Internet services.

According to Lukina (Lukina, 2013), “the development of the sector is also limited because of the lack of legal base in the field, the lack of respect to copyrighting, and poor personal data
protection mechanisms”, and Lukina concluded, “however, despite the objective difficulties the dynamics of Internet usage in Russia is growing positively and contributes to consumers’ demand for online information. This fact encourages renovation processes in Russian digital media”.

In Russia examples of data journalism are rather rare. However, more data developments are presented by civil activists. For example, the data project “Journalism and Public Investigations: new methods, technologies, practices” of the Press Development Institute – Siberia (Novosibirsk), Information Culture (Moscow), and the Branch of the Union of Journalists of Russia in Perm have been working in Siberia since 2014. A data journalism school for youth was opened in 2014 by the Federal Educational Project “Information Stream”. First courses of large and open data, data driven journalism, and PR were provided by the Higher School of Economics (2014, Integrated Communication Dpt.).

One of the first examples of Russian data projects is “Accidents with pedestrians in Novosibirsk”, which began in 2011 (this project won the first prize of Data Journalism Awards in 2012, organized by the Global Editors Network). Journalists and citizens of Novosibirsk collected and analyzed the data of the local traffic police and resources that provide information about the most dangerous streets for pedestrians in Novosibirsk for a year.

This part of our study was designed for the analyses of ten data journalism projects developed in Russia with framework for the project “Great Russia – big data” (RIA “Novosti”¹). These practices are representative of the national model of data journalism and relevant to the global mass media models as well. The chairman of the jury, Antoine Laurent (France), mentioned in an interview with

¹ Now international news agency “Russia Today”. URL: http://www.mn.ru/society/20131120/363029873.html
the author of this study, “Russian projects have demonstrated high level, and international standards” (Laurent, 2013).

In data project “Pulse” (“Big City”) the journalist team presented an interactive map of Moscow, which allows audience to explore, where residents spend their free time. The project monitors the activity of Muscovites on the basis of users’ “check-in” on the Russian social network “VKontakte”. On the map so called power spots, broken down into administrative city districts, were presented. The team has imposed this data on such factors as gender, age, presence of higher education, marital status, and others. As a result, the audience can discover “power spots” existing for different groups of citizens. For example, single men and single women choose different places to rest in the city center. With the information from this project, everyone can test various hypotheses about the life of real people in Moscow.

According to Daniar Shekebaev, the online producer of “Big City”, this project helps Muscovites identify themselves at micro level, for example, as Muscovites of Khamovniki.

The essence of the project “Where in Russia life is good” (“Argumenty i Facty”) is that the user is prompted to answer a series of questions about the criteria for the optimal place to life. After that, there were five options offered to possibly move. The thickness of the lines on the map shows how option meets the criteria of the user. This allows one to see the overall ranking of cities according to the selected criteria. Every user can compare specific datasets of criteria (acceptable distances to a new location, population, rent prices, wages, unemployment, crime, security, kindergartens, hospitals, cultural level, traffic jams and so on) of each city. If a person wants to move to another city, the test will give him or her a rough idea of conditions of the proposed area of relocation.
According to Sergey Kondratyev, head of the design, infographics, and illustration department for the development of new digital media trends of “Argumenty i Factly”, their newspaper has more than 35 regional sites, and their readers were interested in comparing the conditions of their region. Journalists decided to help them to change their lifestyle.

The data project “Emotional twitter background” (“Mercator”) makes an emotional “cardiogram” of Russian language on Twitter. After analyzing several dozens of profiles in the Russian language on Twitter, the team calculates the difference between the number of positive and negative words in the posts. In fact, Internet users can watch the mood changes of the participants of Twitter.

According to Andrey Skvortsov, director and co-owner of “Mercator”, the authors can measure the “positive” and “negative”, “confidence” and “uncertainty”, and political preferences in the direct speech on Twitter in order to find external impact on the audiences.

The project, named “From Moscow. Russia is big, where would you move?” (“Moscow News”), helps users find a location that suits them more than their recent one. For each locality, the team analyzed various indicators and data from certain sources (population density, unemployment, average temperature, and average salary). Industrial cities were selected due to the ratings of the “Urbanik” Institute, and villages were chosen from the ratings of the most amazing villages in “Forbes”. This project will help thinking about the differences of Russian cities, and will help to learn a lot about them.

According to Philip Katz, editor of the infographic studio of RIA called “Novosti”, all the migratory routes in the country clearly go from the smaller villages to the larger ones and towns. According to data (2011), more than half of all journeys in Russia are moving to Moscow, and such a migration path does not allow people to compare settlements. So, the team decided to help this issue and
suddenly found that the provincial town or village will be even better than Moscow to settle.

The team of “the debt load of the population in Russia” (Economic Information Agency “Prime”) presented an interactive map of Russia, which helps to analyze the situation of Russians’ debt over the last three years. The team wanted to show credit figures and levels for the average resident of various Russian regions. The project was based on data of individuals in four groups – consumer, mortgage, car loans, and credit card transactions. This allows citizens to think about the economic risks, and what is dangerous in them.

According to Alexei Timatkov, editor of the design center of RIA “Novosti”, they are alarmed by the abundance of Russian media coverage devoted to problems associated with the loans, due to economic ignorance driving the citizens into a trap. They wanted to explore this theme in order to help people.

Figure 1

Data project “The debt load of the population in Russia”, main visual
The next project, “2013-2030: how the oil will affect the salary” (RosBusinessConsulting, RBC), presents the graph of wages affected by Russian oil prices. The team compares different salaries (2008-2013) with prices of oil at the same period. Users can track how their salary will depend on the fluctuations in the price of Russian oil, and can calculate what their salary will be salary in 2030. The journalists reveal which professions will suffer from inflation as well. So, this project exposes interesting professions that are independent from “oil effects”.

According to Edik Tsvettsih, head of the design and usability department of RBC, this popular topic was chosen when the Ministry of Economic Development published a forecast for the Russian economic development until 2030, and the team decided to explore how people in Russia will spend money and feel in 2030, and to help them defend themselves from these risks.

Figure 2

Data project “2013-2030: how the oil will affect the salary”, main visual

“Rossiyskaya Gazeta” wanted to compare statistical figures with the people of the project “Biographies of Russia”. The program tells the user if he or she lives in better or worse conditions than his
or her peers in a particular region. The program also issues a small prediction based on statistical data. For example, the program shows what the probability that the user will double his or her income in the next three years is, and what the probability that he will live up to 68 years or become a parent of three children is. This project is very interesting and is even exciting for users.

According to Yevgeny Volk, a web analyst (web site rg.ru), the analysts want to create a reliable picture of the Russian population using all these statistical supplements along with users’ input. Volk mentioned, that the main problem of all projects is that the team does not know who their audience is. And the team has created a panel, which gives a clear understanding of where the users live, how much money they earn, whether they have children, what are their educations. Such interactive projects would help to create a better portrait of audience.

*Figure 3*

Data project “Biographies of Russia”, main visual
In the project “Milk for harm” (“RIA Ranking”), the team gives an analysis of the environmental situation in the Russian regions compared to economic parameters. The project helps people to understand if they live in the region with a poor economic environment, and live with or without any compensation. Compensation index (“milk index”) consists of the local authority’s spending for the community, the number of doctors per capita, and the average wage in the region. Looking at the map, many users can recognize that their region is not as prosperous as they thought. After understanding the source of the issue, residents can raise this problem to state or federal authorities.

According to Andrew Manko, head of the financial institutions analysis department (RIA “Novosti”), the project team would like to develop a more detailed map of the regional, municipal, and urban district level. They also had to examine more than a million cells of information. And the problem of data quality is not completely solved. For example, six regions have no data at all, such as Chechnya and Ingushetia. Statistics on these republics began to appear only recently.

On the map of the project called “Decent state. Kindergartens” (“Tatar-Inform”), the user can select a district, such as Tatarstan, and can see how authorities solve problems with state and private kindergartens. This project also can compare the districts by the number of kindergartens. So, people can compare the quantity and quality of kindergartens and local authorities.

According to Maxim Tolchinsky, the executive director of the agency “Tatar-Inform”, the team wants to demonstrate how data projects can truly help people. In the future they will develop a project demonstrating this, and they will uncover all the other social problems in the region.
The project “Changing Russia” ("Esquire"), based on English Wikipedia articles, is devoted to the perception of Russia in the world. On the map of Russia, the audience can see what information about the country and abroad the Russian people are most interested according to Wikipedia foreign users.

Figure 4

Data project “Changing Russia”, main visual
According to Dmitry Golubovsky, the chief editor of “Esquire”, the team has created a working prototype of the services that will be available soon on their website. They have long been interested in Wikipedia as, which organizes human communication and knowledge.
### First Russian data journalism projects. Communication models

(*beta, n=10*)

<table>
<thead>
<tr>
<th>Project</th>
<th>Data</th>
<th>Subject of Communication</th>
<th>Object of Communication</th>
<th>Message: Text, Visual, Multimedia</th>
<th>Type of Interactivity</th>
<th>Model of Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Pulse” (“<em>Big City</em>”)</td>
<td>126,000 “check-in” in 62,000 profiles of “Vkontakte”</td>
<td>Journalist, web-designer, programmer, on line producer. Not obvious interactive</td>
<td>Not obvious interactive</td>
<td>Verbal text, interactive mapping</td>
<td>Interactive programming. Passive subject and object</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>“Where in Russia life is good” (“<em>Argumenty i Factly</em>”)</td>
<td>Rating of Russian regions” of “Urbanik” Institute (St. Petersburg) the datasets of Rosstat</td>
<td>Journalist, web-designer, programmer. Interactive.</td>
<td>Interactive</td>
<td>Verbal text, interactive mapping of different types</td>
<td>Interactive programming. active subject and object</td>
<td>Non-hierarchical</td>
</tr>
<tr>
<td>“Emotional twitter background” (“<em>Mercator</em>”)</td>
<td>The archives of the project “Open Data Hub”, twitter data, library of positive and negative words</td>
<td>Journalist, web-designer, programmer. Not obvious interactive</td>
<td>Not obvious interactive</td>
<td>Interactive diagram</td>
<td>Interactive programming. Passive subject and object</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>Project</td>
<td>Data</td>
<td>Subject of Communication</td>
<td>Object of Communication</td>
<td>Message: Text, Visual, Multimedia</td>
<td>Type of Interactivity</td>
<td>Model of Communication</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>----------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>“From Moscow. Russia is big, where would you move?” (&quot;Moscow News&quot;)</td>
<td>14 cities, 13 industrial cities, 16 provincial towns, 14 villages and seven places for solitary life</td>
<td>Journalist, web-designer, programmer. Interactive.</td>
<td>Not obvious interactive</td>
<td>Verbal text, mapping</td>
<td>Interactive programming, active subject</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>“The debt load of the population in Russia” (&quot;Prime&quot;)</td>
<td>National Bureau of Credit Histories and Rosstat.</td>
<td>Journalist, web-designer, programmer. Interactive.</td>
<td>Interactive</td>
<td>Mapping</td>
<td>Interactive programming, active subject and object</td>
<td>Non-hierarchical</td>
</tr>
<tr>
<td>“2013–2030: how the oil will affect the salary” (RBC)</td>
<td>The data on wages, the stock information on the prices of Urals.</td>
<td>Journalist, web-designer, programmer. Interactive.</td>
<td>Interactive</td>
<td>Diagram, images</td>
<td>Interactive programming, active subject and object</td>
<td>Non-hierarchical</td>
</tr>
<tr>
<td>Project Description</td>
<td>Data Source Details</td>
<td>Role</td>
<td>Interactive Content</td>
<td>Content Format</td>
<td>Interaction Type</td>
<td>Hierarchy</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------</td>
<td>------</td>
<td>---------------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>&quot;Biographies of Russia&quot; (&quot;Rossiyskaya Gazeta&quot;)</td>
<td>Rosstat, Russian census of 2002 and 2010, open data of the Ministry of Health</td>
<td>Journalist, web-designer, programmer</td>
<td>Interactive</td>
<td>Verbal text, images</td>
<td>Interactive programming, active subject and object</td>
<td>Non-hierarchical</td>
</tr>
<tr>
<td>&quot;Milk for harm&quot; (&quot;RIA Ranking&quot;)</td>
<td>Rosstat</td>
<td>Journalist, web-designer, programmer</td>
<td>Interactive</td>
<td>Interactive mapping</td>
<td>Interactive programming, active subject and object</td>
<td>Hierarchical</td>
</tr>
<tr>
<td>&quot;Decent state. Kindergartens&quot; (&quot;Tatar-Inform&quot;)</td>
<td>Data on kindergartens, data of &quot;people’s control&quot;</td>
<td>Journalist, web-designer, programmer</td>
<td>Not obvious interactive</td>
<td>Interactive mapping</td>
<td>Interactive programming, active subject</td>
<td>Non-hierarchical</td>
</tr>
<tr>
<td>&quot;Changing Russia&quot; (&quot;Esquire&quot;)</td>
<td>English-language Wikipedia articles, data editing, and site dbpedia.org</td>
<td>Journalist, web-designer, programmer</td>
<td>Not obvious interactive</td>
<td>Interactive mapping</td>
<td>Interactive programming, active subject</td>
<td>Hierarchical</td>
</tr>
</tbody>
</table>
**Table 2**

First Russian data journalism projects. Functions and type  
*(beta, n=10)*

<table>
<thead>
<tr>
<th>Project</th>
<th>Aim</th>
<th>Function</th>
<th>Type of Involving</th>
<th>Values</th>
<th>Type of project</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Pulse” (“Big City”)</td>
<td>To present and to test free-time life of Muscovites</td>
<td>To inform</td>
<td>Passive, technology mediated</td>
<td>Social Identity of Muscovites</td>
<td>Social</td>
</tr>
<tr>
<td>“Where in Russia life is good” (&quot;Argumenty i Facyt&quot;)</td>
<td>To show real conditions of proposed relocation in Russia</td>
<td>To involve</td>
<td>Active, technology mediated</td>
<td>Complicity, Social Identity of Russian regions</td>
<td>Social</td>
</tr>
<tr>
<td>“Emotional twitter background” (“Mercator”)</td>
<td>To watch the mood of the main characters of Twitter</td>
<td>To inform</td>
<td>Passive, technology mediated</td>
<td>Individual Identity</td>
<td>Social</td>
</tr>
<tr>
<td>“From Moscow, Russia is big, where would you move?” (“Moscow News”)</td>
<td>To help users to find a location that suits them more</td>
<td>To inform</td>
<td>Active, technology mediated</td>
<td>Social Identity of Russian regions and Russians</td>
<td>Social</td>
</tr>
<tr>
<td>“The debt load of the population in Russia” (“Prime”)</td>
<td>To analyze the situation with the debt load of Russians and to choose correct model of behavior</td>
<td>To change type of economical thinking and acting</td>
<td>Active, technology mediated</td>
<td>Economic education</td>
<td>Social</td>
</tr>
<tr>
<td>Title</td>
<td>Main Objective</td>
<td>Method</td>
<td>Domain</td>
<td>Social Purpose</td>
<td></td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------</td>
<td></td>
</tr>
<tr>
<td>“2013–2030: how the oil will affect the salary” (<em>RBC</em>)</td>
<td>To analyze the economic situation and to choose correct model of behavior</td>
<td>To change type of economical thinking and acting</td>
<td>Active, technology mediated</td>
<td>Economic education and professional identity</td>
<td></td>
</tr>
<tr>
<td>“Biographies of Russia” (&quot;Rossijskaya Gazeta&quot;)</td>
<td>To analyze real lifestyle of Russians</td>
<td>To involve into journalists project</td>
<td>Active, technology mediated</td>
<td>Social Identity of Russians</td>
<td></td>
</tr>
<tr>
<td>Milk for harm” (&quot;RIA Ranking”)</td>
<td>To analyze the environmental situation in the Russian regions in comparison with the economic parameters</td>
<td>To involve into social life</td>
<td>Active, technology mediated</td>
<td>Social help</td>
<td></td>
</tr>
<tr>
<td>“Decent state. Kindergartens” (&quot;Tatar-Inform&quot;)</td>
<td>To analyze social problem and help people</td>
<td>To inform</td>
<td>Active, technology mediated</td>
<td>Social help</td>
<td></td>
</tr>
<tr>
<td>“Changing Russia” (&quot;Esquire&quot;)</td>
<td>To show the parameters of Interest of foreigners to Russia</td>
<td>To inform</td>
<td>Active, technology mediated</td>
<td>National identity of Russia and Russians in the world</td>
<td></td>
</tr>
</tbody>
</table>

Data journalism, social
First ten Russian data-journalism projects (beta versions) demonstrate new formats of media content, which are not journalistic (data journalism interactive projects, not texts, 100%; 2 mixed projects – data journalism and social project both) with high level of interactivity (100%) and social integrity (100%).

All the projects are completely technologically mediated. The majority of creative ideas of the projects (80%) depend on original technological solutions even more than on journalistic ideas. Moreover, the majority of projects (90%) may exist only due to interactive programming.

Journalists had chosen economic (4 projects), social (4), and political (1) themes for their data projects. These projects are of city (1), regional (1), interregional (4), national (3) and global (1) levels.

Models of communication are classical and hierarchical (50%) and non-hierarchical (50%). The authors are from diverse creative teams in all the projects (journalist, web-designers, programmers, and producers). But the interactivity of professional authors during the project is not clearly stated (only 70%).

The audience has to be interactive in many ways, but the audience’s interactivity is not obvious (50%). In fact, in these first Russian projects about data journalism has a level of interactivity with audience higher than that of professionals.

The type of content is substantially visual. Visualization models are rather classical (mapping, 70%; diagram, 20%) and not totally interactive.

In spite of a high level of social relevance, professional novelty, and creative interactive formats of communication, the aims designated by professional media teams were rather traditional: to present (3) and to analyze the information (6), as well as to help people (6). Also the functions of data-journalism in these projects
are meant to inform (5), to be involved in social life (3), and to change personal and social life for the better(2).

These are the main characteristics of the Russian data journalism model. As a result, all the cases demonstrate new types and models for media communication, which is rather social than journalistic.

**Conclusion and discussion**

The study of data journalism as a phenomenon, a process, and a type of media communication allow us to define data journalism:

- as a integrated set of specific skills of creative teams consisting of journalists, web-designers, programmers, and producers to search, analyze, and present information from digital and Internet data sources allowing the construction of unique analytical content for effective interaction with the audience;
- as the method of creation, transmission, consumption of large and open datasets, which can be used as meta-analysis for various, new, journalistic genres;
- as a type of integrated digital media text, based on original info and new categories of data.

All of the analysed projects of data journalism are completely technologically mediated. Data journalism demonstrates creative possibilities of methodologically facilitated professional communication. But it also strengthens human, rather than technocratic, trends in media communication, because the creation of original content based on open information requires specific interpretation and analysis.

The communication model of data journalism, presented in the paper, presented transformations of all classic elements and
links in it. As a result, data-journalism is a new type of media communication. Additionally, this would change many features and functions of the media system.

The first ten Russian data journalism projects demonstrate new features of national journalism. New formats of media content from these data journalism projects are more social and “citizen”-oriented than journalistically oriented, and all of them demonstrate a high level of interactivity and social integrity. The communication model is constructed by a diverse team (journalists, programmers, web-designers, and producers). The levels of interactivity of audience in data journalism projects are higher than that of professionals. The majority of these data journalism projects may exist only due to interactive programming. Also, data journalism content is technologically determined. But functions of data journalism developments are still rather classical, such as the main function of data journalists being to inform.

From the theoretical point of view, the paper for the first time supplies a new format of the synthesis of fundamentally different approaches for using facts as a main category of journalism as well as the possibility for synergy of all journalistic genres.

The emergence of data as a new type of factual information and media communication displays that journalism has a new level of functioning in society. This category of media content on the Internet makes the development of a new format of construction not only of opinions, but also of knowledge possible. Data journalism is an effective source of media literacy.

This paper raises several questions for future research: Is data the most useful tool for contemporary journalism? Does the audience select, analyze, and present data with or without a journalist? What are the types of integration (or convergence) of data content? What would be the new types of content? How should we use data in media
economics, media business processes, and internal communication in media organizations?

For clarification, let us note that data journalism corresponds with the government programs of open governance around the world and in Russia, which contribute to the development of these types of databases, and consequently contribute to data journalism as a new and significant leader of social development.

References


