

## **Between Media Datalization and Statistical Literacy: China's logic**

Zhongliang Zhang<sup>1</sup>, Weizi Zhang<sup>2, 3</sup>

<sup>1</sup> National Bureau of Statistics of China, Beijing, CHINA

<sup>2</sup> Mount Holyoke College, MA, USA

<sup>3</sup> Corresponding author: Weizi Zhang, e-mail: zhang24w@mtholyoke.edu

### **Abstract**

In recent years, the media datalization has gradually become a popular trend, with the expansion of official statistics fields and the change of media ecology, More and more statistics data entered public life with the help of media, affecting or even to some extent dominating the behavior choices and mood swings of the public, providing a new path to enhance public statistical literacy. This paper, based on the analysis of the driving forces and operation logic of media datalization, proposes two modes of statistical literacy measurements, and examines interactions around media datalization and statistical literacy. The authors argue that without statistical literacy as a prerequisite, the room for media datalization to play a role will be greatly narrowed down; and without the existence of the media datalization, the efforts to improve statistical literacy will be tremendously limited.

Key Words: statistics data, media dissemination, media datalization, statistical literacy

### **1. Introduction**

In the past ten years, with the expansion of statistics fields and the change of media ecology, more and more statistics data become the content of media dissemination and entered the daily lives of the general public. Overwhelming statistics data, as well as date-based media content, has become a great landscape of media dissemination nowadays: about citizen's living, there are CPI, House Price Index, Precipitation Probability, Security Index; about business, there are Consumer Goods Total Retail Sales, PPI, PMI, Investment Climate Index, TV Ratings; about national economy, there are GDP, Total Imports and Exports, Residents' Income, Unemployment Rate, Stock Price Index, Fiscal Revenue; about social development, there are Tax Misery Index, Residents' Happiness Index, Life Expectancy, ... At the same time, data analysis, interpretation, comments, is constantly being produced, presented in media, and affects around even decides people's behavior choice and mood swings.

The change above is defined as “media datalization” in this article. There are few local studies on media datalization in China, which largely focused on how the media use statistics data, such as how to excavate the value of the data, how to prevent the misuse of data, as well as the knowledge and skills of media practitioners and so on.

Understandably, to a large extent, media datalization depends on the statistical literacy of stakeholders. On the one hand, the higher the statistical literacy of the media jobholders is, the better the presentation of media datalization, and the stronger the attractiveness for the audiences; On the other hand, the higher the statistical literacy of the audiences is, the higher the likelihood for them to accept and understand the datalization content, and the greater the power to motivate the media to deepen

datalization. Thus, it can be inferred that there is a close link between media datalization and statistical literacy.

Taking into account the fact that, so far, there are few studies on media datalization and interaction of media datalization and statistical literacy, therefore, based on analysis of driving forces and run logic of media datalization, this article will examine and discuss the meaning and measurements of statistical literacy, and interactions of the media datalization and statistical literacy.

## **2. Driving forces of media datalization**

Nowadays, media datalization has become a major feature of the media dissemination. In particular, the financial media shows special preference to the datalization content.

In china, media datalization has two main directions: First, the statistical data, especially official statistics data become the focus of media reports. In fact, the time to release vital statistics, such as the monthly CPI, PMI, House Price Index, is when the time media datalization achieves the ultimate. Almost all of the media will arrange a considerable space to report and review data or do deep excavation to data, and present to the audience. Second, the Precision Journalism becomes the new darling of the media. The so-called Precision Journalism refers to depth reports, which use surveys, experiments and content analysis and other methods to collect data, and on this basis, quantitatively describe, analyze and parse the data.

At the moment, almost all the influential media, regardless of the mass media, or niche media, have a preference for Precision Journalism. Some set up a special investigation agency, while some regularly commissioned professional research institutions to organize investigations around various topics.

The cause of the appearance and development of the media digitization is the drive of the following factors:

First, the development of market economy, the awakening of civic consciousness and the rise of consumer culture, are the “hotbeds” of the appearance of media datalization. The Maximum reasonableness of the existence of media datalization is the needs of the audience.

Second, the change of media ecology, especially the digitization of the media, is the most important driving force of the media datalization. On the one hand, new media ecology provides a great space and great convenience for information dissemination; on the other hand, because of its implication value with information as its core, data can be transformed into the niche of media competition.

Third, the way of public thinking and living changes. At the moment, the public’s attention and participation of national economic life is becoming higher and higher. Concern about the changes in the macro and micro economy data has become an important part of everyday life of a significant number of citizens.

Fourth, the expansion of the serving scope of the official statistical agency. In the past ten years, the official statistical agency has paid a great deal of efforts to render services to the public. And, the media, as a connection link between official statistical agency and the public, has become an important support to the official statistical agency for serving the public.

Fifth, the development of data production formats. In recent years, China is having

a great boom in market and public opinion research industry. More and more universities and government research institutions collect data according to statistical methods. The amount of datalization content which can be used by the media now is definitely far beyond what it was in the past.

Of course, in addition to the above changes, there is an important change, citizens' statistical literacy is rising continuously. In 2001, the new "Full-time Compulsory Education Math Curriculum Standards" promulgated, as a result, "Statistics and Probability" become an independent subject throughout the mathematics curriculum in primary and secondary schools. Because of this, in 2001, the statistical literacy of the public stepped into a rising tunnel in China.

### **3. Statistical literacy from media datalization perspective**

Talking about the subject, there are mainly two kinds of people who promote media datalization.

Firstly, the so-called disseminators, that is, media practitioners, such as media reporters, editors and managers, their choices and actions depend primarily on the following four aspects: First, if they have interest in the dissemination of datalization content; second, if they can get close to and make use of reliable data resources; third, if they have a good grasp of audience interest and the ability to understand the datalization content; fourth, if they have necessary statistical literacy.

Secondly, the audience, such as newspaper readers, network surfers, their choices and understandings of datalization content depend mainly on the following four aspects: First, if they have interest in the datalization content; second, if they could contact media which hosts the datalization content; third, if they trust in the data producers and the media; Fourth, if they have necessary statistical literacy to understand the datalization content.

In fact, in terms of media datalization, statistical literacy is the most basic attributes for either disseminators or audiences. Just think, if they do not have any basic knowledge of statistics, how are they able to take any interest and trust in the datalization content? And how are they able to have access to data resources or datalization content?

The problem is that different roles require different level of statistical literacy. From the perspective of the media datalization, we believe, "doing statistics" and "using statistics" must be separated in the discussion of statistical literacy. If to say, for the disseminators who are "doing statistics", the statistical literacy is the ability to "speak with data" in a certain sense; then, for the audiences who are "using statistics", the statistical literacy is merely the ability to understand "the words saying by data".

Taking this point into account, the measurement of statistical literacy of the disseminators and the audiences, can be discussed in regard of the datalization content.

For the disseminators, whether they have the necessary statistical literacy or not can be determined by studying the datalization content they produce. As an attempt, we selected 200 texts from various types of media datalization content, studied on whether there are some statistical concepts and causal inference missing in these texts. It was found that about 65% of the texts had no problem basically, and other texts were insufficient in one way or another. Among them, the common lack is mainly reflected in the following five areas:

The first is some unclear basic concepts of statistics. For example, people lack of necessary understanding of basic statistical concepts of the sample, error, average number, and lead to a misreading of the datalization content.

The second is that conclusions are generated from samples without representativeness. For example, they deal with the results obtained from the nonprobability sample as probability sample, and apply to some population.

The third is invalid causal inference. Some datalization content producers did not make clear the difference between the descriptive study and interpretative study, while the data of a "what" is described as a "why", by mistake, which causing the causal inference in content actually invalid.

The fourth is that content writing is not standardized. Some datalization content has no explanation of the methods and processes of data collection. Or some provides very little associated information, thus it is difficult to make evaluation to the credibility of the content from this.

The fifth is discussing limited in the numbers only. In a certain sense, the data is no more than an abstract symbol. Only linked with explanation, data can become into information. But unfortunately, in many datalization contents, we can often see references limited in data even only an assembled content.

Although the percentage of 65% of the data may not be a representative number, with this, we still have two judgments: First, the majority of disseminators have the ability to "speak with data", having the required statistical literacy of datalization; second, the "words saying by data" of some disseminators has this or that kinds of lack, the statistical literacy of which needs to be enhanced.

For the audiences, to determine whether they have the necessary statistical literacy can focus on whether they can understand the "words saying by data". Therefore, if the audiences are able to understand the statistical terms often occurring in datalization content, they can be considered possessing basic statistical literacy.

In this sense, the measurement of the so-called audience statistical literacy should include two aspects, one is to determine the most common statistical terms, and the other is to know the audiences' understanding of these statistical terms. It is not difficult to assert these days, with the use of Internet search tools, in fact, it is very easy to determine the most frequently occurring statistical terms. Thus to design a way to measure the audiences' understanding of these statistical terms, should not be a difficult implementation.

To verify the feasibility of the above-mentioned measurement mode, we conducted a pilot test over 100 subjects. First, with the help of web search tools, we determined the five high frequency statistical terms in datalization content, i.e. the "probability", "sample", "average", "CPI" and "GDP". Afterwards, the respondents were asked to express their awareness of these terms. The test showed that 45% of the respondents have basic understanding of these terms. Of course, whether 45% is representative remains to be further measured.

However, this test tells us that in the audience level, the usual misunderstanding and misreading are mainly in three aspects:

The first is the lack of basic understanding of the average number. In datalization content, we can often see the average number, but in the case of non-normal distribution,

the amount of information of average number is very limited. Unfortunately, a lot of the audiences are not aware of this, misunderstanding and prejudice consequently sprout.

The second is associating themselves with statistics data. Statistics data is derived from the individual, but not indicating a specific individual. But some audiences will compare statistical data with their individual situations, thus forming incorrect judgments.

The third is some audience misunderstanding of the common statistical indicators. For example, GDP is absolutely a high-frequency word in datalization content, but few audiences can grasp the connotation of GDP correctly.

Therefore, we can assert that, despite the fact that the statistical literacy of the audience has been greatly improved, this improved statistical literacy is also on a relatively low level. Or in another word, there is a certain distance between audience statistical literacy and the required level of media datalization.

#### **4. Media datalization boosts statistics literacy**

We believe that the media datalization provides an all new path of the development of statistical literacy, when opening up a new path of media disseminations.

This path implies three cycles.

The first cycle is the cycle of information. Information dissemination, i.e. datalization content, is the basic mission of the media datalization. The logic is: according to the understanding of the audience interest and statistical literacy as well as data resources at their disposal, the media interpret and package valuable data and present data in the media. In turn, once the audiences contact data content, they have understandings of themselves, part of which feed back to the media, thereby affecting the choices and actions of media about datalization content; at the same time, part of the audiences will be transferred to disseminators, either providing their opinions to the media, or express their opinions on the internet media, becoming the manufacturer of the datalization content.

The second cycle is the cycle of trust. Media datalization depends on the audiences' trust in the media and data producers. Nowadays, the society develops more and more demand of statistics data, in the meanwhile, criticism and disapproval of the data, especially the official data, become more and more common. Why is this? The reason is very complex, and there is a lot controversy over it, but the lack of social trust is one of the reasons seemingly to be the consensus of all parties. Under the social atmosphere of the absence of trust, it is very difficult for data producers to promote trust, not to mention the fact that there are also huge amount of questionable data. However, in this context, the media also maintains and even strengthens datalization content, which is a precious trust and support to data producers.

The third cycle is the cycle of literacy. Media datalization requires the disseminators possessing certain statistical literacy. Once the disseminators recognize their statistical literacy is not enough to meet the requirement of datalization content, naturally, it drives them very hard to improve their literacy. At the same time, the media datalization actually has been nurturing and culturing audiences' statistical literacy. In particular, the booming network media, provide audiences with a situational learning environment. When encountering statistical concepts and methods hard to understand

completely, it is easy for the audience to find answers through linking in the network media. In fact, for the moment, interpretations of almost all the statistical concepts and methods can be found in the network media. Such interpretations, provide a very convenient tools for the audiences to develop statistical literacy.

According to the theory of knowledge gaps, when we feel our knowledge gaps, curiosity will generate which in true trigger action for knowledge search. The problem is, before acting, you must be aware of the existence of the gap. Contacting datalization content, a considerable number of audiences will naturally ask the following questions: Can the data be trusted? Are the samples representative? Is the conclusion from the data credible? Is there a better way to show the data? Is there some kind of mystery hiding behind the data? To answer these questions, some can use existing knowledge, while others need to rely upon new knowledge. Thus, the so-called knowledge gaps will be filled. With the help of the network media, it is not difficult to fill the gaps. With the assistance of network media, when the audience constantly enriches the knowledge of statistics, they can also stimulate their interest in statistics, or even train themselves to think like statisticians.

On our survey to the audiences, many audiences told us: the datalization content gave them knowledge, at the same time, help them know what they did not know before. Particularly, the network media could help turn them from the unknown to the known. And in the process, they could understand the statistical concepts, the “words data saying”, and experience the fun of statistical thinking.

If we say that the literacy cycle directly helps the audiences to develop statistical literacy, then the information cycle and trust cycle, are an incentive to enhance the audiences’ statistical literacy. Information-rich means there are sufficient space for statistical literacy, which should be a positive promotion for the development of statistical literacy; trust-existed, means the acceptance of the datalization content, which, to a certain sense, explains the value of statistical literacy. It is not difficult to understand, without identification, there is no space for media datalization, to develop and no justification to enhance statistical literacy.

## **5. Conclusion**

From the analysis above, we can see that there is a positive feedback between the media data and statistical literacy: on the one hand, the higher the statistical literacy of the disseminator is, the better the presentation of the media datalization, and the greater the motivation to the audiences to contact the datalization content and the better enhance of statistical literacy.

On the other hand, the higher the statistical literacy of the audiences is, the greater the likelihood of exposure and recognition of datalization content is, and the more powerful the motivation for the disseminator to deepen media datalization and enhance statistical literacy is.

Therefore, we believe that statistical literacy is the basis of the media datalization. Without statistical literacy as a prerequisite, the stage for media datalization to play a role will be greatly discounted. Without the existence of the media datalization, the statistical literacy arena will be greatly limited.