What is name identification?

Akiko Aizawa
Professor, Digital Content and Media Sciences Research Division, the National Institute of Informatics

Verification of statements, and costs
When the role of linking records boils down to the verification of people's statements, one cannot help but be reminded of the "name identification" phrase that filled newspaper columns and became a contemporary buzzword during the recent public pensions problem in Japan, in which millions of payment records were lost.

"I haven't been paid the pension I was supposed to be enrolled in," "Who do the disappearing pensions belong to?"

These are simply a question of verifying the veracity of statements about people, and identifying their names to see if they are the same people.

However, the difficulty, which has remained unsolved for over half a century, appears to posses all the essential perplexities of data processing. Vagueness in notation cannot be avoided as long as human intervention is involved, and we have to approach data processing with the premise that this vagueness exists. In doing so, what is important is information about who carried out the verification and in what manner. In other words, the identification of where the responsibility lies, is it with respect to the responsibility that the veracity of names of statements are verified.

However, these costs should be considered an investment in the additional merit of clarifying verification and the locus of responsibility, rather than merely as compensating for the losses caused by incomplete information. Therefore, recording verified statements as if they have actually been verified is liable to mislead, which is exactly what we have witnessed in the case of the pension fiasco.

Furthermore, the fantasy of dramatic cost cuts through automated data processing is, as half a century's research shows, little more than an overestimation of the value of computerization.

Linking each and every one of these geographically dispersed individuals' records is somewhat more difficult than it appears, but it is paramount to civil life because it is the basic data for proving a person's identity and providing the proof by which their family allowances and pensions are paid.

Linking records is the process of checking two pieces of data and deciding whether or not the content confirms. Though this could at first appear to be a task that could be performed easily with computers, this is not actually the case. The Von Neumann-type calculating machine was first conceived in 1946, and the computerization of society subsequently progressed rapidly.

The difficulty of linking records remains unchanged. Let's take a look at a good example of this difficulty by trying to find out about a certain person using search software. First, we enter the surname - resulting in a mountain of data about millions of payment records were lost.

These are simply a question of verifying the veracity of statements about people, and identifying their names to see if they are the same people.

However, the difficulty, which has remained unsolved for over half a century, appears to posses all the essential perplexities of data processing. Vagueness in notation cannot be avoided as long as human intervention is involved, and we have to approach data processing with the premise that this vagueness exists. In doing so, what is important is information about who carried out the verification and in what manner. In other words, the identification of where the responsibility lies, is it with respect to the responsibility that the veracity of names of statements are verified.

However, these costs should be considered an investment in the additional merit of clarifying verification and the locus of responsibility, rather than merely as compensating for the losses caused by incomplete information. Therefore, recording verified statements as if they have actually been verified is liable to mislead, which is exactly what we have witnessed in the case of the pension fiasco.

Furthermore, the fantasy of dramatic cost cuts through automated data processing is, as half a century's research shows, little more than an overestimation of the value of computerization.

The issue of linking records is an excellent example illustrating that the knowledge we automatically obtain from computers and the knowledge verified by experts are both assets, and both essential.
Asao Fujiyama
Professor, National Institute of Informatics

Professor, Graduate University for Advanced Studies

Masago Minami
Senior Editor, The Yomiuri Shimbun (Tokyo head office)

NII Interview: Asao Fujiyama + Masago Minami

The Genome as the Keystone of Integration

Merging Life Science and Informatics

Minami: Taking the integration of life science and informatics as a starting point, I would like to ask you about your areas of expertise, genomics. The Human Genome Project, which represented a major step, could not have been accomplished without computer efforts, could it?

Fujiyama: When tangible work on the Human Genome Project started around 1980, a number of databases pertaining to DNA already existed. The DNA Data Bank of Japan (DDBJ), GenBank in the United States, and the European Molecular Biology Laboratory (EMBL) database are typical examples. Considering the volume of human genome data that was anticipated, it was apparent from the outset that computers would be necessary. But the extent of the need for them was beyond all imagination.

Minami: Compared with the situation then, the information environment supporting research — computers, communications, and so on — has changed greatly, hasn’t it?

Fujiyama: In the 1980s, large volumes of data were shared by putting information on magnetic tapes and then sending them off by airmail. The acceleration of communications and network formation are a result of these efforts. The marriage of biology and informatics was feasible. Minami: In the scenario of a smooth relationship unfolding as envisioned?

Fujiyama: Presently, a number of universities serve as places that offer training in both fields from the outset. When you perform genome research, you work with information that pertains to the entire spectrum of living things, examine relationships among them, compare the genomes of different organisms, and so on. So you’re forced to use computers. The development of people who are capable of doing informatics work in this field has gradually been taking place. Additionally, with a growing need to put the genome and genetic information to practical use at the corporate level, human resources in bioinformatics will be sought not only by universities but also by pharmaceutical makers and other companies. The integrated field of bioinformatics is likely to take hold in this way.

Minami: When did specialists in informatics and specialists in biology begin to cooperate closely?

Fujiyama: Back in the early 1980s, some biologists were also writing computer programs on their own as the need arose. In the eyes of informatics specialists, though, their efforts seemed unpublishable.

Gradually this task was left up to the professionals. But even when biologists and informaticists suddenly came face to face, they weren’t able to converse. It was all they could do to try to share each other’s language and attempt to communicate. There was even debate as to whether the marriage of biology and informatics was feasible.

Minami: Isn’t it the case that the two are becoming inseparable?

Fujiyama: “That sort of thing isn’t science” was the typical comment. Now, though, genome information is there for use by everyone, and you can’t write a paper without using it.

In the case of molecular biology, it probably took about 20 or 30 years to become solidly established. Around the time I was a university student — from the latter part of the 1960s to the first half of the 1970s — early molecular biologists were not accepted by mainstream academia and were struggling hard at schools in outlying areas. At the time molecular biology was a field that integrated chemistry, physics, and biology. The people involved in this field, though, did not particularly perceive it in that way, and they proceeded with a determination to establish a new academic domain. . . . Today the Molecular Biology Society of Japan (MBS) has become an organization with a sizeable membership of 8,000 and is Japan’s largest academic society in the field of basic biology.

Minami: Information also serves as an instrument for promoting integration. Doesn’t it?

Fujiyama: That’s right. Genome data have been made available, and anyone can access this information free of charge without restrictions via the Internet. So this information is playing a significant role in the research activities of life science specialists as well as in the work of researchers in other fields.

Genomics researchers consist of people who have a diversity of backgrounds, including medical scientists, biologists, and zoologists. Plus, researchers in a broad range of fields, for instance, engineering, agriculture, and pharmaceutical sciences, are also engaged in activities based on genome information. When fields differ, communication becomes problematic. One reason for this is the use of different nomenclature for the very same things, including even genes and oxygen. Disparities in terminology could also present a barrier to the integration of multiple databases, genome comparison, and so on.

Consequently, what we are strong to do now is to utilize information theories, including natural language processing and ontology. We are attempting to systematize terminology and concepts and to create a dictionary that people in any field will be able to use. Because an enormous volume of data is being produced, there is also a need to develop technology for the successful extraction of whatever information and knowledge people are seeking.

Minami: Doesn’t information also have another role in terms of informing the public about this field?

Fujiyama: Political forces in favor of sending out information have gained powerful momentum. In reality, however, the people doing research don’t have time for that. We’re creating a site called Jabion, a Japanese-language biotechnology portal. One of the reasons for launching it is our belief that scientific articles in newspapers may be somewhat difficult for the average person. There are also instances when limited page space makes it hard to convey background information adequately. Our idea is to take life science themes that are likely to attract attention within society and provide information as plainly as we possibly can.

A word from the Interviewer:

It’s really something that NF has biologists and that biology and informatics are becoming so integrated. While continuing to conduct experiments, Professor Fujiyama — who has been involved in the Human Genome Project from its inception and has always been a pioneer in this field — has initiated a project for the general public to become acquainted with today’s biology via the Internet. The number of newspaper articles on basic science is still quite low, with some data indicating that they account for a mere 0.1% of articles in general. I think that a portal site for researchers to transmit information on their own is extremely valuable. I’m excited about Professor Fujiyama’s work, and I plan to keep an eye on its progress.

Jabion: Japanese biotechnology portal site

A Japanese-language resource for conveying up-to-date biotechnology information in a way that is easy to understand

Jabion: 日本語バイオポータルサイト

Japanese biotechnology portal site

to the integration of multiple databases, genome comparison, and so on.

Consequently, what we are strong to do now is to utilize information theories, including natural language processing and ontology. We are attempting to systematize terminology and concepts and to create a dictionary that people in any field will be able to use. Because an enormous volume of data is being produced, there is also a need to develop technology for the successful extraction of whatever information and knowledge people are seeking.

Minami: Doesn’t information also have another role in terms of informing the public about this field?

Fujiyama: Political forces in favor of sending out information have gained powerful momentum. In reality, however, the people doing research don’t have time for that.

We’re creating a site called Jabion, a Japanese-language biotechnology portal. One of the reasons for launching it is our belief that scientific articles in newspapers may be somewhat difficult for the average person. There are also instances when limited page space makes it hard to convey background information adequately. Our idea is to take life science themes that are likely to attract attention within society and provide information as plainly as we possibly can.

A word from the Interviewer:

It’s really something that NF has biologists and that biology and informatics are becoming so integrated. While continuing to conduct experiments, Professor Fujiyama — who has been involved in the Human Genome Project from its inception and has always been a pioneer in this field — has initiated a project for the general public to become acquainted with today’s biology via the Internet. The number of newspaper articles on basic science is still quite low, with some data indicating that they account for a mere 0.1% of articles in general. I think that a portal site for researchers to transmit information on their own is extremely valuable. I’m excited about Professor Fujiyama’s work, and I plan to keep an eye on its progress.

Jabion: Japanese biotechnology portal site

A Japanese-language resource for conveying up-to-date biotechnology information in a way that is easy to understand

Jabion: 日本語バイオポータルサイト

Japanese biotechnology portal site

to the integration of multiple databases, genome comparison, and so on.

Consequently, what we are strong to do now is to utilize information theories, including natural language processing and ontology. We are attempting to systematize terminology and concepts and to create a dictionary that people in any field will be able to use. Because an enormous volume of data is being produced, there is also a need to develop technology for the successful extraction of whatever information and knowledge people are seeking.

Minami: Doesn’t information also have another role in terms of informing the public about this field?
Chemical information research can give rise to results that lead to new ways to use natural resources. However, to get to these results, it is necessary to understand the kinds of the other research members to set effectively research cooperation harmony. Here is a case report of genuine integration of chemistry and information science motivated by the CAST (Chemical Information System) method.

The CAST method was a joint research project from the 2000s to 2010s involving Satoh, Uno, and Iwata. While trying to clarify the chemical structures of new substances found in marine life, Satoh and Uno identified all the problems. When Uno thought their efforts would be in vain, Iwata identified the CAST method. This led to the development of the CAST/CNMR system, which uses a coding method for chemical structures.

Starting from three-dimensional chemical structure data, the CAST/CNMR system generates a set of codes that represent the molecule. These codes are then used to predict chemical shifts. The CAST/CNMR system has been applied to a variety of molecules, and its predictions have been shown to be accurate.

The ideal form of integration is one in which the two research members are working together, and the collaborating research is leading to results that set a new direction in research. The CAST/CNMR system has been a success in this respect.
Transdisciplinary research can give rise to results that significantly differ from those achieved by teams working separately. However, to get the benefit of the research collaborations, it is important to understand the fields of the other team members in an effectively comprehensive manner. Here, an example of genuine integration of chemistry and informatics is described.

One of the most important tasks of a computer scientist is information retrieval. While this can be summed up as “molecular structure”, there is more to it than that. There are various algorithms that require the structure around the carbon atom. However, recognizing ring structures included in a molecule is a problem for chemists.

Structure determination is a key problem in the fields of chemistry and informatics. The State of Japan has a major system for the identification of molecules, the CAST/CNMR, which has been applied in academic research. It has been utilized in a variety of fields, including the fields of chemistry and informatics. The State of Japan has a major system for the identification of molecules, the CAST/CNMR, which has been applied in academic research. It has been utilized in a variety of fields, including the fields of chemistry and informatics.

Molecules in CAST/CNMR can be identified through the CAS system. The CAS system is a computer program that can be used to identify molecules. In addition, the CAST/CNMR system is a database that contains information about the CAS system.

In order to identify molecules, it is necessary to use the CAST/CNMR system. The CAST/CNMR system is a database that contains information about the CAS system.

Researchers working on the CAST/CNMR system include Takeaki Uno, who is a professor at the University of Tokyo, and Mr. Shungo Koichi, who is a Ph.D. student at Tokyo University. Uno is developing algorithms for recognizing ring structures. He is working on developing an algorithm that can accurately recognize ring structures.

Uno is working on developing an algorithm that can accurately recognize ring structures. He is working on developing an algorithm that can accurately recognize ring structures.

Uno is working on developing an algorithm that can accurately recognize ring structures. He is working on developing an algorithm that can accurately recognize ring structures.
The New Value in Information Linkage

Today’s researchers are looking for new value in information by using “information linkage”—assembling information about the same item or person and placing it under integrated control. Drawn by the theme of information linkage, what kind of new value are they aiming for?

“Information linkage” is conventionally known as “account consolidation.” (“*1) “While it’s not very noticeable, it’s an extremely important operation,” says Akiko Aizawa, leader of the “Collection and Analysis of Data for Large-Scale Linkage” group at the National Institute for Informatics Transdisciplinary Research Project.*2 Professor Aizawa’s original motive for making use of information linkage was as a form of “databases-cleaning,” namely, “to eliminate duplicate records in databases of academic papers and online catalogs.”

For example, when the same family and personal name appear multiple times as an author record, it’s important for the user of the database to know whether it is a single duplicate record, or a different person with the same name. Using information linkage to improve the quality of the database has great benefits for actual use of the data (see accompanying graph).

With the spread of computers and gigantically growing networks, the volume of data not only continues to expand, but it also is uncontrollably disorder- and unordered. There are indeed innumerable occasions when information linkage is required to organize and order information in response to specific needs.

Information Linkage and Databases

Not all information linkage is the same, however, and the linkage involved can occur in several different patterns. Professor Aizawa states that “the problem setting can be divided into four levels based on disorderliness.”

First, as represented by the example posed earlier, one can consider operations performed within a single database. A database is a collection of data assembled in categories and formats for a specific purpose, with the result that it is relatively easy to compare data items and determine whether they are the same (deduplication).

Next is the case of integration across two or more databases. Information linkage (“account consolidation”) had unintentionally become known recently after the Japanese pension-record mishandling re- ports, in which comparison and identification was performed on personal records scattered across databases in different locations, and those which could be determined to belong by a single individual were then integrated. In that way, integration can be per- formed across the entire database.

Up to this point we’ve been talking about databases with established data formats. But there are great quantities of useful information outside of databases proper, on the Internet, for example. The remaining problem is how to draw that totally heterogeneous, disordered information into linkage. Namely, the third level is the linkage of information in databases with that outside databases, while operat- ions entirely outside of databases represent the fourth level.

Toward Concrete Services

Even taken alone, this degree of information linkage has a great deal of real value, but the purpose of the project is to use linkage to assemble information and then to effectively utilize the “something extra” created from that assembled data. Associate Professor Kei Kurakawa, working at the Research and Development Center for Scientific Information Resources (director, Hideaki Takada) is currently engaged in joint research aimed precisely at gaining concrete shape to that “something.”

“With the purpose of developing academic and sci- entific contents services, we are first doing name identification research,” says Professor Kurakawa. “This will be of benefit, for example, when looking for a joint researcher, or when evaluating researchers’ perfor- mance.” Name identification is being performed on the Ministry of Education, Culture, Sports, Science, and Technology’s researcher number database for grants-in-aid for scientific research, together with databases of researchers at individual universities; this thus can be called a concrete implementation of the “second level” of linkage. Masao Takaku, a re- searcher at the Transdisciplinary Research Integration Center, is focusing on the same research-grant data- base to reformat the data, focusing on links between individual researchers. By paying attention to joint researchers, and the relationships between team leaders and their associates, Takaku says “we may be able to understand, for example, things like who should be made team leader in order to facilitate re- ceiving research grants.” The kind of useful data produced depend on the kind of information link- age employed.

Foreign student interns are also active in Aizawa’s research group. Dang Bac Van is a master’s student at Vietnam National University - Ho Chi Minh City, identification results, where he is engaged in re- search on the extraction of information from text data. This is an attractive research theme with broad practical applications. For example, by extracting the names and relationships of methods and tools de- scribed in an academic research paper, it may be- come possible to apply these in other fields. Van is currently giving his efforts to the theme of how to extract test data without large investments of hu- man labor.

Associates Drawn Together by Information Linkage

The various members of Professor Aizawa’s re- search group were not all originally involved in infor- mation linkage research, and there were few com- monalities between them. Professor Aizawa began from communications, going on to machine learn- ing, optimization, text and natural language process- ing. Associate Professor Kurakawa, by contrast, studied engineering design science and software design. The only point in common with Professor Aizawa was the fact that he was utilizing the same data. Most of the researchers assembled in the group are like this.

Also, the everyday research landscape is different from the usual image of the physical sciences re- search environment, where associates are pushed to gather in the same room each day. Instead, rooms are provided where all members can gather indepen- dently to discuss issues as required. Since Van is a student, he has regular meetings with Professor Ai- zawa, who, he says, “has given me numerous good ideas, thus allowing me to engage in productive re- search,” but in principle, such meetings are held only when considered necessary. But when they do gather, the group members display strong unity. Ac- cording to Aizawa, “Everyone has a different back- ground, so we are stimulated anew each time we have a discussion, making it a very enjoyable envir- onment for research.” Truly, this is a research group brought together through “linkage.”

The Goal: Easier-to-User, More Valuable Data

Professor Aizawa states, “At present, we have cre- ated a system that can handle information linkage without problem up through the second level. But since human resources are required to verify the identification results, we run up against the prob- lem of what degree of quality we can achieve at what cost.” Many issues are amenable to solution by sharing databases, such as the problem of con- centrating research grants on specific individuals, and the problem of pensions noted earlier. The rest is basically up to implementation policies.

When it comes to future directions, Professor Aiza- wa suggests that “by using information linkage, we have basically come to the point of being able to un- derstand to some extent things we were unable to grasp at all before. We next intend to assess the value of that partial knowledge and create a system that will provide truly useful information to society.” On the side of data as well, he is also considering ideas for new kinds of services. It is thrilling to contemplate what kind of new value will be created by these re- searchers, assembled under the umbrella of informa- tion linkage.

(Written by Tomoaki Yoshito)
Building a Proud Heritage: NII Graduate Education

The National Institute of Informatics (NII) has three faces: a center of comprehensive research on informatics, an interuniversity research institute, and an educational institute where individuals receive high-level training in informatics. In order to get a better idea of the current status of the Institute’s program of graduate education, I interviewed Professor Noboru Sonehara, acting manager of student recruitment, together with several recent graduates and current students.

Somehara: I’ve asked you all here today to talk about NII’s graduate school. I’d like to start by asking for self-introductions, and the reasons you decided to apply here.

Ohmukai: I’m studying how blogs and online social networks change our ways of communication. I was in Kyoto through my M.A., but I felt there wasn’t enough information for my research there, so decided to come to Tokyo.

Kanokwan: I came to Tokyo to study the Japanese lifestyle. While searching on the Internet for graduate schools matched with my interest on electronic commerce, I found NII’s students also influenced me to pursue a Ph.D. here. Currently, I have been researching on the factors affecting the success of electronic commerce.

Ph.D. Course for Study and Research

Kanokwan: The research facilities here are modern, and instructors are all professionals. The atmosphere makes it easy to discuss issues with instructors and therefore I never have any uncertainty with regard to my research.

Kajiyama: While being in the Ph.D. course means you’re a student, at many universities you have responsibility for teaching the younger students and don’t have time to devote yourself to your own research. On the contrary, NII has allowed me to fully use my time as a student for effective learning.

Ohmukai: In addition, I like the atmosphere of being able to do free research. It’s so free, in fact, that some people might feel neglected, but it’s ideal for students who want to set their own theme and design their own research program.

Somehara: Fundamentally, research is ﬁred by the possibility of free creativity. We’re proud of the fact that it is our status as a small graduate school that allows us to offer an open environment for research.

In order to encourage our students to push forward with un求助 research, NII’s network and online academic contents are used to publish their research results throughout the world.

Kajiyama: NII has the will to publish research results to larger society; they even made a patent application based on my research.

Ohmukai: I was in school, I started a company to develop an RSS (Rich Site Summary) reader, a tool for reading online blogs.

Sonehara: Of the eighteen central research institutions within the Graduate University for Advanced Studies, we are the only comprehensive educational institution dedicated to research on informatics. Part of the mission of the faculty and staff here is to disseminate Japan’s intellectual property strategy, so we are constantly aware of the situation in the larger world outside.

The Synergy of Diverse Faculty and Students

Ohmukai: Another characteristic of NII is the diversity of the people here. Of the students in attendance, one-third are foreign students, and pre-half already have post-graduate employment experience. This is a balance unheard of at other graduate schools. Many of the instructors have also experienced employment in the private sector, outside of academia.

Kajiyama: That’s why it wasn’t as closed an environment as an ordinary university, and it represented a better situation for me before going out into the world. My desk was near Ms. Kanokwan’s, so we became friends, and I think that we provided each other with good intellectual stimuli.

Kanokwan: What I find interesting now is the fact that NII students are all people who accepted the challenge of ‘changing their environment.’ This goes not only for foreign students and older students with employment experience, of course, but I think all those who choose the Ph.D. program here have challenging personalities. If you change your location—your environment—your way of thinking also changes, and that is linked as well to changes in your research topics. NII is a gathering of kind of all people, so they’re all serious thinkers.

NII’s Broadening Web of Influence

Ohmukai: I’ve been a research staff here for three years, and I feel the environment for research is really tops. I’d like to stay here forever, but on the other hand, I’ll do good research, you also have to maintain connections with the outside world. For the present I’m trying to keep up my contacts while focusing myself here at NII.

Somehara: We tell our graduates we want them to aggressively go out and get jobs in society. The times demand that kind of “centrifugal force.” When I said “centrifugal force,” I mean to transmit the things students learn and acquire here at NII to the outside world at large. After all, that should be the mission of Japan’s highest institution of learning.

Kajiyama: I’ve been employed at Waseda University since this April. I’m a woman, and when I consider that the present time is a priceless period that I can use for my own purposes, I thought I’d like to concentrate on my research. But now I really feel a significance in passing on what I’ve learned to students.

Ohmukai: Most of the graduates from NII take employment at organizations where they previously performed joint research as students. That’s great because the work they did was clearly recognized, so it’s something to be proud of.

Kanokwan: I originally thought I would return to Thailand after graduation, but now I would like to continue my research at either NII or a Japanese corporation—any place giving me an opportunity.

Kajiyama: Although this has the perfect environment for research, it’s unfortunate that the graduate school itself is not very well known.

Ohmukai: I like new things, and I think their name value can be made by yourself. I really feel that the things I’ve done are helping to write the history of NII.

Somehara: With the current era of lowering birthrates, every university is fighting to acquire its share of the student population. We’re also trying to create an environment that is amenable for learning, both for exchange students and for older students who are returning to school after years of employment. But enhancing the brand identity and value of an NII education depends on the activity and results produced in society by people like those assembled here today. NII will continue to evolve and improve so long as the two wheels of research and graduate education continue to turn smoothly. Thanks to all for taking time from your busy schedules to talk with us here today. (Written by Akiko Ikeda)
Weaving Information into Knowledge

record.

The difficulty of linking records

Verification of statements, and costs

Verification of statements, and costs

All sorts of events occur during the course of peoples’ lives—are ordinary things such as births, deaths, marriage and divorce, no adoption and immigration. In all of these cases, various forms of verification have to be made, often happening across a wide geographical range. And since these documents are handled by people and therefore liable to human error, misspellings and data input errors can occur.

The difficulty of linking records

Linking records is the process of checking two pieces of data and deciding whether or not the content conforms. Though this could at first appear to be a task that could be performed easily with computers, this is not actually the case. The Von Neumann-type calculating machine was first conceived in 1946, and whilst this was a huge step forward for the computerization of society subsequent progress rapidly the essential difficulty of the task of linking records remains unchanged.

Let’s take a look at a good example of this difficulty by trying to find out about a certain person using search software. First, we notice the surname—resulting in a mountain of data about the same name and the names of places etc. Entering the surname as keywords about the person’s affiliations etc. results in an additional layer of data about the person’s work history and career etc. This way, the search becomes a matter of sorting through a great deal of data until one finds the correct record.

Verification of statements, and costs

Verification of statements, and costs

When there is a case of linking records, one is concerned with the verification of people’s statements. One cannot help but be reminded of the “name identification” phrase that filled newspaper columns and became a contemporary buzzword during the recent public pensions problem in Japan, in which millions of payment records were lost.

“I haven’t been paid the pension I was supposed to be entitled to. Who do the disappearing pensions belong to?”

These are simply a question of verifying the veracity of statements about people, and finding the person to whom they pertain. Verification is a matter of clarifying who the person is and the purpose of their statement. Verification is a matter of clarifying who the person is and the purpose of their statement. Verification is a question of verifying statements, whether or not the person in one record is the same as the person in another record is a matter of human error, misspellings and data input errors can occur.

The difficulty of linking records

Linking records is the process of checking two pieces of data and deciding whether or not the content conforms. Though this could at first appear to be a task that could be performed easily with computers, this is not actually the case. The Von Neumann-type calculating machine was first conceived in 1946, and whilst this was a huge step forward for the computerization of society subsequent progress rapidly the essential difficulty of the task of linking records remains unchanged.

Let’s take a look at a good example of this difficulty by trying to find out about a certain person using search software. First, we notice the surname—resulting in a mountain of data about the same name and the names of places etc. Entering the surname as keywords about the person’s affiliations etc. results in an additional layer of data about the person’s work history and career etc. This way, the search becomes a matter of sorting through a great deal of data until one finds the correct record.