Science Journal of Business and Management
2015; 3(1-1): 66-72
Published online January 16, 2015 (http://www.sciencepublishinggroup.com/j/sjbm)
doi: 10.11648/j.sjbm.s.2015030101.21
ISSN: 2331-0626 (Print); ISSN: 2331-0634 (Online)

Data, information, knowledge in FUTÁR: Case study of a public transportation information system

László Tóth, Péter Szikora

Keleti Faculty of Business and Management, Óbuda University, Budapest, Hungary

Email addresses:
lacka88@gmail.com (L. Tóth), szikora.peter@kgk.uni-obuda.hu (P. Szikora)

To cite this article:
László Tóth, Péter Szikora. Data, Information, Knowledge in FUTÁR: Case Study of a Public Transportation Information System. *Science Journal of Business and Management.* Special Issue: The Role of Knowledge and Management's Tasks in the Companies.
Vol. 3, No. 1-1, 2015, pp. 66-72. doi: 10.11648/j.sjbm.s.2015030101.21

Abstract: Data is not information, and even information is far from knowledge. On the one hand organizations tend to gather large amounts of data just for the sake of collecting it, without a clear plan, or even concept on of how to use them in the future generating information overload. On the other hand in most of the situations, besides having overload from irrelevant data, crucial information might be missing which makes optimal decision impossible. Present paper endeavors to introduce and analyze a system – namely that of the FUTÁR project – that is well equipped for collecting data and has a well-functioning inner logic to create information from the assembled data. What is lacking is the understanding of the possibilities this system is providing to its users, and the realization of ideas – the application of knowledge – for which it has been established.

Keywords: Information, Information System, Knowledge, Public Transport

1. Introduction

Nowadays in the information society one of our most difficult challenge is transforming data into information, and to convert that information into useable knowledge. Gathering data itself is easy, and is being done by most organizations, for-profit or state run. However, organizations tend to gather large amounts of data just for the sake of collecting it, without a clear plan, or even concept on of how to use them in the future. One of these organizations is Budapest Transport Center, where the recent introduction of an automated, GPS-based traffic management and passenger briefing system called FUTÁR have accumulated significant amounts of data. Providing examples of how to process this pile of information at BKK is my goal in this paper.

2. Pyramid Model of Data, Information, Knowledge, and Wisdom

Although in everyday language they are used as synonyms, there is a huge difference between Data, Information, and Knowledge. Based on Russell Ackoff's theory [1], the relationship between these concepts can be represented with a pyramid, where data is at the base of it, and we can work our way up to wisdom through information and knowledge. Raw data can be collected by automated systems, and it should be used for further processing to synthetize information by putting it in context, and thus giving it meaning. When we collect information for a specific use in mind, we will (or at least can) end up with knowledge. Information is a sequence of data with a specific meaning, while knowledge enables its owner to act [11]. What knowledge gives us is an answer to questions or problems that we have come across in the past, but not to ones we have not yet. We need to have an understanding of a concept to be able to synthetize new, probable knowledge through interpolation of existing ones. Wisdom is achieved when we have understood the principles, and we can extrapolate from the knowledge we have, and have an understanding in a field where there is no data to gather, and no information can be generated. It allows us to be creative, or to make decisions based on our values.

Nonetheless, people often confuse data, information, and knowledge. More data is often equated with higher levels of accuracy, and gives birth to a heuristic that the accumulation of data results automatically in good decision-making. [2] Accordingly not only information is necessary to make a decision, but in many cases, besides the inner logic of the data gathered, the embeddedness of information – in other words, knowledge – is required.

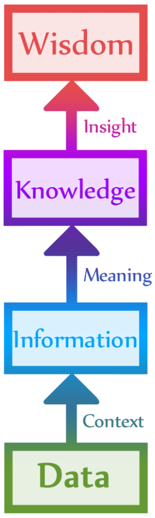

Figure 1. *Flowchart model of Data, Information, Knowledge, and Wisdom.*

Polányi [13] identified two kinds of knowledge, tacit and explicit. Tacit knowledge embodies experiences and the understanding of the situation at hand, while explicit knowledge consists of elements that can easily be articulated (coded) and transmitted. Hence explicit knowledge is directly connected to information and logical, rational thinking, while tacit knowledge requires qualities that necessitate human qualities such as intention, imagination and creativity.

Knowledge does not emerge from itself, it has to be generated. Nonaka and Takeuchi identified [14] four processes of knowledge creation, where they were using Polányi's basic categories. According to their theory information can be internalized by humans, meaning that they try to understand the message the data and information tried to capture, and by creating an understanding, they already created knowledge.

The antonym of this is externalization, where someone's knowledge has to be incorporated in words and/or numbers, so it has to be captured in information pieces.

It is very hard to externalize tacit knowledge, since most of the time one doesn't even know the name of the action, or the cause of the gut feeling that is directing his/her actions.

Externalization of explicit knowledge on the other hand seems easy. Information based knowledge with a clear inner logic is easily transmittable by first explaining the underlying logic and then providing the missing data and information. Even new knowledge can be created this way, by combining the logic with new data, or assembling the already existing information with a different logic in mind.

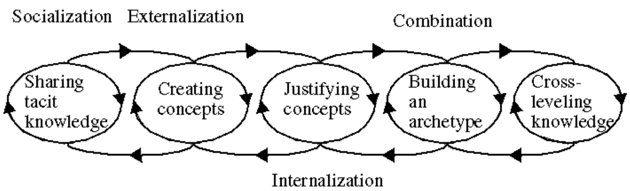

Figure 2. *Source: Nonaka Takeuchi, 1995*

While life is a chain of decisions, this belief can be dangerous. The most important parts of the decision as a process are the assembly and analysis of information. The more information is available in relation to the problem at hand, the better one can define the options for the actions, as well as their assessment. [11] However, complete information can never be achieved. We can never gather each and every piece of the necessary information, since there is not enough time for it, or the information is not accessible for us, or we do not have the financial resources to be able to obtain them. What is more, even if we could do so, the amount of such information would be so excessive that we would not be able to cope with them or process them. In line with this, all decisions' inherent feature is the lack of information, and thus uncertainty.

However, the lack of time and resources to gather all the information for a given problem is not our only concern. The other side of the coin, namely information overload is also an imminent threat in some cases. The amount of information available to people is growing rapidly. The Internet and the 24/7 accessibility of it bombards its users with information. Hence people have to deal with escalating amount of information in their everyday life. However, there is a limit to information processing. According to Miller [12], a human is only able to perceive and process at most 7+/-2 digits of information. What is more, sociologist George Simmel already in 1950 – well before the IT revolution – has pointed out [15] that information overload takes away people's ability to react and renders them too tired to make decisions. And this is only about processing information, and not about storing or recalling them. Storing and recalling would need an inner logic, a so-called human "database structure", where information about related topics are related, hence the occurrence of a situation can trigger the recollection of all information that is somehow connected to the situation at hand. Accordingly, one of the biggest deficiencies of information processing is the lack of applicable schemes or the lack of ability to create such.

While information is necessary for good decision-making, alone even in big mass it is insufficient. Information must be tailored to the knowledge and experience level of the user. Hence good information is:

- Exact
- Timely
- Operational
- Comprehensive
- Accurate

In short, information helps us decide what to do, not how to do it. The "how" requires knowledge.

In addition to this, our life with its confusing inner logic is much more complicated than what we could possibly understand and simulate, so a perfect model can never be prepared. With the help of the internet unlimited amounts of data is accessible for us, whether they are providing information, or are ready to be internalized, or can only be recombined but not fully understood is up to the individual's characteristics – skills and competences. So a further limit to information processing can be the cognitive limits of humans on attaining and using information. [16]

3. Current Situation of the FUTÁR

Since Hungary with its less than 10 million inhabitants is heavily centralized - one-fifth of its inhabitants living in or around the capital, Budapest – there is a big need for a well-organized public transport that can satisfy the needs of such a big audience. [3][4][5][9]

The FUTÁR project has been initiated by the Budapest Transport Corporation (BKV) in 2009 because changes in legislation have forced the company to abandon its analogue trunked radio system and transfer to a digital radio system. BKV estimated the transition to cost about 2.5 billion HUF, and they've started to look for a way they can involve some EU funding in the project. They have succeeded in this search, and the replacement of the radio system was connected with the development of a traffic management and passenger briefing system with a total cost of about 6.7 billion HUF, 4 billion of it coming from the EU.

BKV had a pilot project preceding FUTÁR on bus line 86 installed in 2007, which has been used to gather experience with GPS-based vehicle tracking systems, and the integration of their data in passenger briefing systems. This project used an off-the-shelf device made by a Hungarian company, but it was eventually abandoned, and was replaced by FUTÁR. The pilot project also included four so called "smart tables", which displayed the position of the buses using a map and some LEDs on the route, and a small screen telling the distance of the nearest bus and the estimated time of its arrival. This LED assisted map was discarded in the final system, and present displays only show the time the next bus will depart from the stop, not the distance it need to travel.

Based on the experience acquired in this pilot project, BKV has been able to specify its requirements for the final system. These were:

- text and audio connection between driver and traffic control using digital radio
- online tracking of vehicles using GPS data
- displaying deviations from the timetable to both the driver and traffic control
- handling traffic disturbances, and maintaining passenger briefing during these
- handling canceled starts and breakdowns happening on the line
- controlling automatic passenger briefing such as vocal and textual information about the next stop, headsigns, and special announcements
- handling real-time displays and loudspeakers in stops
- controlling onboard ticket machines
- giving preference to public transport in selected intersections
- provide information about the routes and vehicles to passengers online

4. Gathering Data and Information

The FUTÁR system's On-Board Units (OBUs) send data about their status every 30 seconds via a GSM/3G datalink. These packets include data on the current position of the vehicle, whether or not the vehicle is staying in a stop with doors open, the vehicle's registration number, the current line and trip number, and the driver's ID. Some of this data is used in passenger information, some, like the driver's ID, is only visible to dispatch. All data generated is stored and backed up, and later can be used to fine tune timetables, or verify passenger complaints.

This database/set is not publicly available, and I couldn't get access to it from BKK yet. The closest we can get is through a 3rd party website web-schedules.rhcloud.com, which acquires its data from the official online site at futar.bkk.hu. Unfortunately this way we can only get data from the current day, not the ones before.

To work around these issues, I wrote a shell script that downloads every trip from a given stop on the specific day, and scheduled it to run every day at 11 pm. I had to choose a specific line I wanted to investigate because this third party website isn't very fast, downloading large amounts of data wasn't an option. I've decided on bus line 5 for various reasons:

- this is a line I've been using almost every day since it was extended in 2008,
- it is considerably long: 19.5 km, 46 stops, 60 minutes in one direction, and
- it has a diverse route featuring the suburbs, a part where it stops in the same stations as trams, the city center, and Rákóczi way, which has the most busy bus lane in the city.

The downloaded pages have been processed by another shell script, generating a CSV file with the following data structure:

- sequence number of the stop (as in first, second, …)
- ID of the stop (e.g. F03033)
- date
- route (in this case 0050 for bus line 5)
- direction
- trip
- start time (departure from first stop)
- difference from timetable

This CSV file was then imported to Excel, where the difference of the timetable differences for each stop was calculated to see how a vehicle's delay changed through the stops. The resulting spreadsheet was then imported to an Access database. I've run some queries on the data, grouping them by direction and time of day: dawn, morning rush, midday, evening rush, night. The queries included an average of the delays and the delay variations for every stop for a given direction and time of day. The queries' output was then brought back to Excel, where I've generated graphs from it.

5. Analysis

For this demonstration, I'll be using the morning rush graph from the second week of November 2014 in the direction of Buda.

The graphs show the stops of the line on the horizontal scale, and the delays in seconds on the vertical scale. There are some interesting points we can already see here, such as how the average bus departs from the first stop half minute late, and arrives in the last station with a six minutes delay.

As I've started my analysis, the first thing that came to my attention was the significant spike at the middle of the route after Mexikói út M, circled with red in Fig. 4.

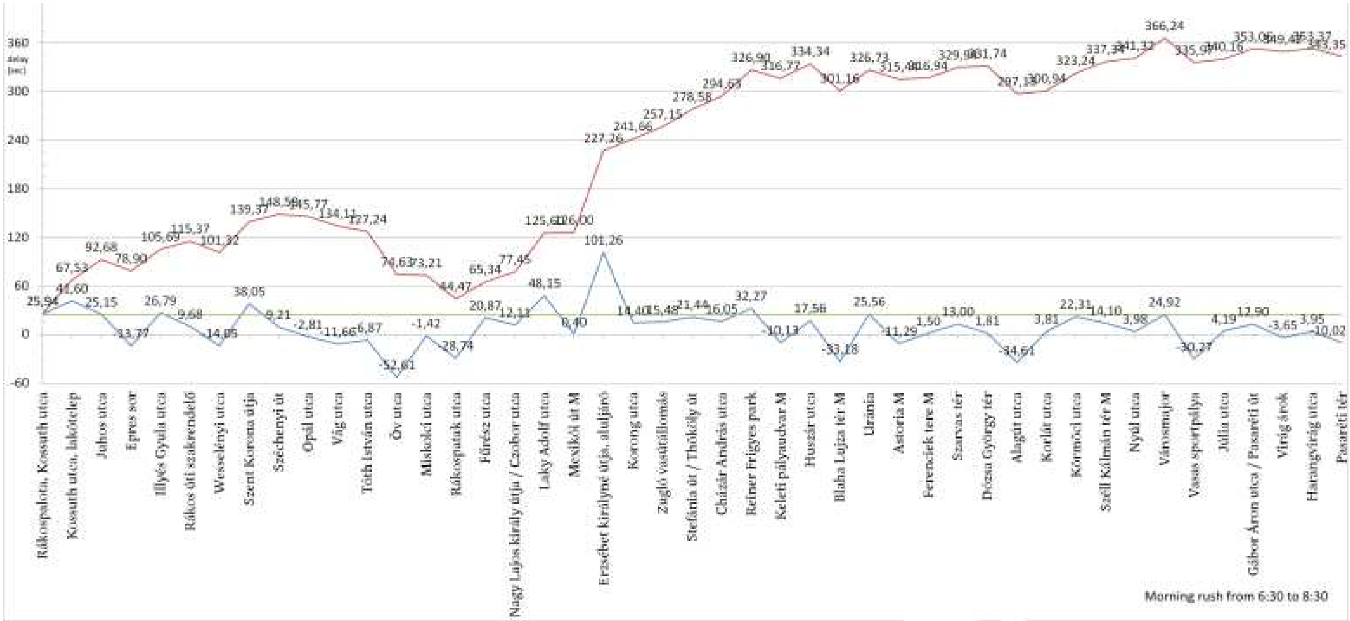

Figure 3. *Visualization of data.*

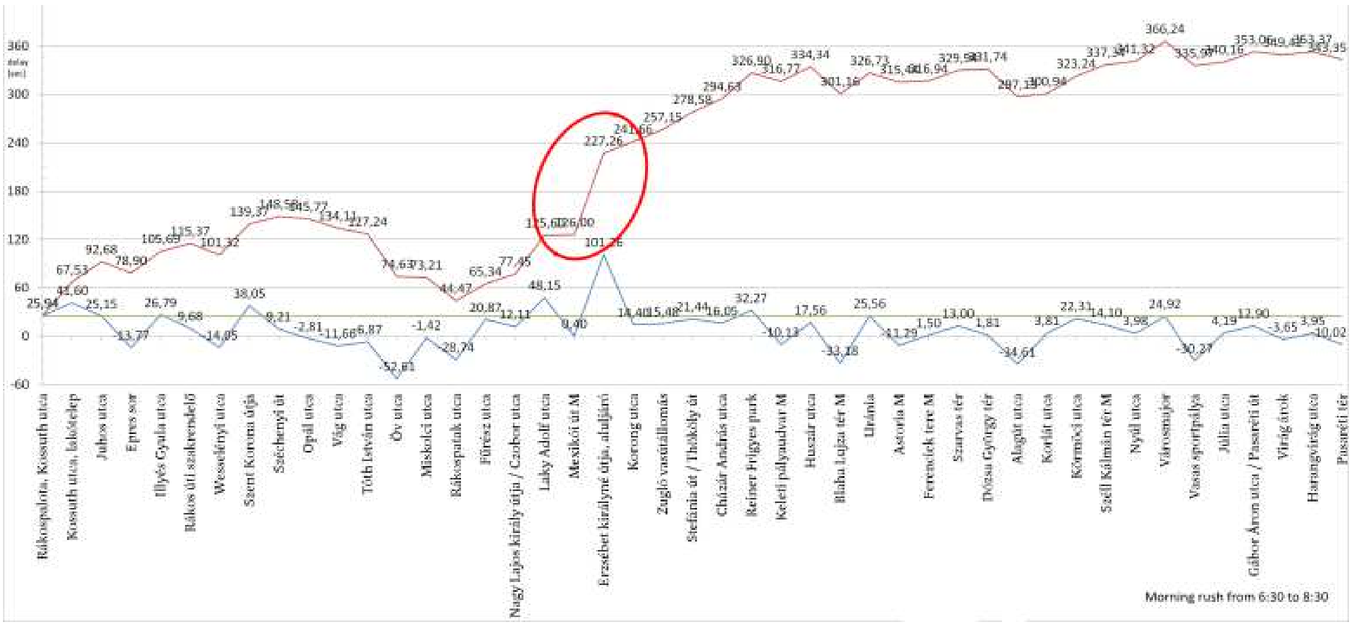

Figure 4. *The most noticable spike.*

After some investigation it turned out to be the caused by the construction that has been taking place in this area, and it only affected a single day, and for some strange reason, only in this direction.

Moving on with the analysis, I've focused on the smaller increases in delay that emerged, circled with orange on Fig. 5.

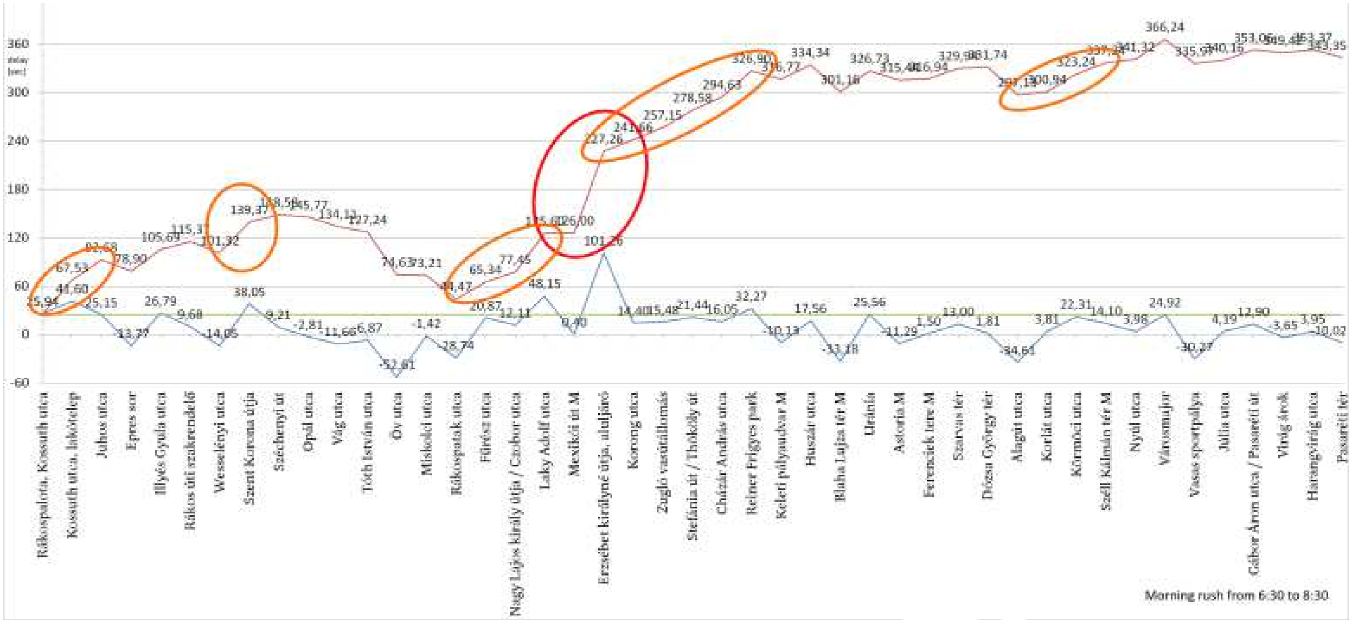

Figure 5. *Adittional increases*

There is a long streak of growing delays on the stops on Thököly út, and some smaller ones elsewhere. These stops cover about half of the route's stops! That means that either the timetable needs some significant corrections for rush hour travel times, or BKK needs to come up with a way to prevent buses from being late. There's already a bus lane on the Rákóczi út, the longest streak, so there's not much we can do there to speed up buses other than to fine-tune traffic lights, but the other places of interest should get some attention from the agency.

But buses not only gather delays, they can also work them off. The green circle on Fig. 6 shows an area like this, where the timetable have some spare time, so buses running late can catch up to their designated time.

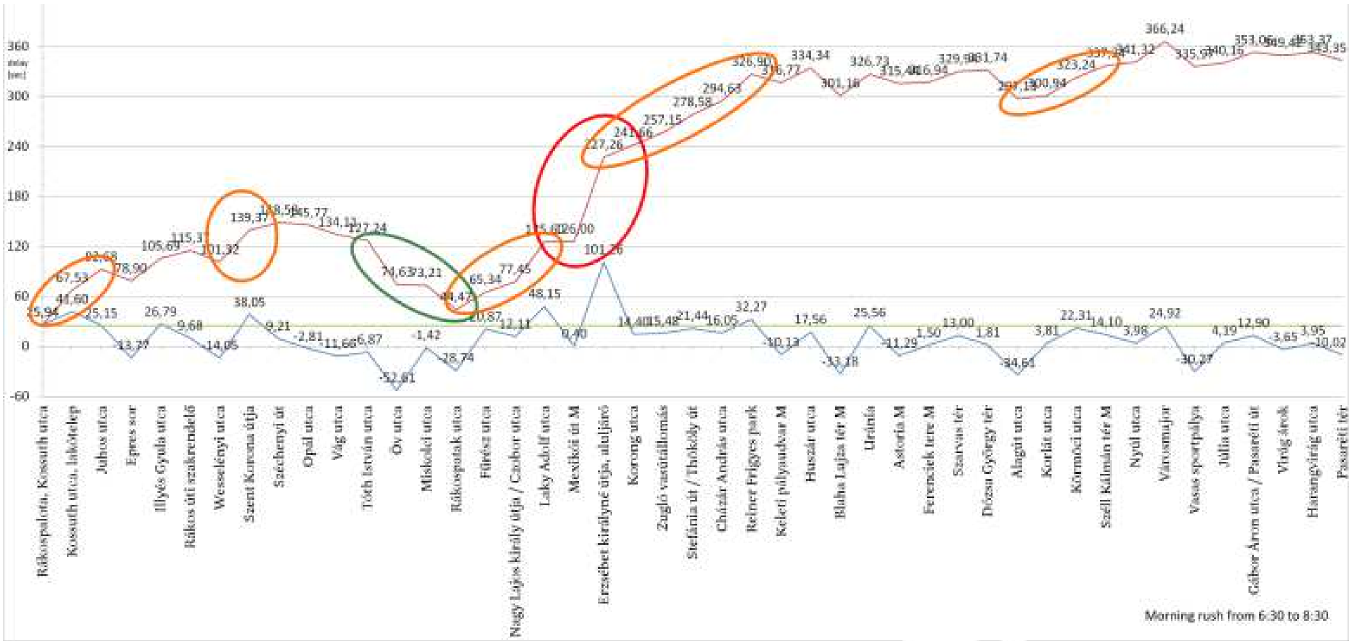

Figure 6. *Catching up.*

Finally, there are small anomalies like the one in the blue circle on Fig. 7, where the bus gathers some extra delay in one stop, but loses it in the next. Places like this exist because not every stop takes exactly a minute or two, but timetables can only use minute precision. When two stops take 90 seconds each, and the timetable gives 60 and 120 seconds for them, we get anomalies like this.

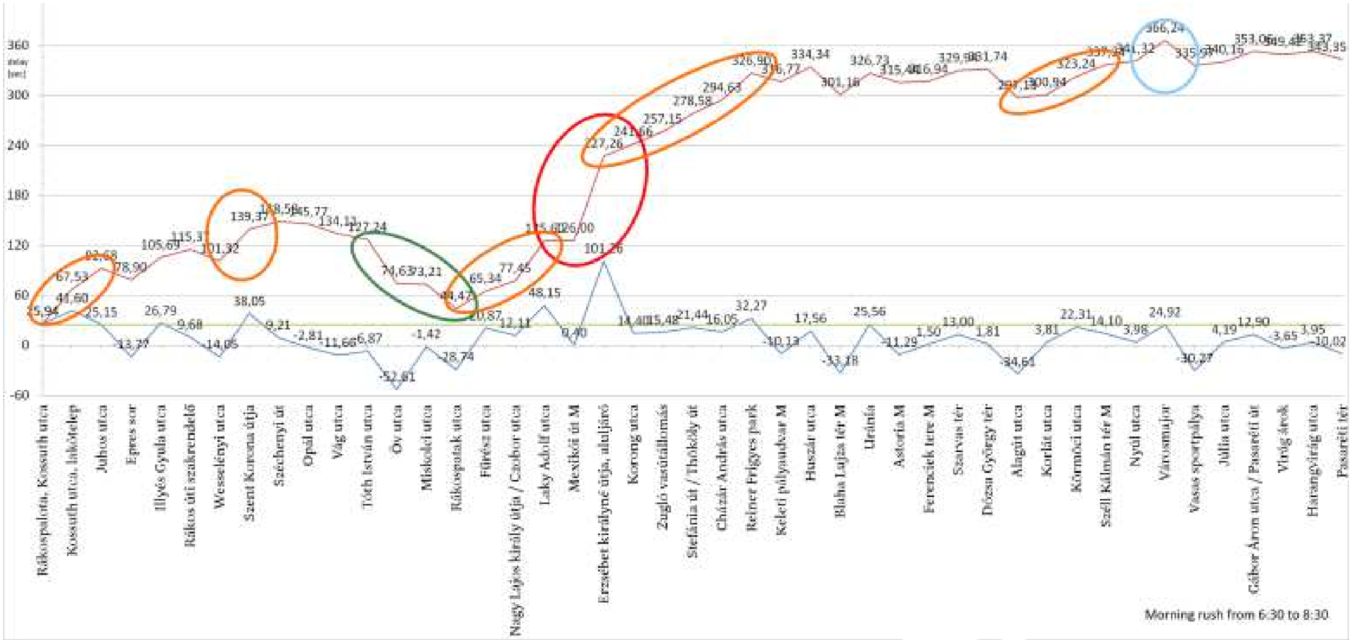

***Figure 7.** Small anomalies.*

6. Using this Knowledge

The analysis showed us that the places where delays occur can be found using the data recorded by the transport agency. Developing automated methods to find these anomalies strike one as a worthwhile effort. The application of such methods can reveal significant reserves in the public transport system. As people generally dislike uncertainty, making public transport vehicles more accurate or providing passengers accurate information about arrival times can help make public transport more predictable, and thus more appealing. [6][7][8]

However, public transport is not the only one that can profit from such improvements. A more appealing public transport can change the modal split large enough to noticeably reduce private car usage, which in turn will make both public and private transportation faster, potentially making public transport even more appealing. As Enrique Peñalosa said [10] in his TED talk, "an advanced city is not one where even the poor use cars, but rather one where even the rich use public transport."

Paying attention to small things like delays and uncertainties in public transport can help us make public services better, and eventually make our cities more livable.

To put these ideas into practice, transport agencies will need to implement passenger information systems where vehicle arrival data is calculated by taking past behavior of these vehicles into account, and deliver this information to passengers in the right time, place and form.

7. Conclusion

Budapest Transport Centre (BKK) has implemented a vehicle tracking system on most of its fleet. This provides data real time, which can be used to inform passengers and affected employees. By using past data, we can make timetables and arrival time predictions more precise, identify problematic areas, and make changes in the city's traffic system more effectively.

Putting data into context can yield us information, and by presenting that information to people in a timely manner they can draw conclusions from, and give meaning to it. By doing this, they acquire knowledge, and with some insight, it can become wisdom, that can help them make competent decisions.

By utilizing these possibilities we can create public transport that is more accurate, and adapts to the ever-changing environment of a busy city.

Making public transport vehicles more accurate or providing passengers accurate information about arrival times can help make public transport more predictable, and thus more appealing, and can change the modal split large enough to noticeably reduce car usage, which in turn will make both public and private transportation faster.

Summary

Transforming data into information, and converting that information into useable knowledge became one of the most difficult challenges of information society. Most organizations gather large amounts of data, without a vision on how to use them.

Budapest Transport Centre (BKK) has installed the FUTÁR system on all of its buses, trams, trolleybuses and ships, which gathers temporal and spatial data on them, and every 30 seconds, the on-board unit sends them to the dispatch center via a GSM/3G datalink.

Currently BKK only use these datasets internally, and the general public can access the processed data through BKK's website. Through the 3rd party website web-schedules.rhcloud.com, which acquires its data from the

official online site, we can access this data a bit easier. This website's pages are easily convertible to raw data, as it uses a plain HTML format. By using simple text-processing methods I was able to convert these HTML pages to CSV files.

However it is quite hard to extend the scope of the analysis to a lot of lines, as it takes a considerable amount of time to download all the trips in a day. It has another downside, namely the fact that we can only get data from the current day, not the ones before.

I had to make a choice, and select a single route I wanted to analyze. This route was bus line 5. Reasons for choosing it included its almost 20 km length, its diverse route featuring various parts of the town, and the fact that I've been using it since its extension in 2008.

This dataset was then imported to Excel, and after some processing there, the resulting spreadsheet was then imported to an Access database. Running some queries on the data, grouping them by direction and time of day: dawn, morning rush, midday, evening rush, and night gave me data that can be best illustrated with graphs.

Analysis of these graph showed the problematic areas of the route, where a small intervention can give us significant benefits. I've concluded that developing automated methods to find anomalies like these are worthwhile as they can reveal previously unseen potentials in public transport systems.

Unlocking these hidden potentials in public transport can attract car users to put down their cars, which in turn can make both public and private transportation faster, saving time for everyone, and making our cities more livable.

To put these ideas into practice, transport agencies around the world should implement passenger information systems as described, and deliver useful pieces of information to their passengers.

References

[1] Ackoff, R. L., "From Data to Wisdom", Journal of Applies Systems Analysis, Volume 16, 1989 p 3-9.

[2] G. Bellinger, D. Castro, A. Mills, "Data, Information, Knowledge, and Wisdom", 2004 http://www.systems-thinking.org/dikw/dikw.htm

[3] A. A. Nunes, T. Galvão, J. F. e Cunha, "Urban public transport service co-creation: leveraging passenger's knowledge to enhance travel experience", Procedia - Social and Behavioral Sciences, 111, 2014, pp. 577–585.

[4] F. Filippi, G. Fusco, U. Nanni, "User empowerment and advanced public transport solutions", Procedia - Social and Behavioral Sciences, 87, 2013, pp. 3 – 17.

[5] S. Pensa, E. Masala, M. Arnone, A. Rosa, "Planning local public transport: a visual support to decision-making", Procedia - Social and Behavioral Sciences, 111, 2014, pp. 596 – 603.

[6] D.-T. Le-Klähn, R. Gerike, C. M. Hall, "Visitor users vs. non-users of public transport: The case of Munich, Germany", Journal of Destination Marketing & Management, 3, 2014, pp. 152–161.

[7] T.M. e Sousa Gonçalves, "A smartphone application prototype for exchanging valuable real time public transport information among travellers", Faculdade De Engenharia Da Universidade Do Porto, 2012

[8] B. Ferris, K.E.Watkins, A. Borning, "OneBusAway: Behavioral and Satisfaction Changes Resulting from Providing Real-Time Arrival Information for Public Transit", 90th Annual Meeting of the Transportation Research Board, Waschington, D.C., Aug. 2006, p. 18.

[9] P. Newman, "Why do we need a good public transport system?", PB-CUSP Parsons Brinckerhoff, NSW Government, 2013, Sydney's Ferry Future, NSW Government, http://www.curtin.edu.au/research/cusp/local/docs/pb-cusp-research-paper-section-abcd.pdf (accessed: 30 Nov 2014)

[10] Peñalosa, E. (2013). "Why buses represent democracy in action". Downloaded from: http://www.ted.com/talks/enrique_penalosa_why_buses_represent_democracy_in_action (accessed: 30 Nov 2014)

[11] Zoltayné Paprika Zita "Döntéselmélet", Aliena kiadó, Budapest, 2005

[12] Miller, George A.: "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", The Psychological Review, 1956, vol. 63, pp. 81-97

[13] Polányi, Michael, "The Tacit Dimension". Garden City: Dou-bleday and Company, 1966

[14] Nonaka, Ikujiro; Takeuchi, Hirotaka (1995). "The knowledge creating company: how Japanese companies create the dynamics of innovation", New York: Oxford University Press. p. 284.

[15] Simmel, G. (1950). "The metropolis and mental life", In K. H. Wolff (Ed.), "The sociology of Georg Simmel",.New York: Free Press.

[16] S. March, A. Hevner, S. Ram (2000). "Research Commentary: An Agenda for Information Technology Research in Heterogeneous and Distributed Environments", Informs Pubs Online. http://dx.doi.org/10.1287/ isre.11.4.327.11873 (accessed: 30 Nov 2014)