

Teemo Tebest

# Open Data – Fuel for Data Journalism

## *On incentives and challenges of using open data*

Open data, created by public authors, private companies or the third-sector actors, is the fuel for data journalism. In the Nordic countries, legislation provides journalists and citizens with a good access to data created by the public sector. A journalist can ask for public data and, in most cases, the authorities release the data as requested. In this way, open government forms the basis for the existence of open data – and for data journalism. In this article, I will demonstrate the uses of open data in journalism created at the data journalism unit Plus Desk at the Finnish Broadcasting Company Yle.

Data journalism by definition is journalism that specifically makes use of data.<sup>1</sup> Data is considered to be raw, and by itself doesn't tell a story. Instead, the story is created using journalistic means and in close collaborative work of a reporter, a coder and a graphic designer.

Open data<sup>2</sup> can be, and is often, published in journalistic platforms as is; unemployment rates and export statistics may be interesting as such. However, open data shows its true force only when the data is analyzed and combined with other data sources. In this way, open data allows journalists and active citizens to create new information and to understand the world better.

### Advantages of opening the data

In Finland, authorities have actively opened data sets since 2012.

---

**Teemo Tebest**, a data journalist and social media web developer who has worked at the Finnish Broadcasting Company Yle since 2012. A full-time data journalist since 2013. Previously he worked as a researcher and a postgraduate student at the Tampere University of Technology, where he studied information visualisation.

The first actions were to open up map and weather data.<sup>3</sup> This was a Ministry of Transport and Communications initiative, which was later strongly supported by the Ministry of Finance.

Also, third-sector actors like the national section of the international Open Knowledge International, Open Knowledge Finland, played a significant role in supporting this initiative. Open data was seen among other things as a possible source of support for small and mid-sized businesses. Today, there are over 1,300 open data sources available in the Finnish Open Data Portal.<sup>4</sup>

Before these developments, the main challenge in opening data was a lack of resources. Without good examples, the positives of open data, such as the direct economic benefits, were hard to point out. Nowadays, the benefits of opening data are more widely recognised.

According to the European Data Portal, open data enhances the performance of public services, benefits the private economy and improves social welfare systems.<sup>5</sup> For example, according to the Mayor of Helsinki, open budget data has led to significant savings in the city's budget.<sup>6</sup>

## “Thunderbird”, “Thundebird”, “Thunderbird”

Opening data requires effort, and data is considered open only if it is usable. According to the Open Knowledge Foundation, the key features of openness are availability, reusability and universal participation.<sup>7</sup> Making data usable requires cleaning the data, validating the data, describing the data, keeping the data up-to-date and making the data available. Cleaning the data may include fixing misspellings, removing duplicate entries or handling synonyms. In 2013, when the Finnish Transport Safety Agency released their automobile registry, the data included a car model called Ford Thunderbird written in at least 10 different ways, which made the data in many ways unusable.

Data should also be machine readable. For example, PDF documents are rarely considered as open data. Furthermore, legal issues such as personal privacy have to be taken into account when opening data. Licensing the data is vital, as this defines how the data can be reused. In 2016, to support these matters, the Ministry of Finance, the Ministry of Education and Culture, the Association of Finnish Local and Regional Authorities and the Finnish Consulting Group organised a seminar at the Aalto University specially tailored for Finnish public actors interested in opening data.<sup>8</sup>

Data publishers should also work in a close connection with the users. Open data users may include journalists, experts and active citizens.

There are a number of questions to be discussed together: What kind of data is relevant for the users? Which file formats should be used? Which coding languages should be supported? Is a full-scale API required, or is an Excel file the best way to deliver the data? In a collaboration with the users, the data publishers can become more aware of the benefits associated with their open data, which, then, helps to adapt and justify the efforts made.

Open data should also be prominently displayed and promoted so that users will be able to find it. Open data is not the end in itself but it is a way to make things better for the users.

## From open data to the most-read stories at Yle

At Yle, where I am employed at, we have a team of eight staff members who work in the field of data journalism at our Plus Desk unit (<https://plus.yle.fi>). We have three coders, three graphic designers and two producers in the team. The producers' job is to co-ordinate our work with other news desks.

For every data project, we try to form a team consisting of a reporter, a coder and a graphic designer. For example, if we want to cover budget data we would team up with someone from the economic desk. The reporter is responsible for the content as a whole, the coder tackles the technical and statistical issues and the graphic designer decides how to present the data. Our team is involved in about 150 stories each year.

In our model, in which professionals from different areas work together with data, it is important to define a common goal. Sketching ideas has proven to be very efficient. We also try to meet as often as needed and, whenever possible, work in the same physical space. It is important that everyone knows their role and the status of the project.

Our projects vary in scale, but what is common to the most data stories is that they are very popular among the readers. People tend to enjoy stories where they can obtain personalised and detailed information that relates to them. Data stories are also often visually impressive.

For example, the story we published in 2013, based on the inconsistent automobile data obtained from the Finnish Transport Safety Agency was one of the most-read stories that year.<sup>9</sup> Although it required a lot of effort to clean the data, in the end we converted the Excel file with 6,700 rows into a story and a news application. In our application, one could input their own vehicle's information and get out more information related to it.

More recently, in 2018 we published a story with the goal to help people applying to universities where we combined data from the Finnish National Agency for Education and Statistics Finland to rank examinations.<sup>10</sup>

The screenshot shows the 'Kommunradarn' (Municipality Radar) interface. At the top, the URL is 'svenska.yle.fi'. The main heading is 'Kommunradarn' with a sub-heading: 'Hur mår din kommun, vilka är framtidsutsikterna och hur går det för grannarna? Vi tog fram uppgifterna, nu kan du forska i dem. Nu med mera valinformation och uppdaterad statistik!'. Below this is a search bar with 't.ex. Raseborg' and a 'Välj' button. A section 'DELA RESULTATEN: PÅ FACEBOOK, PÅ TWITTER ELLER SOM LÄNK.' is visible. On the right, a list of categories is shown with five-star ratings: Livskraft (5 stars), Ekonomi (5 stars), Hälsa (4 stars), and Boendeklimat (4 stars). Below this is a 'Valfrågor' section with the text 'SKOLNÄT, ENTREPRENAD, SKULD TAGNING, MÖGELSKOLOR' and a paragraph: 'I fjol lyckades Borgå vända ekonomin på plus. Samtidigt finns en lång rad bygg- och reparationsprojekt inplanerade för totalt 35 miljoner euro. Hur det här ska finansieras är en stor fråga men dessutom får fullmäktige ta ställning till skolnätet och om social- och...'

The Municipality Radar (2017) was one of the largest open data projects created by the Yle Plus Desk. The goal was to create a service that provides statistics and data on each of the 297 Finnish municipalities.

This allowed people to determine the kinds of education programmes more likely to lead into unemployment and what levels of salary people could expect to earn for each examination passed. This story was especially useful in reaching younger audiences, which is the goal for our company.

Often, we also convert data sets into a calculator format where users can input their own information. One of the best examples of this was the Retirement Age Calculator, published in 2016, where we used data related to the retirement age for each year of birth.<sup>11</sup> Our implementation calculated the retirement age for the user-entered year of birth and birth month. This was the most-read story of the year, and remains among the most-shared Facebook content in Finland.

One of the largest open data projects we

have worked on in recent years was the Municipality Radar.<sup>12</sup> The goal of the Municipality Radar was to create a complete view of the status of each of the 297 Finnish municipalities from a data point of view.

We made use of various open data sources, including the Statistics Finland, the National Institute for Health and Welfare and LocalFinland, to gather thirty-nine data variables and combine them into four categories, which we presented through a five-star rating.

The outcome was a story where the user entered the municipality in which he or she was interested, to see how the municipality was performing according to 1) the future of the municipality, 2) the municipality's economy, 3) the health of the citizens and 4) the atmosphere in the municipality. The data was visualised in a way that was clear to under-

NYHETER

LOGGA IN SÖK MENY

UTRIKES SPORT KULTUR HUVUDSTADSREGIONEN VÄSTNYLAND ÅBOLAND ÖSTERBOTTEN ÖSTNYLAND

## Drömmer du redan om pensionen? Kolla när du får pensionera dig

© PUBLICERAD 27.12.2016 - 14:12. UPPDATERAD 30.03.2017 - 10:44

DELA: 5696

Bild: Yle Nyhetsgrafik

**Vid årsskiftet 2016-2017 förändras pensionssystemet. För de flesta av oss höjs pensionsåldern. Målet är att förhållandet mellan den tid man tillbringar i arbetslivet och tiden som pensionär ska förbli det samma också i framtiden.**

### Pensionsräknare

Födelseår

Födelsemånad

**Du får gå i ålderspension om 33 år och 4 månader**

Dela på [Facebook](#) eller på [Twitter](#).

Som tidigast börjar din pension i september 2051. Du är då 66 år och 11 månader. Din förväntade livslängd efter pensioneringen är då 22 år och 9 månader.

Fullständig pension får du i juli 2054 då du är 69 år och 9 månader. Då förväntas du fortsätta leva i 20 år 7 månader. Uppgifterna kommer från Pensionsmyndigheten, individuella skillnader kan förekomma.

*"Du har 12 167 dagar, 21 timmar, 10 minuter och 50 sekunder kvar till pensionen."*

Calculators are an efficient way to simplify data sets. The Retirement Age Calculator (2016) allows the user to calculate his or her estimated retirement age.

stand and allowed for easy comparisons. It was also possible to obtain the detailed data on demand. Based on the analysis, textual content was provided for each municipality to explain their status further.

Behind this the idea was that, often, only one data set is looked into, and that single point of view may not tell the full story. The Municipality Radar made it possible to look into municipalities' matters in a deeper but simpler way through using open data. We also released the data that we gathered and explained the process of how the analysis was conducted so that anyone could replicate or improve it.

The Municipality Radar was adopted by

citizens, professionals and municipality leaders; our readers spent over 10,000 hours online using it and reading related stories. It was also referred to by other media companies several times, and it was a nominee for the Nordic Data Journalism Award.<sup>13</sup> The service made municipalities' matters relevant and accessible at a time where municipality elections were imminent.

### “What is the meaning of nothing?”

There are, however, a number of challenges associated with using and combining data from

## Open Data Sources in Finland

- **Ministry of Finance:** <http://vm.fi/avoin-tieto>
- **Finnish Meteorological Institute:** <https://ilmatieteenlaitos.fi/avoin-data>
- **Open Data and Interoperability Tools**, a service for standards and guidelines: <https://www.avoindata.fi/fi>
- **The Association of Finnish Local and Regional Authorities:** <https://www.kuntaliitto.fi/tilastot-ja-julkaisut>
- **National Land Survey of Finland:** <https://www.maanmittauslaitos.fi/asioi-verkossa/avoimien-aineistojen-tiedostopalvelu>
- **Statistics Finland:** <https://www.stat.fi/org/avoindata/pxweb.html>
- **Finnish Transport Safety Agency:** [https://www.trafi.fi/tietopalvelut/avoin\\_data](https://www.trafi.fi/tietopalvelut/avoin_data)

different sources. Although a single data set may look clean and simple, often a lot of effort is required before data sets can be combined. The challenges include, among other things, handling empty and missing values and figuring out what zeroes mean. There are no standardised ways to present anonymised or missing data.

When gathering data sets for the Municipality Radar, one of the most common issues encountered was the problem of an empty cell used to present at least anonymised data, missing data and a zero. In some data sets “-” and “-1” were used for similar purposes. Differing statistical years were also a huge issue when combining data sets for the Municipality Radar when the number of municipalities varies from year to year. When combining dozens of data sources, it is vital that the data syntax is coherent.

Ultimately, you have to know your data and validate it, and there is no quick fix for this. Each data set has its tricks, but the basic functions for each data set are the minimum and maximum values. Sometimes statistical knowledge like standard variation is required.

Journalists should avoid becoming blind to the data. If something looks obscure, one should recheck, recheck and recheck because most of the time there is an error somewhere, either in the data or in your analysis. Common sense and journalistic doubt are required

when dealing with data. It is important to remember that data publishers can also have agendas that they want to advance, which can collide with journalism.

Online societies, such as the Facebook groups “The Finnish Open Data Ecosystem” and “Datajournalistik”, are good sources of support. There is also literature available, including the *Data Journalism Handbook*.<sup>14</sup> I myself have opened up the processes behind our data stories in my personal blog that can be read at <http://datajournalismi.blogspot.fi>.

## With open data comes great responsibility

Openness benefits society in various ways, and we journalists act as intermediaries that work to make the open data accessible. Thanks to our efforts, it can be said that open data leads to information that helps us know the world better.

New business opportunities appear when data is available and not restricted to only the few, for example when journalists can more easily point out misuses of taxpayers’ money. Data also allows for the creation of more personalised services that benefit users, and for telling stories that are more meaningful from the readers’ point of view.

However, with benefits comes responsibility, as open data can also be misused. For

this reason, open data publishers have great responsibility to define what their data is and what it is not. In addition, how the data is formed and what is left out, which is referred to as metadata, is vital for understanding and using open data. This responsibility is then transferred to the data user, whose job is to ensure they do not draw false conclusions. In most cases, open data publishers in Finland understand the importance of metadata.

The hype around data and data-driven decision making is slowly starting to show real benefits.

However, the ethical, technical and practical challenges have to be discussed in public by the data openers and the users of the data. Some public data sets remain unopened, and in journalism we are always interested and keen to get our hands on a greater amount of accurate data.

## Notes

1. Wikipedia (2018a).
2. Wikipedia (2018b).
3. Lehto (2012); Parikka (2012); Linnake (2012).
4. Avoindata.fi (2018).
5. European Data Portal (2018).
6. Hänninen (2014).
7. Open Knowledge International (2018).
8. Aalto University (2016).
9. Yle (2013).
10. Yle (2018).
11. Yle (2016).
12. Yle (2017).
13. NODA (2017).
14. Gray et al. (2018).

## References

- Aalto University (2016). Open Data and Open Access Publishing. *News & Events* 29 March 2016. [Online.] Available at <<http://www.aalto.fi/en/current/events/2016-03-17/>>. Accessed 15 April 2018.
- Avoindata.fi (2018). *Open Data and Interoperability Tools*. [Online.] Available at <<https://www.avoindata.fi/en/>>. Accessed 15 April 2018.
- European Data Portal (2018). Benefits of Open Data. [Online.] Available at <<https://www.europeandataportal.eu/en/using-data/benefits-of-open-data>>. Accessed 15 April 2018.
- Gray, Jonathan; Bounegru, Liliana & Chambers, Lucy (eds.) (2018). *Data Journalism Handbook*. [Online.] Available at <<http://datajournalismhandbook.org/1.0/en/>>. Accessed 14 April 2018.
- Hänninen, Jyri (2014). Pajunen uskoo rahaliikenteen avoimuuteen. *Helsingin Sanomat* 29 November 2014. [Online.] Available at <<https://www.hs.fi/kaupunki/art-2000002781549.html>>. Accessed 15 April 2018.
- Lehto, Tero (2012). Lisää dataa avataan – nyt Ilmatieteen laitokselta. *Tivi* 8 August 2018. [Online.] Available at <<https://www.tivi.fi/Arkisto/2012-08-08/Lisaa-dataa-avataan---nyt-Ilmatieteen-laitokselta-3144451.html>>. Accessed 15 April 2018.
- Linnake, Tuomas (2012). Startup-yhteisö palkittiin vuoden it-tuotteena. *Yle Uutiset* 21 March 2012. [Online.] Available at <<https://yle.fi/uutiset/3-6323540>>. Accessed 15 April 2018.
- NODA, Nordic Data Journalism Base (2017). Community Radar. *Nordic Data Journalism Base* 2 January 2017. [Online.] Available at <<http://nodabase.net/cases/community-radar/>>. Accessed 15 April 2018.
- Open Knowledge International (2018). *Why Open Data?* [Online.] Available at <<https://okfn.org/opendata/why-open-data/>>. Accessed 15 April 2018.
- Parikka, Outi (2012). Urpilainen: Yle voisi avata koko arkistonsa ilmaiseen käyttöön. *Yle Uutiset* 5 October 2012. [Online.] Available at <<https://yle.fi/uutiset/3-6323540>>. Accessed 15 April 2018.
- Yle (2013). Näin moni ajaa vanhemmalla autolla kuin sinä. *Yle Uutiset* 16 October 2013. [Online.] Available at <<https://yle.fi/uutiset/3-6880326>>. Accessed 15 April 2018.
- Yle (2016). Haaveiletko eläkkeestä? Katso laskurista, milloin eläke alkaa ja kauanko se kestää. *Yle Uutiset* 15 November 2016. [Online.] Available at <<https://yle.fi/uutiset/3-9273402>>.
- Yle (2017). Kuntatutka. [Online.] Available at <<https://yle.fi/uutiset/3-9490180>>. Accessed 15 April 2018.
- Yle (2018). Yli 47 000 opiskelupaikkaa tarjolla – Ylen korkeakoulukone kertoo, mille alalle on tiukin seula ja mistä valmistutaan kovapalkkaiseksi. *Yle Uutiset* 17 March 2018. <<https://yle.fi/uutiset/3-10116390>>. Accessed 15 April 2018.
- Wikipedia (2008a). Data Journalism. [Online.] Available at <[https://en.wikipedia.org/wiki/Data\\_journalism](https://en.wikipedia.org/wiki/Data_journalism)>. Accessed 15 April 2018.
- Wikipedia (2008b). Open Data. [Online.] Available at <[https://en.wikipedia.org/wiki/Open\\_data](https://en.wikipedia.org/wiki/Open_data)>. Accessed 15 April 2018.