

Ethics of data journalism

Four ethical phases in the working process

Heikki Kuutti

PhD. research coordinator, journalism

Department of Languages and Communication

University of Jyväskylä, Finland

A paper presented at Nordic Data Journalism Conference NODA 2016, 21 April 2016 in Helsinki, Finland

Summary

Data journalism has emerged remarkably during the last years. Discussions of data journalism have mostly concentrated on access to data, preparing a data story and providing truthful information based on data. However, ethical requirements and responsibilities of data journalism have not been widely discussed. For instance, press councils do not particularly discuss data and its impact on journalism ethics. Data when mentioned in the codes usually refer to information and its accuracy requirements to be taken into consideration in information gathering and publishing.

This paper analyses ethical issues of data in conventional journalistic working process starting at access to data and followed by data analysis, publishing a data story and collecting feedback from the public. Each of the four phases contains several important ethical points of attention. Phases have also consequences to subsequent phases.

Keywords: journalism, data, journalism ethics, editorial working process

Introduction

Data journalism is journalism grounded on digital data material. In contrast to traditional journalism, data journalism contains a different and particular set of working processes such as data inspection and cleaning, data analysis and inspection of data analysis, which follow data gathering and access and precede story publishing. The ethical aspects in data journalism relate to these parts of working process and data based materials and their subsequent processes about access to data as a part of information requests (tinted boxes in the Figure 1).

In spite of its digital character, data journalism ethics do not relate to technical issues but journalists' choices in their work. Data is just a tool in journalism and does not include ethical content. The main issue is the ability of journalists to examine news topics by means of data and their expertise to interpret data and to consider the importance of the information they have gathered (Uskali & Kuutti 2016:93).

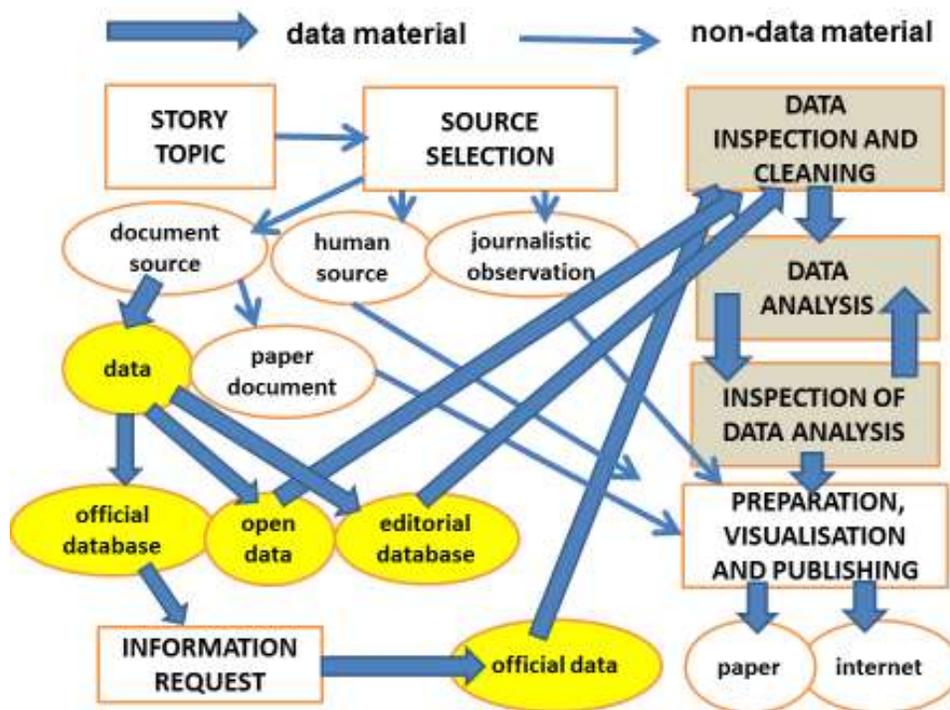


Figure 1. Data journalism in the journalistic working process. Rectangles describe working processes and yellow ovals show their data based materials. Tinted parts relate to data journalism analysis activities. (Uskali & Kuutti 2016:93)

During the last few years, advances in information and computer technologies have strongly influenced journalistic practices and emphasized the importance of data journalism. Accordingly, data journalism has established its position as an important seminar topic in several conferences (*Nordic Data, National Institute for Computer Assisted Reporting, Data driven journalism, Data harvest, Global Investigative journalism conference*). However, only a few journalists use data in their everyday work. In many cases, journalists shun technological features in data journalism and therefore are not very familiar with the use of data. (ibid)

Actually, data journalism may contain mysterious aspects for less technologically advanced journalists. Data journalism has certain reflections with the *deus ex machina* (“god from the machine”) phenomenon where an external actor solves, (both unaccountably and unexpectedly), people’s problems. Journalists with insufficient knowledge of technological issues may be surprised by the results of their analysis. These are caused by algorithms and computer runs and therefore beyond their “personal reach”.

However, data journalism practices from the ethical viewpoint are not discussed very often. For instance, *Data Journalism Handbook* (Gray et al. 2012) concentrates on preparing a data story and does not discuss data journalism ethics. Data usually refer to information and its accuracy requirements, which journalists should take into consideration in information gathering and publishing. Ethically, it is more difficult to assess the use of data in editorial activities as a whole. For instance according to the ethical guidelines of *the Council for Mass Media in Finland* main ethical concerns relate to journalists’ professional status, obtaining and publishing information, the rights of interviewees and interviewees, corrections and the rights of reply, and private and public issues. According to the codes “journalist must aim to provide truthful information” but they do not make any distinction between data and other forms of information. (CMM 2014)

It is difficult to separate the importance of data from the ethics in data journalism as a whole and compared with other journalistic activities in producing a data story. Ethical problems in a data story are difficult to discern, unlike misspelled interviewee’s quotes or obviously false information in a story. Whenever suspicion of a journalist’s analysis or interpretations is awakened, the search to corroborate incorrect information in a data story is a time-consuming exercise, which also requires the ability to reanalyse the original data.

Ethical aspects of data journalism

Reliability of data journalism relates to the reliability of the data. According to Bradshaw (2013) ethical considerations in data journalism do not differ significantly from those in any other area of journalism. In principle, data journalism encounters the same ethical problems as traditional journalism regarding, for instance, endangering privacy and untruthful presentation. However, information technology used in this kind of journalistic practice has created additional concerns relating to access to data material, use of data in analyse, transparency in publishing the results and openness to get public feedback and correction of mistakes in a story.

Journalist may ask irrelevant questions of data, analyse incorrect data or leave out important data material in their stories. Problems may arise if data is missing, analysis methods are insufficient, conclusions are not justified, or the overall picture of the issue is not broad enough. However, errors are to get out if editorial staff studies and discusses the source data and working methods and if the data and analysis used are made transparent in a story for the public. (ibid.)

Ethically, one important point of attention in data journalism is journalists' attitude to their own work and the way they should conduct their profession and maintain their watchdog role. Data is a convenient tool to "find out the truth" which is considered one of the most important working ethos of journalism.

Actually, the format of *data-based information* opens up new perspectives and enables journalists a kind of *birds-eye-view* on story topics. Information created this way gives journalists "new horizons" in their work. Data journalism approach enables journalists to build new kinds of entities and to see the broader patterns of the covered topics. When using data journalism practices, journalists are able to uncover the connections and reasons behind seemingly separate issues not exposed by only individual data. Data analysis as such also redirects more in-depth information gathering to the issues that analysis results have awoken.

Journalists' dependence on authorities as a source of information decreases by using public data while their own expertise about the topic increases. Public service institutions lose their previous exclusive right to data material interpretations and their monopoly producing statistics and reports of their own activities.

In this paper, ethical problems of data journalism are divided into four phases in the journalistic working process: access to data, data analysis, publishing a data

story and asking the public for feedback. Each of these phases contains several ethical point-of-views, some of which are common and have their own implementation during the working process. For instance, dirty data in the access phase refers to the content of data and data sets, in the analysis phase possible errors in analysing data material and in the publishing phase ethical claims to inform the public about possible problems caused by dirty data. The same variations are included in a journalist's interpretation depending of its object: what does dirt in data mean in the contexts of access, analysis or publishing. (Figure 2)

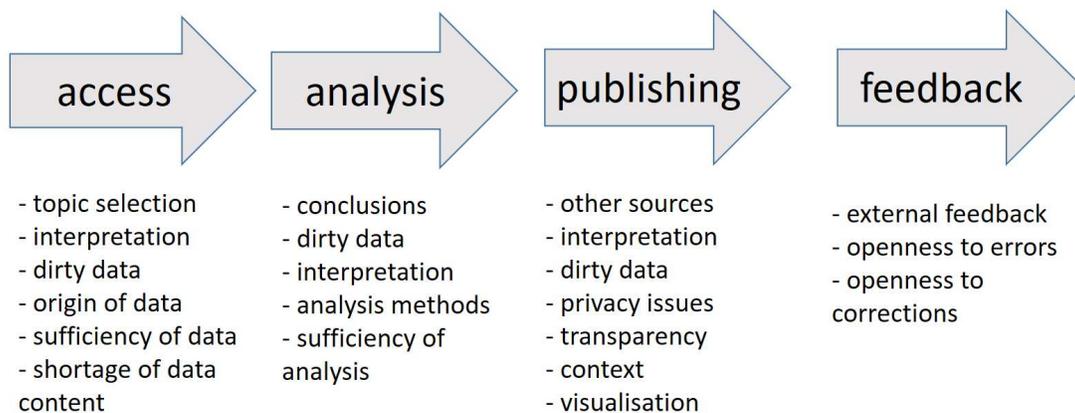


FIGURE 2. Ethical phases of access, analysis, publishing and feedback in data journalism working process. Each of the four phases require specific ethical considerations.

Access to data

As more and more government records are produced digitally, it becomes ever more important how they are used and the context, in which the data is given. In principle, digital data as a kind of documentary record is relatively reliable in journalism. Unlike human sources, data do not change, forget or incorrectly recall information having reported in the news. Data is journalistically sustainable stored in databases as records of (real-time) events and decisions. Data as a piece of public information is also official. Moreover, data do not contain deceptive PR-dimensions having not created specifically for journalistic purposes.

However, technical characteristics of digital data may cause several ethical problems. Ethical issues in data journalism relate mostly to the accountability of data and data sets. In examining data sets, it is important to notice the origin, sufficiency and shortage of data. Journalists should ask data the same questions as any other source and evaluate data ethics in several viewpoints: whether all the data needed in a story has been found, how valid and sufficient is the analysing method, how justified are the conclusions, is the over-all picture broad enough and is the context accurate. (Bradshaw 2016)

Bradshaw (ibid.) emphasises journalists to ask the following questions:

- What is the vested interest of the person providing data?
- How has this information been collected, and what or who (or when or where) might be missing from it?
- How were the questions phrased, and what questions were used to frame the data beforehand?
- Is there a second independent source of the same information, or a different interpretation?
- What is the margin of error?
- Does journalist have the knowledge to ask the right questions of all these sources

LaFleur (2014) advises to check the integrity of data with following steps: read the (original) documentation, check know how many records you should have, check counts and totals against reports and look for possible missing data, duplicates and internal problems.

Dirty data refers to several errors in data and is widely used expression among data journalists. Problems of dirty data usually have originated in storing, transferring and registering processes and resulted in disappearing or mingling connections between data fields and data sets. Moreover, errors in collecting and transmitting data may cause problems. Data manipulation in various operational phases is easy since those who register, store and transmit data are human beings. Connections between data fields and databases may vanish or get mixed-up or variables used in analysis are outdated.

Webster (2014) warns that if the data does not remain in its raw state, it may be modified and therefore contain programming errors. Dirt can be in data itself or in larger data sets. Data is dirty if it is incorrect, stored in a varying format or if it is out of date. Dirt in data sets refers to databases in which a relevant data or whole data fields may be missing, data are stored in incorrect fields or the same data are stored more than once. A good starting point is that all data is dirty. Accordingly, the last question is how dirty data actually is. Reasons for dirty data usually lie in

different motives of using databases. When authorities normally store data for their own varying pick-up purposes, journalists go through the whole data set to search news items or strengthen their story hypothesis.

Freedom of information (FOI) requests of open data differ from traditional FOI requests. Data requests refer to large information entities and the intention is to reuse the data. National political strategies aim to open up data to be used for varying purposes without specific regulation and therefore to foster the information industry with new applications.

However, open data is not a solution to data access since it is just one piece of all public data and published voluntarily. Even if it does not contain a clear PR dimension, owner authorities have checked and cleaned it. In addition, open data publishing may be a way to hide journalistically more important data material. Public data contain varied information, and authorities usually do not have exact understanding of its overall content. Therefore, they may be reluctant to release it. Usually authorities only collect data with no specific responsibility to analyse it. Data requests may raise fears among authorities about possible harmful results exposed by “unpredictable” analyses by unknown journalists.

According to a Finnish study (Kuutti 2011) several problems were exposed regarding the response behaviour of public service institutions (PSIs) to data requests. Even if the information is of a public nature, authorities may categorically deny the mass release of data. The main problem in data access seems to be the electronic format of information and the comprehensiveness of requests considered more complicated tasks than individual requests for information or documents. PSI's computer systems are not designed for servicing FOIs. Access tests by the Finnish study have revealed several defects in document administration and implementing information systems. When rejecting requests to release data, authorities refer to the inflexibility of the information systems. In many information requests for data the PSI conducts fee-attracting “special tasks” to prepare and edit the material containing both public and confidential data. According to several PSIs, it is not possible to separate confidential elements from the public elements, or separation can be conducted only by manually i.e. by time-consuming and expensive way. In order to be safe in ambiguous situations, PSIs prefer not to provide information. Even though it can be challenged, any refusal to surrender also public material is as such a legal interpretation and the right of the PSI concerned, whereas providing confidential material in a request can lead to criminal prosecutions (ibid.)

On the other hand, access to data may lead to its limitless use and invite journalists to take advantage of using data hastily and carelessly without considering restrictions of the data in a story. In addition, this kind of *data caused blindness* may marginalize information gathering only to databased materials and those (less important) story ideas where data is available.

Data analysis

Ethical problems in access to data also have their own consequences in data analysis. According to Bradshaw (2013), probably the most basic ethical consideration in data journalism is the need to be accurate, and provide proper context to the stories that journalists tell. That can influence how journalists analyse the data, report on data stories, or publish the data.

Accuracy is perhaps the central concern of journalists working with any form of data. Houston (2004) reminds journalists to be accurate and fair and not to prevent themselves from seeing only what they hope to see. That is possible since data journalism works with numbers which can be used as back up own interpretations. The only good story is one that summarises accurately what you have found.

Assumptions and findings are be challenged. Journalists using data need to be careful not to make false comparisons. For instance, LeFleur (2014) advises journalists to look for possible alternative results and ask what else could explain their findings, and check that all the necessary data are collected. It is journalists' responsibility to make certain that data used in a story actually tell what it should do and correspondingly do not tell what it cannot do. Cairo (2014, 25-27) points out, that even if a chosen variable does have a connection to the story, it may not be the best possible alternative to tell the story and therefore needs to be changed.

Scepticism is also required when journalists gather data themselves. Depending on how the results are used, gathering of data for instance by crowdsourcing may need a check on unusual patterns of behaviour, which may suggest that human sources are gaming the system. Also useful is a second check mechanism where some submissions are vetted by other users or even require some proof to be submitted alongside the data. (Bradshaw 2013)

O'Neil (2013) criticizes journalists for thinking that everything is measurable and that the variables they have chosen are viable for uncovering particular issues. By contrast, journalists are not necessarily technically oriented to discover the main

notion in data and its analysis. They do not know how to convert journalistically important questions into mathematical models.

It is a big challenge for the analogue journalist to be mathematically accurate and tech savvy to get what they want. As Otieno (2013) puts it, "every algorithm, however it is written, contains human, and therefore editorial, judgment. The decisions made about what data to include and exclude add a layer of perspective to the information provided".

Numbers, charts and maps possess an air of authority that other types of information often lack and yet they are equally subject to manipulation. Journalists need to be careful both in the credibility they place in numerical and graphical sources and in the way they present their own stories numerically and graphically. As an increasing proportion of journalists' sources involve data, *numeracy* becomes as important as literacy. The act of confusing "percentage increases" with "percentage point increases" should be as shameful as spelling someone's name wrong. Journalists should be as concrete in language regarding the data as they are concrete in describing events and people, where there is no room for vagueness or confusion. (Bradshaw 2016)

Data journalism also requires careful thinking about biases built into the original data and its interpretation. Inherent unconscious biases can affect decision-making, skew interpretation and create blind spots in data collection, analysis and its reception by the public. Journalists should develop an awareness of their own biases for instance by making a commitment to ethical work by knowing their own limitations and the limitations of the data. (ibid.).

However, hectic journalistic activities can spoil data analyses that require accuracy as Segnini (2015) points out. According to him, misuse of statistics can occur easily and risks increase when journalists attempt to draw causation from large or unexplained datasets.

As a sum, it is an incorrect hypothesis that data is always truthful or computer analysis does not make mistakes. Journalists should be as sceptical with data as with human sources. Interviewing data needs rough questions to data material: how journalists carry out the analysis in detail, what kinds of results they eliminate in the story and what has been the interpretive logic behind journalists' own conclusions.

Publishing a data story

As in traditional journalism, editorial judgements in data journalism are between public interest and individual privacy. In every situation the journalists face there will be unique considerations about whether and how to publish data.

Data used in a story should have a clear link with the results. Data can be erroneous, misleading, harmful, embarrassing or invasive. Presenting data as a form of journalism requires that journalists subject the data to a journalistic process. (Sonderman 2013)

Even if data is precise, its context to reality may be insufficient and it may not correlate adequately to the reality it should describe. This may be a serious problem in stories about opinion polls carried out using insufficient samples or wrong questions. The power that journalists wield as professionals is all attention. Journalists bring attention to a thing, and that attention has good and bad consequences. Decisions journalists make are often about what happens when attention is in this thing. (ibid.)

Even anonymized data can reveal private information. Journalists may also need to be careful about protecting sources in the way that they publish leaked data. Metadata stored in files concerning the date and location of access, the computers and accounts used, and other data may identify the source. (Bradshaw 2016),

Invasion of privacy is one of the major ethical problems in publishing a data story. Data may be confidential or connected to the wrong person. Data as a part of larger data sets may contain additional information that is actually not necessary for a particular story. This kind of *data-overdose* may offer journalists a kind of *peeping-tom-situation*. Publishing human data may endanger privacy in a story or in a separate database published along with a story.

Besides these kinds of data-overdose situations, journalists may indulge in *over-interpretations* when publishing their results. Data should not be forced to tell more than it is able to do and journalists should ask themselves do data tell precisely what they should do, is analyse comprehensive enough and in context with reality. Journalists are eager to search causalities between variables but may not take into account the not-so-visible third element that could explain the exceptional results.

Another ethical problem in publishing data relates to incorrect interpretations of the results. This takes place when journalists write their stories solely based on data and without using other sources or without requesting outside experts to

comment analyse results. If data is the only source material in a story, journalists may be tempted to *over-report* analysis results and at the same time neglect some important and relevant issues in the topic.

Sonderman (2013) points out that journalists should not assume that the data is inherently accurate, fair and objective. Neither, they should not mistake their access to data or their right to publish it as legitimate rationales for doing so. In publishing, journalists should think critically about the public good and potential harm, the context surrounding the data and its relevance to other reporting.

Journalists dealing with surveys should always request access to the original data, including all the questions asked. If that information is not forthcoming then the truth of the claims are not established and the journalist will need to take the decision not to publish. If the survey is already in the public domain, however, the ethical decision then concerns whether the journalist should report on the resistance to publish more details or data, and criticisms about the methods used. (Bradshaw 2013)

In a *cautious data story*, the question is how much journalists can stretch data analysis if data are not hundred percent clean. Dirt in the data is relative and certain amount of dirty data does not necessarily pollute the whole data set. However, journalists should take into account even small error threats caused by dirty data in the results and explain them to the public in a story. One aspect supporting hasty publishing in this kind of situation might be the importance to publish even approximate results.

Predictions are one type of data where the conflict between the principle of accuracy and minimising harm comes to the fore. The publication of predictions can be self-fulfilling or self-cancelling, as well as dangerous to publicise. Journalists should be especially wary when other predictions do not indicate the same thing, as can happen in political voting intention polls. (Bradshaw 2013)

In presenting data, context is the key. Measured results alone do not tell anything about whether those numbers are higher or lower than they should be, going up or down, or the best or worst in the region, country or world. Presenting them within a historical context, by person or by date helps make the numbers more meaningful. Personalization, however, can present problems of its own. If journalists tell users how things affect them, they should also have a sense of the larger picture. (ibid)

One ethical issue is the level of detail, at which a journalist is actually required to tell the story. Aggregate and less personal information may provide a clearer story

about broader trends, while random checks of the validity of the data may turn up stories about flaws in such publicly available information. (ibid)

Adding context is a vital part of the data journalism process. Journalists should put absolute figures into the context of the size of the local population, historical patterns, and even differing demographics. Moreover, they should check trends against changes in boundaries or data collection and classification methods. Journalists should adhere to the ethical principle of transparency in attributing sources and linking to the full data where possible, with the exceptions detailed both above and in the section on protecting sources below. (ibid.)

When a story is complete, journalists need to go back and check facts on a line-by-line basis. Whenever possible, facts should be traced back to either or both data and documents. According to Houston (2004), journalists need to play the role of *devil's advocate* with the data and come up with arguments against any initial finding. Houston argues journalists should share the highlight of own findings during their interviews and to listen and consider any criticism. As he summarizes "it is much better to find out you are wrong before you publish or air a story than after the public sees it". In addition, journalists need constantly think of the "lurking variable" referring to other factors that could cause the data to look the way journalist is seeing it.

Another ethical issue refers to the dumping of a raw database of public records onto the Internet. If journalists are not doing any kind of analysis or any kind of value-added work to the data, this kind of publishing activity enlightens and educates the public. To ensure journalists publish data responsibly, Sonderman (2013) suggests journalists ask themselves about the need to publish something, reasons for not publishing and how best to publish.

In considering whether to publish, journalists should have a clear idea of what they are trying to accomplish by publishing the data: what effects do they intend to have, does data really create value for readers, or does it relate to the other elements of reporting. Journalists dump data when they cannot come up with a more valid reason than "because we can" or "because we think it would look cool". Considering the reason for not publish something, journalists should spend some time thinking about likely problems that could arise from publishing a certain set of data. Questions about the harmed ones are especially important if the data set includes information about specific individuals. (ibid.)

Another concern relates to data accuracy. Even if data comes from a government source, there is a chance it contains inaccurate data. Any database is not really a

database but a set of data containing last-known recorded information that may be outdated, inaccurately recorded or inaccurately provided or contain a lot of missing information. Consequently, the data point of a story may be misleading. The third concern relates to whether data is relevant to a story. Journalists should ask themselves whether they have added enough context, whether they are presenting the data and how the reader should interpret it. (ibid.)

Finally, journalists have to decide how to present the data in a way that maximizes the benefits and minimizes the harm. The issues are about what facets of the data are truly essential, and which the journalist could restrict or redact. Journalists writing articles frequently have to decide whether to use a quote verbatim, or to paraphrase it. The same is true in presenting data. Journalists can manipulate the raw source data to enhance clarity, context or other principles. For instance, the U.S. Census Bureau while gathering, analysing and mapping all sorts of personal data in aggregated tables or maps, never identifies any individual. (ibid.)

In the age of instant information, journalists carry an ethical obligation to present data clearly and in context. Since simple charts and infographics only offer a snapshot of the data, the appropriate background and any guides to interpret the information are essential. It is true that story visualisation causes its own ethical problems. According to Cairo (2014, 25-27) visualising data is convincing and a proof of accurate information in a story and therefore people are less sceptical about new information. Visualization creates an impression among the public that the information in the background is reliable. To avoid these kinds of ethical problems, visualization and data should have a clear connection between each other. In addition, journalists should be aware of what kind of visualisation is the most suitable to what kinds of presentations. Unethical visualization may cause confusion among the public and lead to false interpretations. It is easy to lure the public to read a story with a good instance of visualization but not at the cost of clarity.

Bradshaw (2013) suggests data visualisation ought to be clear. For example, charts and tables should generally have a baseline of zero to avoid misrepresenting changes as being more severe than they are. Journalists should choose timescales to represent long-term trends rather than misrepresent results by starting or selecting from an all-time low or high. Portelance (2014) criticizes data visualizations to mislead with insufficient data points and a truncated graph that made superficial trends appear in uncorrelated data.

Both sources and journalists can manipulate visual representations of data. Baselines that do not begin from zero can be particularly misleading. Line charts

that begin from the lowest or highest point can suggest a much bigger drop or rise than the long-term reality. The use of 3D effects can actively distort proportions in a chart and adds meaningless noise to it. A pie chart that recedes into the distance suffers from the same problem as anything in the distance: a slice that is further away is smaller than the same slice in the foreground. (ibid.)

In publishing a data story, it is important to remember that data is just an object to analyse and require additional sources to complete the story. In many situations, additional sources may reveal loopholes in the analysis and show in a story important points of comparison with reality. However, data collected from the public may produce false results due to the possibility to fabricate answers and problems to identify responders.

For instance, LeFleur (2014) argues that analysis is just the start in the data journalism process. Reporters have to ask themselves whether their analysis is consistent with the findings. They should publish a detailed methodology about data and the journalistic process. Also, it is important to invite feedback and corrections from the public.

Conclusions of results should be discussed openly among colleagues in editorial offices and journalists are advised to look for outsider experts to evaluate the results. Getting a response on data before publishing is a vital step in checking its accuracy. (Bradshaw 2013)

In summary, journalists cannot withdraw from their ethical responsibilities regarding dirty data and must do everything to clean it before analysis. In certain minor unsolved problems of dirty data, it might be acceptable to provide the public in treating analysed results with caution. It is not acceptable for journalist to try to solve editorial problems in dirty data just by publishing the data and informing people about everyone's personal needs to decide how trustworthy the story might be.

Feedback from the public

External feedback received from the public is the last essential point in evaluating ethical standards in data journalism. In order to convince the media public of the analysis results, journalists should make their activities as transparent as possible. A convincing data story should contain a description of the used data and include the origin, content and character of the material. Important information to the public would also be the analysis methods used in the story. Interpretations and

conclusions of journalists should contain a kind of disclaimer and an invitation to the public to inform journalists about possible errors or absences in data material used in their stories.

An important ethical question relates to the role of journalists in providing required corrections to their stories. The public should have a possibility to notify editorial offices of likely errors in the data that warrant further checks, especially when it relates to them individually. When the story is based on publicly available data, also the original data provider should be notified of the errors. In these cases, there may be a follow-up story about flaws in the data itself.

Conclusions

In spite of the increased importance of data journalism, ethical discussions about its working practices have been scarce.

The ethical dimensions of data journalism can be traced through four phases of the journalistic working process: in access to data, in data analysis, in publishing a data story and in asking the public for feedback. Each of these phases contains several important points of attention and their possible consequences to subsequent phases.

In access to data, journalists have to assess the origin and content of data sets and their sufficiency in the planned story. Especially dirty data with its varied false content may jeopardise story preparations whenever journalists have not cleaned data sufficiently.

Analysis of data results needs accuracy and careful examination in order to avoid biases and to prevent journalists uncritically seeing what they hope to see in the data. Moreover, it is imperative for journalists to make certain that the interpretative logic behind their own conclusions is correct.

In publishing a data story, it is important to remember that the data is just a journalistic tool and analysis is just the start of the process. Journalists should challenge all their assumptions and findings by comparing these to reality. Data and analysis findings must have a clear connection with the story context. Data visualisation and privacy issues need specific attention.

To convince the media public of the validity of both analysis results and the story in general, journalists should make their activities as transparent as possible by describing data and their analysis methods. In addition, journalists should actively

ask the public to indicate possible errors in data collection or interpretation and be ready to provide needed corrections in the stories.

The frequency of data journalism is expected to increase significantly in the future. Consequently, ethical problems are expected to become more common regarding journalistic practices and therefore they need specific attention to be avoided. Ethical problems in data journalism may be more visible than in traditional journalism. Therefore, public credibility in data journalism and journalists needs broad transparency about journalistic activities.

References

- Bradshaw, Paul (2013). Ethics in data journalism: accuracy. Online journalism blog. [Blog] September 13, 2013. Available at <https://onlinejournalismblog.com/2013/09/13/ethics-in-data-journalism-accuracy/> [Accessed 25 November, 2017]]
- Bradshaw, Paul (n.d.). Data journalism. ONA Ethics. The Ethics & Excellence in Journalism Foundation. Online News Association [Blog] Available at <https://ethics.journalists.org/topics/data-journalism/> [Accessed 26 November, 2017]
- Cairo, Alberto (2014). Ethical infographics. *IRE Journal*, Spring 2014: 25-27.
- CMM (2014). About Council for Mass Media in Finland. [online] Available at http://www.jsn.fi/en/guidelines_for_journalists/ [Accessed 25 November, 2017]
- Gray, Jonathan & Chambers, Lucy & Bounegru, Liliana (eds.) (2012). *The Data journalism handbook*. European Journalism Center. Available at <https://datajournalismhandbook.org/> [Accessed 21 November, 2017]
- Houston, Brant (2004). *Computer assisted reporting. A practical guide*. Bedford / St. Martin's
- Kuutti, Heikki (2011). *Julkisuusjournalismi*. Jyväskylän yliopisto [Access to information in journalism] University of Jyväskylä
- LaFleur, Jennifer (2014). Data journalism tips from data driven ethics. Conference at Markkula Center for Applied Ethics [Blog] 1 January, 2014. Available at <https://www.scu.edu/ethics/focus-areas/journalism-ethics/resources/data-journalism-tips-/> [Accessed 20 November, 2017]
- O'Neil, Cathy (2013). On Being a Data Skeptic. O'Reilly Media Available at <http://www.oreilly.com/data/free/being-a-data-skeptic.csp> [Accessed 25 November, 2017]
- Otieno, Samuel (2013). Numbers don't lie: my experience with data journalism. Data Journalism Fellowship Blog [Blog] Data Dredger. Available at <http://internewskenya.org/fellowshipblog/numbers-dont-lie-my-experience-with-data-journalism/#sthash.dkVisHqo.dpbs> [Accessed 25 November, 2017]

Portelance, Eric (2014). Misleading with statistics. How journalists make arguments with distorted data. Medium.com [Blog] Available at <https://medium.com/i-data/misleading-with-statistics-c63780efa928> [Accessed 10 November, 2017]

Segnini, Giannina (2015). Investigating with databases: Verifying data quality. Craig Silverman (ed.): *Verification handbook*. [online] European Journalism Center. Available at <https://gijn.org/2015/05/14/investigating-with-databases-verifying-data-quality/> [Accessed 25 November, 2017]

Sonderman, Jeff (2013). Programmers explain how to turn data into journalism & why that matters. Poynter institute. Available at <https://www.poynter.org/news/programmers-explain-how-turn-data-journalism-why-matters> [Accessed 25 November, 2017]

Uskali, Turo & Kuutti, Heikki (2016). *Datajournalismin työkäytännöt* [Working practices in data journalism] Tampere: Vastapaino

Webster, Maryjo (2014). Data State of Mind. *IRE Journal*, Winter 2014:11.