media numeric

Co-funded by the Erasmus+ Programme of the European Union

BY NATALIA BERGER, JOKE HERMES, FAYE MERCIER, EMMA STOOP

MEDIANUMERIC

THE POWER OF STORYTELLING

WHAT KNOWLEDGE & SKILLS ARE NEEDED IN DATA-DRIVEN STORYTELLING TO EMBOLDEN STUDENTS OF JOURNALISM, MEDIA & CREATIVE INDUSTRIES?

CENTRUM CYFROWE · Storytek_ · SWPS University · SOUND&VISION · AFP · ina · nh · EU SCREEN

Colophon

About this report

The Power Of Storytelling is published by MediaNumeric, an initiative of the Netherlands Institute for Sound & Vision, InHolland University, L'Institut National de l'Audiovisuel, Agence France-Presse, SWPS University, Storytek, EUScreen and Centrum Cyfowe.

The ***Netherlands Institute for Sound & Vision*** is the institute for media culture; an inspiring, creative and accessible place for private individuals and professionals, focusing on current developments concerning people, media and society from a media-historical perspective.

InHolland University is a large university of applied sciences located in eight main cities in the Netherlands, offering practice-oriented education and research opportunities. They educate independent, critical professionals to make a meaningful contribution to the inclusive world of tomorrow, with a focus on fostering a sustainable living environment and a resilient society.

L'Institut National de l'Audiovisuel is a repository of all French radio and television audiovisual archives. Since 1974, Ina has been responsible for preserving, promoting and transmitting France's audiovisual heritage. Ina is also an international training and research centre for digital media and content.

Agence France-Presse (AFP) is one of the world's three major news agencies. Its mission is to provide rapid, comprehensive, impartial and verified coverage of the news and issues that shape people's daily lives.

SWPS University is a private non-profit university in Poland established in 1996, exploring the human mind and applying this expertise to address practical challenges in society, focusing on new technologies and dynamic social change.

Storytek Innovation & Venture Studio is a creative meditech and storytelling accelerator in Northern Europe, offering innovative business models and format options, as well as concrete needs for professional organisations in the European media ecosystem.

EUscreen is a network of European broadcasters and audiovisual archives, media scholars, and technical experts, facilitating access to and engagement with archival audiovisual content through their independent online portal.

Centrum Cyfrowe is a Polish think-and-do-tank supporting openness and engagement in the digital world by changing the way people learn, participate in culture, use the internet and exercise their rights as internet users.

About the authors

Natalia Berger is a research fellow lecturer on Inclusion and the Creative Industries at the Creative Business knowledge centre of InHolland University.

Joke Hermes is a lecturer on Inclusion and the Creative Industries at the Creative Business knowledge centre of InHolland University. Her research focuses on insights into how the creative industry can contribute to putting social issues on the agenda and solving social problems.

Faye Mercier is a media researcher and lecturer in television and cross-media at the University of Amsterdam, focusing on topics as media and representations of precarity on-screen.

Emma Stoop is

Reviewers

Johan Oomen, Jacqueline Pietsch, Rachel Somers Miles.

Cover Image

Designed by Rebecca Haselhoff, 2023.

Co-funded by the Erasmus+ Programme of the European Union

The MediaNumeric project has been co-funded by the European Commission under grant agreement No. 621610-EPP-1-2020-1-NL-EPPKA2-KA. This web site reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Table of Contents

1. Executive Summary

This report consists of a 'needs analysis' that aims to address the central question: what should a course on storytelling with data consist of, according to experts? It offers suggestions from more than 50 international experts consulted by MediaNumeric partners regarding benchmark examples, software packages to be used, and constructive didactic approaches. The material, consisting of the full transcriptions of all the interviews, was analysed using Grounded Theory. This is an inductive methodology that proceeds through three stages of coding. In 'open coding' researchers take their cue from what informants say and produce codes that summarize arguments, examples and evaluations given. In 'axial coding', the open codes are brought together into broader themes. The third and last stage is 'selective coding' in which researchers focus on those themes that stand out and connect these with available academic sources. Selective coding then results in the theorization of the data. In the selective coding phase of this needs analysis, we have focused in particular on 'super codes'. They are the *impact of data*, the *affect of data* (the feelings experienced by individuals when thinking of data and data analysis), and *perceptions of data*. 'Impact' is a well-known term in classic mass communication research and will not come as a surprise, it denotes the effects the use and analysis of data have. 'Affect'[1] is a less well-known term that has its origins in the humanities and in philosophy. It denotes, in this case, the feelings and emotions that are linked to data. These can be seen as a form of connection and energy that exceeds rational arguments. Especially important is the fear that exists widely when it comes to data, working with figures and programming. 'Perception' again is a well-known psychological term that denotes interpretations by individuals of a phenomenon.

Together the super codes cover talk of data across what is expected of data, the feelings data produce, and how others likely understand data and what data can do. It will be clear that the super code *perception* points back to the *impact* and *affect* of super codes. Together they map an interpretative regime[2]. The super codes also make clear that discourses on storytelling with data are structured by various social scripts. The script here denotes a relatively fixed way of thinking about data and how and why data-driven work is important today in relation to disinformation and fake news. These social scripts must be incorporated into the design of the course in order to ensure that it not only meets the needs of students and (future) professionals but also guarantees that the course is shaped in a deeply informed

[1] For more on this account see: Goldberg, G. (2012). Negotiating Affect in Media/Cultural Studies; Hemmings, C. (2005). Invoking affect: Cultural theory and the ontological turn. *Cultural studies*, 19(5), 548-567; Ruddick, S. (2010). The politics of affect: Spinoza in the work of Negri and Deleuze. *Theory, Culture & Society*, 27(4), 21-45; Cole, D. R. (2011). The Actions of Affect in Deleuze: Others using language and the language that we make. *Educational Philosophy and Theory*, 43(6), 549-561.

[2] The underlying methodology of coding will be explained in the dedicated chapter.

way and that it both addresses and tackles the inevitable reality that data is often perceived to be scary and/or overwhelming.

2. Introduction

Inholland University of Applied Sciences (Inh) and Agence France Presse (AFP) took the lead in researching the *state of the art* when it comes to data analysis in and for non-fiction media production, as well as in conducting a *needs analysis* from the perspective of the professional fields involved (creative industries, journalism) regarding the contents and the didactic form of a course to train students in media and creative industries in how to tell stories with data. This is the introduction to the analysis of the expert interviews carried out by all partners in order to provide a solid foundation for the development of such a course.

3. Methodology

This chapter offers a brief overview of the research questions and instruments of the field research. The interview procedure is described and substantiated. It also explains why full transcripts were used and how these were analysed.

3.1. Main Research Questions

After a brainstorming session with the INH, AFP and NISV teams, the following main research question was formulated: *"What knowledge and (teachable) skills in storytelling are needed to embolden students of journalism, media, and creative industries to take on the opportunities that data-driven innovations bring?"* Thereafter, the following sub-questions were formulated:

- What are the relevant educational options offered by European higher education institutions (HEI) and commercial enterprises?
- How do the experts value the relevant higher education (HE) offerings in their countries?
- What are the essential knowledge, skills and tools (for the search and exploration of data - storytelling - checking and debunking misinformation) in their professional fields, according to experts?
- What is the best way to teach (learn) these skills in a higher education setting, according to experts?

3.2. Procedure

As documented in the project proposal, two research methods were chosen for the field study — desk research (mainly for the State of the Art report) and expert interviews (mainly for this Needs Analysis report). The Inholland team worked with 56 interviews (among these 15 interviews were with the members of the MediaNumeric Stakeholder Board) in part delivered by the French and Polish partners. These were transcribed in full using a transcription tool kindly made available by AFP. The interviews were checked by the interviewers and collected in a shared document folder. Based on the grounded theory methodology, Inholland's team then analysed the material; the results are offered below. The analysis was used to curate the pedagogical and didactic suggestions that were offered in the interviews to develop the training course curriculum, as well as provide a first analysis of which perspectives are relevant to the course.

The average length of a cleaned-up interview transcript was twenty-five A4 pages so the total amount of textual data to be processed consisted of some 1,200 pages. To reduce the data and gain insight into underlying themes and commonalities across the expertise of the

interviewees, grounded theory was used (Glaser and Straus, 1967; Strauss and Corbin, 1990). Grounded theory prescribes a logic of open coding, an intuitive procedure in which, in our case, three coders started selecting relevant quotes that they named (or coded). By blind-coding other coders' transcripts, the coding was formalised into a lengthy draft codebook. As saturation was reached (no more new codes could be developed out of the material), the codebook was used to take grounded theory's next step which is to construct 'axial codes' which are built on the open codes and allow for thematic structuring of the now vastly reduced data. The axial codes allowed the Dutch team to present a number of *design requirements* for the course under development. The grounded theory then prescribes a third step, which is to develop 'selective' codes in order to theorise the material. As theorisation was not the goal of this exercise, we chose to treat the selective codes as 'super codes': perceptions of data, affect of data, and power(s) of data. As overarching themes, the super codes helped organise this needs analysis document and the design requirements for the course. By calling them 'super codes' we want to point to their value in framing how to think about data analysis. They are more than a thematic organisational tool, they point to how thinking, feeling and working with data are connected. The term 'super code' needs to alert us against understanding the quotes presented here in an overly descriptive manner or to think of design requirements in too mechanical a manner. . The super codes help us comprehend learning how to work with data from a broader perspective.

These super codes will be particularly helpful in developing the course as they demonstrate the various scripts that are available when discussing data, and therefore they elucidate what perspectives on working with data to use when building the course.

3.3. Research Instrument: Interviews & Selection Procedure

Prior to the field research phase, INH and AFP held a series of online consultations and brainstorming sessions with all MediaNumeric partners. All project members were asked to fill in a form and share their ideas regarding the needs research . Five of seven partner teams delivered more or less detailed information and suggestions, which have been reviewed and summarized into one integrated document . After a Zoom discussion, the designated interviewers settled on a simple and two-fold set of open questions . The names of members of the Stakeholder Board were divided among the partners for an interview. The rest of the interviewees were selected by the partner organisations themselves. They approached potential interviewees based on their experience, affiliation, and their relevance for the objectives of MediaNumeric. The basic principle of the selection procedure was to ensure a balance between the experts from the relevant fields — creative industries, mass media, data science and archival work. Most interviews took approximately one hour, though a few were shorter at about 30–40 minutes. There were also a few longer interviews lasting 90 minutes. Noteworthy is the enthusiasm of the interviewees for this project and the enormous number of examples and suggestions given.

Informants were approached by three national teams. All three teams employed an international perspective. In as far as there are 'national' differences, these mostly manifested as slightly different views as to which professionals to include, and what professions to address. The *super codes* provided a way to work from a deeper logic and address three essential crossovers for today's highly diversified, cross-platform, and cross-media industry when it comes to non-fiction storytelling, journalism and, potentially, marketing. Although journalists were the target professional group for most of the interviewees, in general, MediaNumeric aims to appeal to a broader group of media professionals, if only because a significant part of the paid work in the media industry has shifted away from 'pure' journalism to mixed forms of non-fiction media storytelling. In addition, the use of data for marketing purposes could be understood as *in-house storytelling*. In any case, the training of future professionals should allow them the widest possible range of choices rather than restrict them.

Importantly, *three perspectives* on dealing with data in journalistic and media industry environments revealed themselves in the material. Firstly, interviewees mentioned that among journalists and content producers there is a need for stronger knowledge and know-how in working with data. As one of the experts put it,

> *"What we don't have are training courses that bring together all these different disciplines [journalism, elements of data and social sciences], which would be absolutely necessary for students who are planning to go into these new professions."* (Sophie Jehel, Lecturer at University of Paris 8 Saint-Denis)

The example of improved fact-checking skills and processes comes across strongly in the French team's interviews (AFP) (although this is, at least in part, due to the fact the AFP team emphasised this topic in their interviews). Secondly, others stressed that data analysts should be hired in the media industry but only when they have been trained to understand how data analysis can be meaningful in journalism and other media non-fiction content production. Thirdly, the lack of knowledge of data analysis among managers in media companies was regretted. Managers need to be taught basic skills, so goes the argument, in order for data analysis to become properly assimilated in media content production. The reasons behind this need have a practical explanation:

> *"[...] it's more complicated to convince managers of the interest of establishing long-term tools and of exploring databases without knowing what the results will be in advance. In other words, it's complicated to find the same flexibility that can be given to investigators who work on a lead*

> *for three months when you don't know if it will lead to anything. I would say it's a bit of a 'trick' that can work, but that's supposing that there are sufficient resources and that the managers have the bare minimum amount of confidence in it all."* (Maxime Vaudano, Journalist of Le Monde, Les Décodeurs)

It would be safe to suppose that this lack of confidence among managers is due to their relative lack of knowledge in data-driven approaches to storytelling.

3.4. Data Analysis Approach

Qualitative research depends in large part on whether interviewers establish a rapport with their interviewees. The 'tone and feel' of the interviews matters. As Bryman (2008) explains, qualitative methodology leads many researchers to a "commitment to viewing events and the social world through the eyes of the people that they study" (p.385). The epistemological meaning of this kind of research has been explained by Lofland and Lofland (1995): (1) … face-to-face interaction is the fullest condition of participating in the mind of another human being, and (2) … you must participate in the mind of another human being (in sociological terms, "take the role of the other") to acquire social knowledge" (p. 16).

In the case of the 56 expert interviews that comprise the data, there is enthusiasm, willingness to share, and enormous generosity in taking time to talk about data analysis and multimedia storytelling. All of the interviews mirror a sense of urgency for the type of training course MediaNumeric aims to develop. Even though many examples were given of existing courses, often in slightly more specific domains (such as fact-checking), the necessity of our intended study programme has been acknowledged by all experts. One emblematic utterance of this sentiment came from Johan Linden, Project Manager of 'Future of News' at SVT (SE):

> *"When you ask that question: What is the perfect course? I would say that it is a course that I haven't seen yet, but that I would like to create. It has kind of the perspective of creative storytelling, like in text, in sound, in visuals and visualisations. With storytelling, you know the basics, go back to the Greeks, how do you create a story, a narrative? So you need that."* (Johan Linden, Project Manager Future of News at SVT, SE)

Whenever we found a critical note, it was usually to express concern that it would not be possible to carry out such key teaching within as short a timeframe as one week provided for in this programme.

The three perspectives mentioned above show us how:

- Data can be an entry point for a hybrid representation of reality – as the media landscape has changed, alongside changes in structures of belief and authority, (reliable) data is increasingly important;
- Professional storytelling, therefore, needs a new combination of skills (for manipulation, interpretation, visualisation);
- As data needs interpretation and visualisation, media management now demands a new set of competencies when it comes to hiring, coaching, and assessing the work of different kinds of content makers. In particular, the teams in which these content makers ideally work need facilitation.

3.4.1. The Super Codes

In order to give a theoretical perspective to the data analysis and construct a theory *grounded* in the collected evidence, we chose three super code labels. The three super codes revealed themselves when we focused on the open codes with the three perspectives outlined in the previous section. The super codes are:

- Perceptions of data;
- Affect of data;
- Power(s) of data.

To demonstrate how the data collected in the interviews were interpreted, in the below sections we share some examples of quotes and labels.

3.4.2. Perceptions of Data

Typical examples for 'perceptions of data' hinge on whether data is seen as a material that can be used to manipulate, or whether data is a force for truth-telling. Keywords in such cases would be: 'engineer', 'lying with statistics', and the way in which journalism is understood as a key pillar for democracy.

Data is for Everyone

- *"So, I think that having better quality data is useful for anyone, not just for the professionals."* (Marco Rendina, Istituto Luce - Cinecittà, IT)

Data is Human

- *"[...] because, at the end of the day, a lot of data sets are about humans and human lives and how human relationships are. And I think because of that, I do feel kind of weird about automating that; the idea that there is no human touch in the data collection process when what we're trying to tell stories are about humans. Maybe in*

five years I will have changed my mind and now I'm going to just have all of the machines do it for me, but as of right now with the state of the technology as it is, I think I'm cautiously optimistic about where it could be in the future. Now I'm very much scared and very concerned about using automation in our data process and the data collection process." (Shirley Wu, Independent creator of data visualisations, freelance, USA)

Talking to Data/Data as Language

- *"That's how I discovered data, because in sports, economics and ecology, I was confronted with data and, I must admit, I didn't know what to do with it. I had been taught to interview people, situations, to do reports, but I didn't know how to interview numbers, so I sort of started out on my own, asking myself the following questions: What does it mean? Can I trust this figure? What's the source? And to not just copy and paste press releases. I had more of a scientific background in high school, so I had an appetite for it, but during my years of political science, I had completely skipped those parts."* (Karen Bastien, WeDoData, Co-founder, FR)
- *"I think that it sounds very old fashioned, but I think it's really, really important to focus on language. Also, if you're talking about data journalism language and visuals. I mean you shouldn't lose that at the end, that's how you communicate."* (Daniela Kraus, Presseclub Concordia, journalist, AT)
- *"It also illustrates the ability of data journalists from different newsrooms around the world to speak the same language and to work together on databases."* (Karen Bastien, WeDoData, Co-founder, FR)

Data Magic/The 'Awe' of Data

- *"I think that there is also a great fantasy around data, in the world of big data, which tends to lead people to believe that all you have to do is retrieve a file, press three keys and lines of code appear, etc. But it doesn't work like that at all. It's a job where the human element is very important. It's not the machine that does everything, it's the human who configures it, who checks, analyses and interprets the data, etc."* (Karen Bastien, WeDoData, Co-founder, FR)
- *"Because there are many different algorithms that work in many different ways, in my opinion, it's very important to create a canon or a syllabus for laypeople to better understand the way those other works work, and how they make those decisions so that it will be less of magic, and more of something that can be more or less understood."* (Bartłomej Balcerzak, Polish- Japanese Academy of Information

Technology, PL)

Data is Delicious/Juicy

- *"If you look at it like that, the really kind of delicious data moments though are where you get literally spreadsheets with numbers on it. And you interrogate that in a way that creates something, that can illuminate an issue. That's when it gets really powerful."* (Peter Rippon, editor of BBC Online Archive, UK)

Data is Numbers & Graphs

- *"I still often see that data journalism is completely mixed up with something different, which is infographics."* (Daniela Kraus, Presseclub Concordia, journalist, AT)

Data is Dystopian

- *"Data can be manipulated for malicious intent. For example, deep fakes, face swaps and other applications using known algorithms can ultimately threaten the safety of society, whether these actions are carried out by conspiracy, or simply just for simple goals like money, and the pursuit of power."* (Bartłomej Balcerzak, Polish- Japanese Academy of Information Technology, PL)

3.4.3. Affect of Data

Affect of data is important in how informants talk about the MediaNumeric course. Typical keywords are 'fear', 'fun', and 'visualisation'. The following are some examples from the interviews:

Data is Beguiling

- *"I think it's that people who want to present something in as positive a light as possible can manipulate it very easily, in a very, very beguiling way. And again this is the journalistic challenge, which is to spot when that's happening."* (Peter Rippon, editor of BBC Online Archive, UK)

Data is Intense/Exhausting

- *"I think there is a fear of engaging in this field because there is a lack of data culture among editors-in-chief. They don't really understand what it is — they only see the*

time-consuming aspect. It requires work from a lot of people, it's very expensive, and they ignore the editorial aspect that could never have been produced without all that work." (Karen Bastien, WeDoData, Co-founder, FR)

Data Work is Dangerous

- *"I think I mostly mean that you need to be very careful with the outliers of the data, with the insecurity of the data to be honest. If you want to define what data means, you always need to be very careful about insecurity, confidence intervals, and those kinds of things. So I think that data journalism requires a little bit more of a scientific approach. And in a way as journalists, we of course want to write the story, but that story needs to sell, and therefore we like to, if we have a fact, we like to over-exaggerate it a little bit so the story is nicer. With data, it just requires a little bit more carefulness."* (Tom Claessens, data journalist, Follow the money, NL)

Data is Scary

- *"I think there is a fear of engaging in this field because there is a lack of data culture among editors-in-chief. They don't really understand what it is — they only see the time-consuming aspect. It requires work from a lot of people, it's very expensive, and they ignore the editorial aspect that could never have been produced without all that work."* (Karen Bastien, WeDoData, Co-founder, FR)
- *"I think that there is also perhaps a generational issue and that until we have a generation at the head of newspapers that 'got their hands dirty' with data in order to really understand what it is and to master it from an intellectual and practical point of view, data will remain this sort of 'black box' that scares people."* (Karen Bastien, WeDoData, Co-founder, FR)

Data is Overwhelming

- *"I think because big data sets have so much information, there is a big chance that I might get really overwhelmed or side-tracked, so it's really helpful for me to have a goal and an intended audience so that I can focus my data exploration."* (Shirley Wu, Independent creator of data visualisations, freelance, USA)

Data is Collaboration

- *"For example, we analysed the Lubrizol Chirac media phenomenon, where on the one hand there was a fire at the Lubrizol factory in Rouen, and a few hours later Jacques*

Chirac died. There are two concomitant media phenomena, yet one (the fire) is almost totally overshadowed by the other (Chirac's death). The idea was to measure the gap between the concerns of Twitter and the editorial choices made by the news channels. What I took away from this was really the collaborative approach with the journalist, which would have been very different from the approach I could have had with researchers in the humanities." (Nicolas Hervé, INA, Senior researcher, FR)

3.4.4. Power of Data

The power(s) of data is signalled by the often-used terms 'fake news' and 'interpretation'. To be clear, however, data is nothing without interpretation — and therefore requires more than a technical approach.

Data is the New Seat of Power

- *"You want to be able to hold powerful people to account for the decisions they make that affect everybody in society and believe me the power of data is going to be what the powerful people…the people that control the data are going to be in power. And so your ability to understand that and to interrogate that is absolutely fundamental to your purpose as a journalist."* (Peter Rippon, editor of BBC Online Archive, UK)

The Power of Data Lies in its Format

- "*I think good data is something that really needs to be improved. So, if we want inputs to the fact-checking process we need high-quality, well-published data that we can trust. And the amount of good data on the web, by which I mean openly published, openly licensed data, with good metadata around it so that we can understand what we can and we can't use it for, that is available in a machine-readable serialization; so, it's not that a government or statistics office had published a PDF, it's that they've published some data that is of the web and can be interacted with. It's a very unglamorous kind of world to work in, but the fixing the plumbing of the web, of making sure that this data is published in a way that can be used and is interoperable with more of the fact-checking process, is a technology thing, and is also a cultural thing that will take time to fully see through."* (Andy Dudfield, Full Fact, Head of automated fact-checking, UK)

Data Can Be Powerful Even if it is Not Visible

- *"So it may be that you use zero numbers, zero graphics in your final article, even though there may have been important data work behind it."* (Cédric Lombion, Open Knowledge Foundation & School of Data, FR)

Data Skills are a Legitimising Force

- *"Generally speaking, journalists are not held in the highest regard by scientists and researchers, and I think that working on data has made it possible to re-establish a much more constructive level of exchange. I think that data journalism brings a lot to the table through collaborations outside the office, and in particular the world of research; we saw this with the Panama Paper leaks."* (Karen Bastien, WeDoData, Co-founder, FR)

Data (Visualisation) Brings New Perspectives

- *"Indeed, at the time, no journalist was really going to dig into the figures, so it opened up storytelling perspectives that other journalists didn't have. I saw the potential, and then I saw, above all, that if I could put data in my papers, I noticed that this data took on more strength when we managed to tell it visually."* (Karen Bastien, WeDoData, Co-founder, FR)

Data is a Way of Understanding the World

- *"So, for me personally, I guess this is not as much about moving conversation. I've always used data as a way for me to understand the world. I think that in any data visualisation, the first part is for me as a creator to understand the data set well enough to be able to communicate what's important or insightful within it."* (Shirley Wu, Independent creator of data visualisations, freelance, USA)

The Power of Data can be Abused

- *"I don't know whether I'm optimistic or pessimistic. Most days I'm quite pessimistic because I think the powerful people who control the data have such an advantage that it's going to be very difficult for journalism to perform the role that it performs in society as effectively as it has done up until now."* (Peter Rippon, editor of BBC Online Archive, UK)

3.5. Quality Control

Validity and representativity in this type of qualitative research are given with the transparency of the research procedure (reliability), iterative study design, use of long quotes (validity), and the use of purposive sampling (Tongco, 2007) for the selection of informants (representativity and validity). Representativity and validity were further buttressed by discussing the patterns identified via the super codes and the selection of quotations that ground them with the interviewers from the different national teams.

4. Results & Analysis

In this chapter, we present the results of the first iteration of the analysis. The priority of this stage was given to the completion of task T2.1, which includes reviewing relevant curricula in higher education, collecting experts' opinions on the needs of data-driven solutions for storytellers, and understanding the local specifics in this regard. Additionally this section also explores the conversation of who the target group of this course is, as discussed with experts, as well as outlining the topics that have been presented as necessary for such a course on data-driven journalism and creative storytelling.

4.1. Task I: Reviewing Relevant Curricula in HE

"What is currently crucially missing in the available programmes is skills development that allows the study of contemporary multi-platform audiovisual culture, including the incorporation of technology and media ethics." This statement, posited in the MediaNumeric Detailed Project Description, has been supported during our desk and field research, while at the same time acquiring a more specific and nuanced meaning.

All partner teams were asked to fill in a questionnaire about the relevant educational offer in their countries. Together with a small sample of relevant training programmes in the EU, the overviews delivered by FINA, INA, SWPS and INH formed the empirical basics of our analysis. At this moment our observations of the relevant educational offer in Europe should be considered as indicative, and not representative, in a statistical sense.

There are different educational options where data-driven technologies in storytelling are included in a study programme. At first glance, the offer is very broad and fits different categories of enrollees: from school graduates to professionals who want to obtain specific data-related skills in different phases of storytelling (data gathering, processing and presenting). Also, the range of enterprises that aid such programmes varies between the classic state, applied sciences universities, and commercial parties. Figure 1 below shows these options summarized:

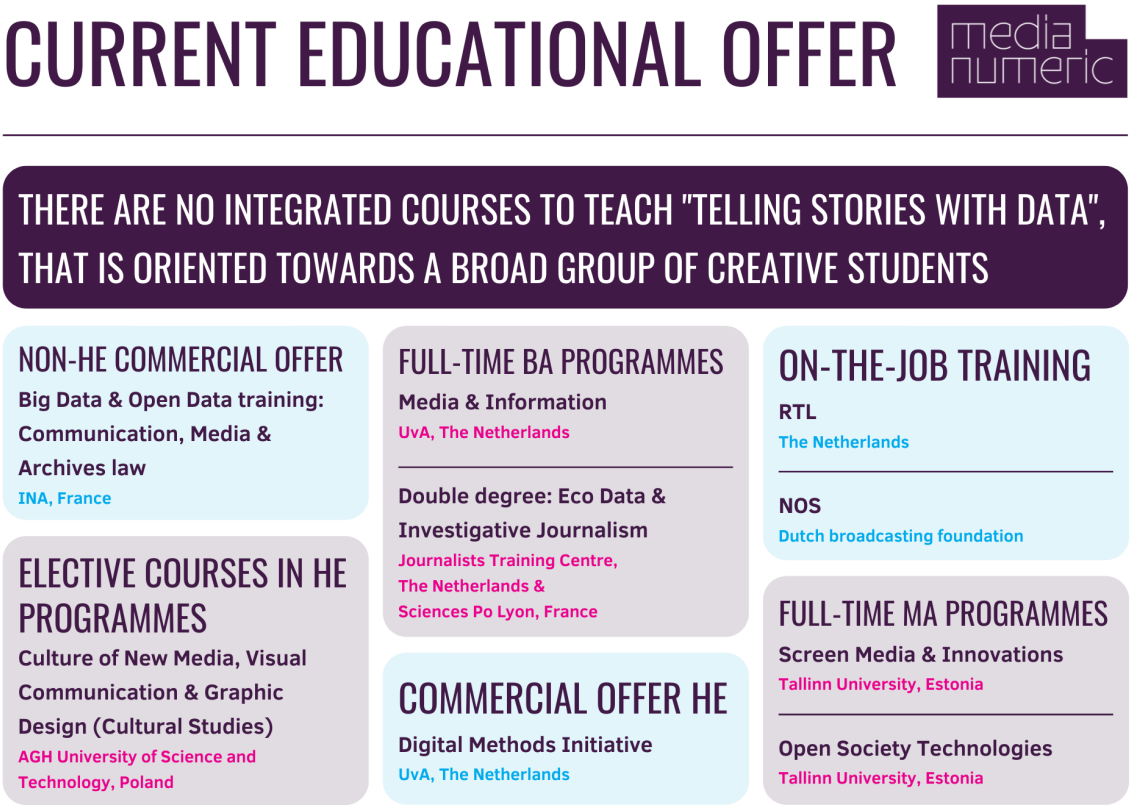


Figure 1. *Summary of the current educational offer.*

Full-Time HE Programmes

Let's look at the educational offer in more detail. In all countries and to our knowledge, alliance students can follow **BA and MA programmes** in Communication and Media, and Creative Business (with different modifications in the programme's title), where the data-driven component is presented in the form of a major or elective course. To show the architecture of the relevant educational offer in one given EU country, we will take the Netherlands as an example.

Both types of regular higher education institutions in **The Netherlands** — research universities and universities of applied sciences (Hogeschool, in Dutch) have some components relevant for our review. For instance, the Netherlands Erasmus University Rotterdam[3] has a Bachelor of International Communication and Media programme, where the following data-related courses are offered: New Media Technologies, Digital Content; New Media and the Creative Industries. Master programmes are Media & Creative Industries (with the courses Innovation in the Creative Industries and Production Cultures in the Streaming Industries); Media & Journalism (Digitalization and Network cooperation);

[3] https://www.eur.nl/

Psychology of Digital Media (Text Mining, Content Analysis, and Data Visualisation using R; Applied Multivariate Data Analysis).

The University of Groningen[4] offers the full-time Media Studies bachelor programme, where the relevant courses are: Imagining the Digital; Digital Cultures (elective); Web Design; Media Theory II: Form and Technology; Transmedia Production. There is an entirely data-oriented master programme titled Digital Humanities. The Media Studies master programme entails different tracks such as Datafication and Digital Literacy, Media Creation, and Innovation.

At the faculty of Media and Information at the University of Amsterdam[5] there is a BA programme called New Media & Digital Culture, with courses in Digital Methods, Information Analytics, and Digital Journalism. Besides this, there is an elective course on Digital Practices. UvA offers the two-year international Erasmus Mundus Master's programme in Journalism, Media and Globalisation, where a course in Digital Journalism is included.

The Faculty of Creative Business of the Amsterdam University of Applied Sciences[6] (HvA) offers the following data-related courses: Digital Media: Content; Digital Media: Concept and Creation; Visual Storytelling; Crossmedia Work; Digital Press Room; Digital Design; Digital Media and Society.

The Faculty of Communication & Multimedia Design at the Avans University for Applied Sciences[7] has a programme where technical and creative skills can be obtained simultaneously during projects tied to various themes, such as Social Design and Storytelling. All students follow their own path, designed through the different LABs (Technology LAB, Minor Research in Immersive Storytelling).

The faculty of Creative Business at the Breda University of Applied Sciences[8] is designed for media professionals. It seems that data-driven skills can be obtained during the 3rd & 4th years of this study programme through participation in one of several of the following projects: ‘trend watching’, ‘cross-platform thinking’, ‘data-driven decisions’, and ‘creating & producing innovative concepts’.

The Faculty of Communication & Multimedia Design at the Hanze University of Applied Sciences[9] (Hanze UAS) offers Visual Design (the programme includes ‘Digital Design’, ‘Programming Basics’, ‘Usability & Interface Design’, ‘Transmedia Storytelling’ and ‘User-centred Design). The full-time programme Creative Media & Game Technologies (starting in September 2021) includes, among others, the following courses: Virtual reality, Design & Tech, UX/UI.

[4] https://www.rug.nl/
[5] https://www.uva.nl/en
[6] https://www.amsterdamuas.com/
[7] https://www.avans.nl/international
[8] https://www.buas.nl/en
[9] https://www.hanze.nl/eng

Inholland University of Applied Sciences[10] has a faculty and a BA programme called Creative Business that offers rather limited data-related content included in the following modules: Streaming Emotions (creating value with data).

The Faculty of Creative Business of the University of Applied Sciences Utrecht[11] offers for second-year students the following relevant courses: Technology; Data Storytelling; Cross Media Fundamentals; Interactive Experiences (elective). The School of journalism of the HvU offers the full-time programme Master Data-Driven Design, oriented for those who want to be a 'data journalist' or a 'digital creator'.

Saxion University of Applied Sciences Leiden[12] has a Faculty of Creative Business that offers a Media, Information and Communication program. Based on the official website information, it would be difficult to say what data-driven themes are included. The same could be stated about the Faculty of Creative Business at the Steden University of Applied Sciences[13] and their Media Management programme.

To summarize, at this moment the Dutch HE system has a varied offering of data-related courses addressed to the BA and MA students, who follow programmes in Creative Business, Communication and Journalism. However, knowledge and skills related to storytelling with data are scattered throughout courses that are usually focused on specific aspects of digital technologies in a specific (narrow) domain.

According to the information delivered by FINA, SWPS and INA, the situations in **Poland** and **France** seem to be comparable to the current situation in the Netherlands. Yet, we still have very fragmented ideas about the relevant educational options in these countries.

The catalogue of the University of Warsaw[14], namely the Faculty of Journalism, Information and Book Studies[15], listed a few relevant studies for the MediaNumeric programme, such as Architecture of Information Space, Journalism and Media Studies, and Big Data Management. As FINA's team reported, some other universities in **Poland**, including but not limited to the University of Wrocław[16] (UWr), the Jagiellonian University[17] and the SWPS University of Social Sciences and Humanities[18] have programmes or courses that could be of interest to our project. However, more detailed revisions of their curricula could be useful.

Colleagues from INA focused on the specific educational options localized in journalism schools in **France**. As such, there are: double degree programmes in data and investigative

[10] https://www.inholland.nl/inhollandcom/
[11] https://www.internationalhu.com/
[12] https://www.saxion.edu/
[13] https://www.nhlstenden.com/
[14] http://www.uw.edu.pl
[15] https://www.wdib.uw.edu.pl/
[16] https://uni.wroc.pl
[17] https://en.uj.edu.pl/
[18] https://english.swps.pl/

journalism delivered by the Journalists Training Center (CFJ)[19] and Sciences Po Lyon[20]; the Paris Journalist Training Center, ESJ Paris[21], offers in the Master programme Digital Media and Journalism a module on "data journalism."

> *"I have the impression that the training courses that exist for data journalism in France are rather superficial and are limited to the DataViz aspects, how to create graphs, etc."*
>
> *- Sophie Jehel, lecturer, University of Paris & Saint-Denis*

As in the Netherlands, the regular HE offer in France and Poland serves primarily the interest of students who want to specialize themselves in a particular (narrow) application of data in media and creative practices. There is no universal course oriented *to all creative students* and that would cover the demand for basic knowledge and skills in telling stories with data.

To compare, there are a couple of examples from the other EU countries. The Faculty of Arts at the Aarhus University[22] in **Denmark** provides a very comprehensive full-time Bachelor's programme in Information Science, which is unique as normally this type of study programme would be facilitated by technical faculties. This same department also has a BA programme in Media Science, where the first-year students follow the module Digital Methods. The Danish School of Media and Journalism[23] (DMJX) has a programme in Interactive design that includes the study of data visualisation. Tallinn University[24] in **Estonia** has a few BA programmes related to storytelling (The Press, Audiovisual Media and Cross Media); yet only the curriculum for the cross-media programme includes modules related to work with data -- "transmedia design" (at least, this is what we could conclude based on the website information).

During the expert interviews, the question of higher education for journalists and creative professionals, specifically its content and quality, has been touched upon by the interviewers. The overall expert opinion about the level of data-related qualifications of young professionals is not high. "I have the impression that the training courses that exist for data journalism in France are rather superficial and are limited to the DataViz aspects, how to create graphs, etc.," noted Sophie Jehel, lecturer at the University of Paris 8 Saint-Denis

[19] https://www.cfjparis.com/
[20] https://www.sciencespo-lyon.fr/
[21] http://www.esj-paris.com/
[22] https://international.au.dk/
[23] https://www.dmjx.dk/
[24] https://www.tlu.ee/tallinna-ulikool

(FR). Furthermore, she mentioned “fragmentation of the university courses” related to data, “lack of tools” and “somewhat homemade way” of working with data and statistics. At the same time, Sophie explained that the situation in Schools of Journalism is much better than at universities: they have “fairly comprehensive programmes on the subject [data journalism].” Peter Burger, founder of the nieuwscheckers.nl and the associate professor at the Leiden University (NL), complains about “both sloppiness and lack of training for journalists working in mainstream media” when it comes to skills in assessing and analysing data:

> *“Quite a few of them tend to accept, for instance, press releases at face value; will just use one source and not dig into the original publication; lack the skills to judge statistics.”* (Peter Burger, nieuwscheckers.nl, ass. professor LU, NL)|

According to this expert, this lack of basic knowledge and skills in journalistic work comprises a “major problem”:

> *“Quite apart from anything to do with bots and trolls and information operations and stuff like that, journalists, in general, could do better.”* (Peter Burger)

Close to this general issue, there is a shortage of historical and sociological knowledge demonstrated by young content creators. Louise Broch, an archivist with 20 years of experience, shared with us her concern: “Sometimes I think that the brand new journalists don't know enough actually about how the society is built.” This means that archivists must increasingly answer impossible requests such as “is it possible to get a sound recording from a king in the 18th century? for example!”

Commercial Offer (HE and Private Parties) of Vocational Training

As our desk research revealed that the educational options in data-driven storytelling are not limited to conventional HE institutions. There is a plethora of **commercial training** courses, workshops, etc. offered by regular universities, schools, and private bureaus. Usually, these programmes have a broad postgraduate professional group as their target group. A couple of examples: Media and Information Literacy at ESJ Lille, the oldest journalism school in France; Training - Datajournalist at CFPE, France; INA provides an Understanding with the Universe of Data training. The University of Amsterdam and its Amsterdam School for Cultural Analysis houses the Digital Methods Initiative (DMI), the leading research and training lab in digital humanities. Twice a year DMI organizes a series of intensive training sessions with different themes. The Summer School 2021 is called “Fake Everything: Social Media's Struggles with

Inauthentic Activities." Utrecht University has its own Data School, which also provides seasonal training sessions; the programme this summer is titled "Exploring Culture Through Data." The Digital Humanities Center of the Institute of Literary Research of the Polish Academy of Sciences[25] also runs the Center for Digital Humanities but its priority lies more in academic research and as such, has less direct relevance for our review.

According to the information delivered by the SWPS's team, in Poland, there are a number of labs and initiatives in media and journalism education, which, as in the Netherlands, are connected to universities: Medialab Gdańsk, whose events are dedicated to new media and technology in science and culture, the responsible use of new media and technologies. The lab has the New Media Academy, a new project about the "potential of new media for culture staff." Notably, the Academy's[26] webinars (with rather narrow practical themes like "Personal Branding in Social Media – for Creators') are free.

The University of Applied Sciences Utrecht (HvU) sells post-bachelor courses such as Data Visualisation and Infographics, Data-Journalism, and the Masterclass: Introduction to Data-Driven Design. INA invites "documentalists, archivists, librarians, managers of audiovisual and multimedia collections, manager of audiovisual and multimedia content, museum professionals" to follow the one-day training called Understanding the Universe of Data. For "written press, TV, radio and web journalists", INA also has a two-day course entitled Data Journalism Training: Collecting and Analyzing Data. Additionally, INA organizes training in Big Data and Open Data: Law for Communication, Media and Archives.

The other French institution, ESJ PRO, provides a short training in data journalism and information visualisation; the aim here is to "Learn to find, sort and use databases. Knowing how to present information in a visual, attractive and interactive way."

Data visualisation as a component of storytelling, but also as an independent specialisation, seems to be very popular among those who seek a professional switch or want to improve existing skills.

According to the Nationale Beroepen Gids (National Occupation Guide), in the Netherlands, there are 65 'best' educational courses and training options for data visualisation offered by commercial enterprises (we suggest that there are more such courses, this is merely an indication). The duration of these courses is between 1 hour and 16 days, with the price varying from EUR 89 to EUR 5,454 for a course. Besides visualisation, training programmes can also include "data science", SQL, Python, Power BI, and other extra data skills. Here is an example of one of these well-rated course providers:

The Graphic Hunters[27] (GH) is an initiative of Goof van de Winkel. He has many years of experience in developing, organising and implementing training sessions. The portfolio of GH

[25] http://chc.ibl.waw.pl/pl/
[26] http://www.akademia.medialabgdansk.pl/
[27] https://graphichunters.nl

is very multifarious in the choice of topics and trainers. Beginners and advanced professionals in DataViz would find something relevant here.

Of course, we should not forget the very extensive possibilities of data- and storytelling-related online commercial courses, as well as free Massive Open Online Courses (MOOCs) available for everyone. Think, for example, about Coursera[28] whose courses are not expensive and certificates are recognized by many employers.

Non-Profit Offer

For those who don't want to pay or have no means for it, there are free courses in data and its applications in creative work. To name a few, the Poland Media School of Critical Thinking[29] accommodates an educational programme for students and journalists with the aim of improving their skills in the critical reading of media messages and in dealing with fake news. In the Netherlands, anyone who wants to learn about artificial intelligence can follow the National AI course[30]. A comparable course was created at the University of Helsinki – Elements of AI[31]. Great sources of learning materials about data and access can also be found on the US-based web source Open NY[32]. For more advanced professionals The New York Times offers their NYT Open platform[33], where journalists and other specialists share with an audience "how we design and build digital products." The newspaper even makes available to everyone, the three-week programme[34] they use to train journalists in-house. The London School of Economics and Political Science (LSE) shared open access to their AI Journalism Starter Pack[35], created for the JournalismAI, a project of Polis – the journalism think-tank at the LSE. The project is powered by the Google News Initiative[36].

On-the-Job Training

This kind of schooling is typical for large media enterprises such as Dutch broadcasters RTL and NOS, Britain's Daily Telegraph, national broadcaster the BBC, and many other mass-media companies who hope to provide their employees with up-to-date professional education where data and its use is becoming more relevant.

Analysis of commercial, non-profit, and on-job training offers was out of the scope of our research; therefore, the above is merely the result of a brief observation. However, for the purposes of MediaNumeric, namely, the development of study programmes, it would be

[28] https://www.coursera.org/
[29] https://media.ceo.org.pl/
[30] https://www.ai-cursus.nl/
[31] https://www.elementsofai.com/
[32] https://www.ny.gov
[33] https://open.nytimes.com/
[34] https://drive.google.com/drive/u/0/folders/1ZS57_40tWuIB7tV4APVMmTZ-5PXDwX9w
[35] https://docs.google.com/document/d/1pWwbqPERg0bUbMHMbYYDWmFQmWJYvK8N2Dmbenp4Qu0/edit
[36] https://newsinitiative.withgoogle.com/

useful to learn about relevant non-HE practices in greater detail.

4.2. Task II: Expert Interviews: Revision of Needs in HE

The key questions used in all expert interviews were:

- What skills, tools, and knowledge are necessary (for the search and exploration of data – storytelling – checking and debunking of misinformation) in the given professional field? What would be a bare minimum to ensure efficient daily work with data?
- What is the best way to teach (learn) these skills in a higher education setting? For example, what would be included in a list of "dream teaching formats" or the methodological setup of the course?

4.2.1. General Advice per Module

Due to the three core modules defined in the MediaNumeric detailed description, the interview questions and discussion with experts evolved around the three topics: 1) search and exploration of data; 2) tracking and debunking misinformation 3) telling stories with data. One may note that the order of the modules in the original project description document (p. 15) is different. However, it seems more logical to follow the architecture of data processing used in data science (as shown in Fig. 2) which also fits the logic of the journalistic workflow (Fig. 3).

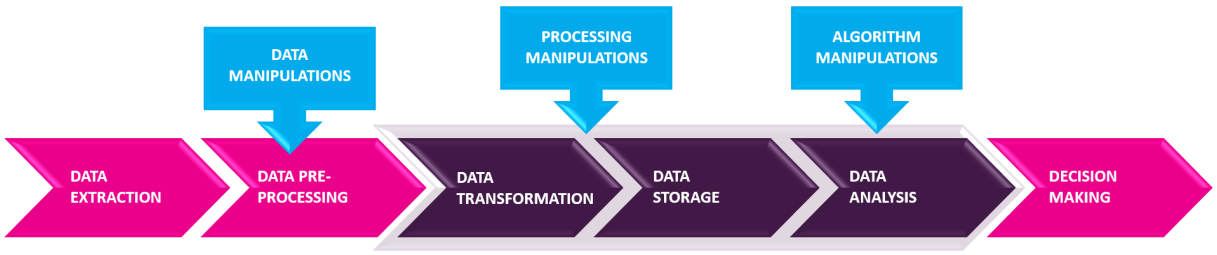


Figure 2. Re-visualised version of the flowchart from: Data Analytics Pipeline In: L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017). Machine learning with big data: Challenges and approaches. Ieee Access, 5, 7776-7797.

It seems that the process of tracking and debunking misinformation belongs to the stage which precedes the final – reporting and presenting – phase of the storytelling.

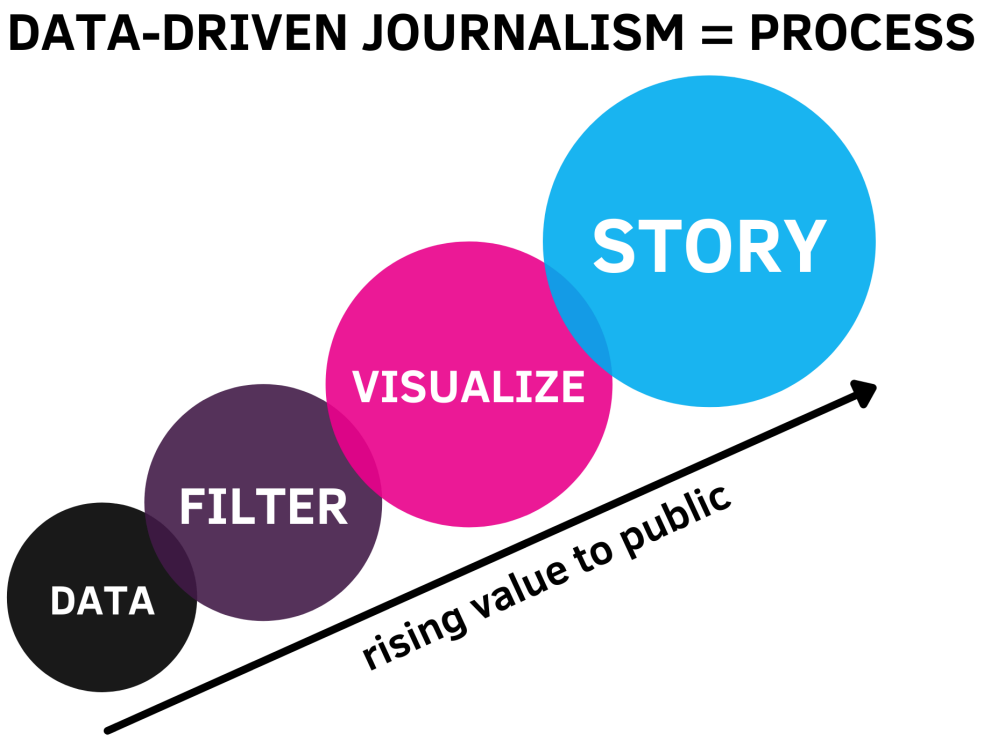


Figure 3. *Re-visualised version of the Data-driven journalism workflow in: Lorenz, Mirko. Data-driven journalism: What is there to learn? Conference materials, based on presentations of participants, August 24, 2010, Amsterdam, The Netherlands. http://mediapusher.eu/datadrivenjournalism /pdf/ddj_paper_final.pdf*

Regarding the direction of Module I (search & exploration data), the experts were in agreement on the following:

- Basic journalistic/storytelling skills, principles etc. are not new and still essential in data journalism and data-driven storytelling
- Data literacy: overcome the fear of numbers & technology; basic statistics; data manipulation skills; common ground with tech professionals
- Tools & databases: basic principles and choice of apps.

"The journalist should still be relatively clear, I think, about the overall approach and what they want to get out of the data, even if they are not the one manipulating it, even if they are working with a data scientist, a data engineer, etc."

- Nicolas Hervé, senior researcher, INA

While developing Module II (tracking & debunking misinformation) we were advised to consider:

- Making critical thinking an essential part of the programme. Teach/let students learn to ask questions. What is the truth?
- To not deny deontology as a basis of the journalistic profession.
- Using open sources and initiatives such as FactCheck.org, Poynter.org, Snopes.com

"...if they use Facebook data and Google Analytics, they also need to know about all of these companies. I really think they also need to know the logic and the market situations of these companies. Because there are also ethical questions. And if you use a Google search, then do you even know how it's ordered?"

- Tamara Witschge, professor HvA

For the creators of Module III (telling stories with data), the following ideas expressed by the experts are worth keeping in mind:

- Basic skills are important. Like the other steps of the storytelling process, the presentation and reporting (writing, telling, etc.) phase demand the mastering of ABC knowledge about the audience, genres, colours, psychology of text and image perception.
- Numbers and statistics always need explaining and contextualization (in line with academic research: Corner, 2021).

- There is a need for collaboration between different content creators – this request is applicable for all units of the storytelling with data process, also on the stage of creating the final product.

"... if you ask me if I had to quickly prototype for you a good story, I probably would like to have this 5% data layer represented in some really cool way and really meaningfully, similarly as in a good, long read." (Alek Tarkowski, Open Future Foundation, Pl)

"It's always a combination of many different tools. There is no one golden solution. It is a combination that people use." (Goof van de Winkel, Graphic Hunters, NL)

"I think it's very important that students have a very broad knowledge of what is possible, kind of a mindset. And then in-depth knowledge of the different specializations because nobody can do everything as we know." (Daniela Kraus, journalist, PL)

"[...] I think that there should be a basic knowledge of statistics, i.e. knowing and mastering all the fundamentals of statistics: "respecting the data." (Karen Bastien, WeDoData, FR)

4.2.2. Advice in Teaching Approach, Didactics, and Ideas for a "Dream Course"

Before we elaborate on the recommendations regarding the teaching approach in more detail, we will first summarize the most common ideas retrieved from the conversations with experts about a 'dream course format' (Fig. 4):

- A multidisciplinary approach to the content, assignments and the teaching team (coaches, guest lecturers from different fields);
- Show the correlation with basic skills in storytelling (first, the basis should be in order – knowledge and skills in finding the story, collecting data through the "old school" investigation: interviews, searching for documents, etc.; verification and presentation).
- Organize a course around real-life cases, use small-scale data sets, let students collaborate with professionals, show them a process and include a playful element in

the course structure.

While discussing the constraints of designing a course on data journalism, Cédric Lombion explained how to provide further learning opportunities to students who are especially motivated to work with data:

> *"We are structuring the second year with minors, so they will be able to have, say, a television major and a data journalism minor. Making it work with the constraints that we have is a big, big, big nut to crack. We had started doing interviews with colleagues from Brazil, Turkey, Greece, London on the topic of what the constraints are for the delivery of a data journalism curriculum in their universities. And everywhere they are often the same: there's one constraint of the kind of pipeline of students coming in because there's still this idea of 'I'm going to study journalism because I'm not good at math.' And then there's the constraint of, for example in France and in Brazil, by law there are a number of mandated hours of sociology, of whatever other courses the journalism curriculum should include, meaning that you have very little margin to include new topics, including data journalism. So what they do is that they cut a little bit of the course around investigation and put one part of it that is focused on data journalism. But these constraints are there, and the question is how do we build strategies that optimise the time that we have within the curriculum, but that also potentially provide stuff that is extra-curricular to organise things where students who have the drive to go farther in data journalism are given the space. There's not that space for students who are very driven within the main curriculum. And so I think that anyone around those data journalism courses has to include those constraints and propose some kind of maybe different solutions or packages that universities could take and adapt for themselves that would include both 'this is what you could do within your curriculum depending on how much space you have."* (Cédric Lombion, Open Knowledge Foundation & School of Data, FR)

Experts and practitioners about teaching approach:

> *"I would rather get them to ask an interesting question on a small data set and understand manipulation skills and then when they're ready enough, let them learn SQLs."* (Daan Odijk, data scientist RTL, NL)

> *"...just make the course really practical […] most jobs in journalism won't give you 5 days for a project […] it's just really important that you can get*

it into your daily routine, to think about data, tell a story about data. But also think about how I can maybe in one or two days make a story out of data?" (Jochem Bruins, TV-journalist, NOS, NL)

"[Money is] the most challenging thing because ... commercial data courses [for journalists] is one for the long run. It's not one you can use immediately, but if you want to make sure to make journalists more data literate, we have to give them more data skills. It's not something you learn in one workshop." (Yordi Dam, data-journalist, NL)

"What is the perfect course? I would say that it is a course that I haven't seen yet, but that I would like to create. It has the perspective of creative storytelling, like in text, but in sound, in visuals, and in visualisations" (Johan Linden, Project Manager Future of News at SVT, SE)

Figure 4. *Summary of experts' views on the teaching approach MediaNumeric should take, with word size corresponding to its prevalence across interviews.*

As detailed below, experts shared with us a few specific recommendations regarding the didactic side of the planned course in storytelling with data.

The most prevailing advice was to be **practice-oriented**. In summary, this means:

- Use real-world cases and questions;
- Work on projects / concrete examples related to news or students' daily life;
- Learning by doing;
- Study visits;
- Simulate the work field by 4 pillars: Practical/hands-on projects, set in real and tangible contexts/topics, quick production (like in a newsroom), work in (interdisciplinary) collaborations (no one can do it all).

A young journalist occasionally working with data gave the following advice regarding teaching students to deal with **time pressure**:

> *"I think you just make it really practical. I think most jobs in journalism won't give you 5 days for a project. So I think, first of all, for many journalists, it's just really important that you can get it into your daily routine, think about data, tell a story about data. But also think about how I can, in one or two days, make a story out of data? I think that for, let's say 95% of the journalists, that's way more interesting than doing a two-week project."* (Jochem Bruins, TV-journalist, NOS, NL)

The study cases should be **small-scale**; one doesn't need to process huge data sets to learn how to read and interpret statistics, for instance. The practical aspect is not so much about learning every skill but learning the process. According to the experts, this will make it easier to imagine possible approaches/solutions and collaborate with other disciplines. In short, this will help students to know where to look, and to understand what is possible and what is not.

> *"It's also about the can-do mindset, right? So if you have never created anything, never done anything with code before, you're completely dependent on the person who can do it. But even if you have solved some simple problems with code, or created something, then you have some feeling of the process and the possibilities, and you kind of learn that a no might not mean a no, that it might mean we need to do it differently, or in more time, or something like that. So it helps to have a little bit of hands-on experience with tooling."* (Koen van Turnhout, professor HU, NL)

In line with the "real world" approach, the experts insisted on good "best practice" examples, which should be incorporated into all elements of the training programme.

As we already pointed out, the word "**collaboration**" has been intensively used by almost all informants in varied contexts. This is due to the essential nature of this requirement for data work in a storytelling business. This is a big theme across all interviews. Experts say 'no one can do it all alone, you need to collaborate with people,' especially if you want to make new innovative formats. But you need to learn *how* to work together with people from other disciplines. That is, you need to learn what to ask for, how to speak their language, and how to manage a collaborative project. The following are some quotes related to this point:

> *"I think it's too much to ask if one person needs to do all of it. So I think it's really about the right collaborations. [...] I think in the course if you have different students from different disciplines coming together in a classroom, and they create something. So it's really about making from the start. ... And then you guide them in the interdisciplinary collaboration as a teacher. Rather than telling them how to do it, you really use the strengths of each discipline and see that every discipline has the same right to co-form the story."* (Tamara Witschge, professor HvA, NL)
>
> *"The collaboration is interesting (between the person responsible for a format and the other responsible for a subject), and this is something that we don't really hear about. We still retain the image of a journalist who works alone, whereas data journalists need to collaborate (with the editorial staff, or researchers, etc.) and to share certain practices, certain technologies while keeping their sources."* (Alexandre Léchenet, Independent, La Gazette des Communes, Data journalist, FR)

"We still retain the image of a journalist who works alone, whereas data journalists need to collaborate."

- Alexandre Léchenet, data journalist

Working on collaborative projects can lead to **reducing fear of numbers and technology**. This fear is often named as one of the main reasons that many creative professionals are reluctant to engage in data work. According to observations of the INA interviewer, the following ideas are relevant for developing the MediaNumeric course:

- Data journalism is at the crossroads of different disciplines and approaches: journalism, computer science, graphic design, data ethics, data literacy, project management, product owners;
- Data is the common language between data scientists, researchers in social science, and journalists.

Demand for common ground was often articulated by experts from different fields during the interviews – it was suggested by data scientists, archivists, academics, and journalists.

As to the forms that collaboration could take, the INA observers extracted this advice:

- Involve professionals from different sectors and backgrounds (journalism/design/IT);
- Encourage practical group work, for instance, hackathons on concrete investigations, accompanied by developers;
- Spotlight live feedback formats with experienced journalists from different countries;
- Create a supplement for online coding classes adapted to journalism.

In every interview, the question of tools was raised and discussed. As the researcher from INA put it, the experts emphasised **the importance of having simple and versatile tools.** In this regard the following topics should be kept in mind:

- There is currently a lack of tools in universities of social sciences;
- Priority must be given to tools that do not require coding skills;
- Students must test new tools and question themselves constantly;
- Tools must be easy and fun to use.

According to our informants, the proposed course should focus on breaking down essential components of data retrieval – that is, a **deconstructive approach**. The logic of this approach is demonstrated in the following quote from the interview with Johan Linden, journalist and trainer:

> *"You need to do some lab-ing with it in the class to understand what you can do. What part of the data process is this? What do I need here? And then you, of course, start in the beginning. If you search for data, you need to have a thorough understanding of web crawlers and how they rank searches."*

Emphasis was also placed on **being creative in your approach to data** – tailor it to the creative skills of the students and **tie it to storytelling**:

> *"...you can make coding, and the whole notion of the tech world, the programming, really creative and interesting, if you can tie it to storytelling. If you are early on in teaching, say that 'you need to code to be able to tell this story'."* (Johan Linden, Project Manager Future of News at SVT, SE)

Different experts mentioned that the **'play' element** incorporated into lessons and assignments would be effective:

> *"Involve them in some kind of fun, so that it would be pleasant. So that they would also know what it could be useful for."* (Beata Biel, Director of Premium Content, Discovery / TVN; Journalist, academic teacher, PL)

This research shows that the creators of the MediNumeric programme will be challenged by the short duration of the course, and will have to consider these specific **time constraints**. On one side, we will have comprehensive and in-depth training, on the other, the course should take the form of two-week-long sessions per year. Another solution, advised by the experienced lecturer in digital humanities, professor of UvA Richard Rodgers: use a flipped classroom!

> *"...we do most of the training by doing, and most of it's a flipped classroom. So you're asked to follow video tutorials prior to coming and then we do projects."* (Richard Rogers, professor HvA, NL)

Another expert, Cédric Lombion, outlined how to construct a course for a tight timeframe based on his own experience:

> *"We've split it into five days with homework in between going through the data pipeline with students explaining the key concept. Basically, we explain the key concept in the morning and practice in the afternoon. We go mostly technical until they are at ease with using spreadsheets, Python tables, and tools like this in their work. They learn a bit about scraping, they learn about data extraction from PDFs, they learn a bit about the common tools used for data visualisations. But of course, in the time that*

we have, we can only cover the basics. And then the second week is mostly focused on delivering a data project. Basically, they have to write a dossier on the topic of their choice. Often it comes back to the same topics that are easy for students new to data journalism, such as sports, transport, policy renewal, elections. Last year we had something around climate change that allowed us to do more work with geo-data, so that was interesting. But generally, we often come back to these similar topics because we aim also to have these projects be both around data, but also on local subjects." (Cédric Lombion, Open Knowledge Foundation & School of Data, FR).

"What works well with my students, is to make them work on controversies and debate methodologies by forcing them to defend different points of view and by structuring the debate"

- Sophie Jehel, lecturer, University of Paris & Saint-Denis

Not all experts could answer questions about specific didactic elements. However, those who teach themselves expressed some valuable suggestions:

- **Offer introductory courses** to the various different facets of working with data that young creative professionals need to be aware of; it's not about making them experts, but giving them an awareness.

  "I think that having these introductory courses in the different domains is quite important. That we see programming, statistics, design, storytelling. They are important. (Marco Rendina, Istituto Luce - Cinecittà, International projects coordinator, IT)

- **Curate a mindset/perspective.** This falls somewhere between a didactic approach and course content, but ultimately the point is that right throughout teaching you need to ensure that students understand the kind of data work that they're going to be doing. Cultivating the right perspective will ensure a better relationship to

data-driven work overall. This also means emphasising what they're **not** doing as well as what they **are** doing.

> *"A key aspect of the courses is at the beginning when we are in the defining phase, and we insist a lot with students on the need to find a conflict. So when students start to think about data they tend to have a sort of mindset where they are going to make a report. So they are going to collect data about a topic and make a kind of repost about the topic. But then we tell them all the time that they're not social scientists. So they're not going to make a report about the sociology of sport and who has success etc, etc. They're not public administration officials, so they're not going to make a report on the impact of a law. Instead, your starting point should be a conflict. [...]*
>
> *So every year, we've been kind of iterating on the way we try to basically optimize the time that we have to have students reach a level where if they are interested in the topic, because that's a major point, they are able to autonomously continue their learning. If they're not interested in that topic at least they have the basics of navigating spreadsheets and being able to understand some technical terms that are now common vocabulary when talking about journalism investigations."* (Cédric Lombion, Open Knowledge Foundation & School of Data, FR)

According to Richard Rodgers, **constraints and templates** are vital success factors for the kind of training that we want to develop.

> *"You have to provide constraints, rather than really open-ended stuff. It's like, 'Here's the recipe. And here's the research question." (Richard Rogers, professor UvA)*

Besides being provided with the templates for research questions, preliminary results, feedback, final report and presentation, students should be given clear timeframes and deadlines. For instance, the students could be advised to use The Harvard Misinformation Review format for writing a report: "a weird format, but ... it **combines journalistic techniques with a sort of social science**" (Richard Rogers).

Recommendations regarding the ideal profile for teachers has been highlighted by a number of our interviewees.

Nicolas Hervé, a senior researcher at INA identifies the importance of having a multidisciplinary team, namely, a diversity of profiles. For example, people who are self-taught and those who have learned on the job; who have a twofold skillset of journalist and IT (ideally binomial rather than one single person); and who have both technical data and business competencies on the floor.

> *"The ideal model would therefore be to have a pair with, on one hand, a journalist who is a specialist in their field and who has a good data culture, and, one the other hand, an engineer/data scientist/statistician who also has a basic knowledge of the field in which they work."* (Nicolas Hervé, INA, senior researcher)

The expert who for years organizes DataViz seminars with leading specialists as speakers advised us to seek out trainers who will be able to inspire the students:

> *"...especially now when we are all online, it's really nice to have the opportunity to have international names that could be part of your course, that could be part of your curriculum, so they could give some inspiration."* (Goof van de Winkel, Graphic Hunters, NL)

The other informant emphasized the importance of involving practicing specialists from a given professional field in a feedback process:

> *"Personally, it is this type of format (live feedback) that inspires me the most and that I think could be useful to students: you have a journalist who comes and explains how he carried out his investigation in his country, the information and data he used, how he built his databases, what computer language he used, etc."* (Alexandre Léchenet, Data journalist, NL

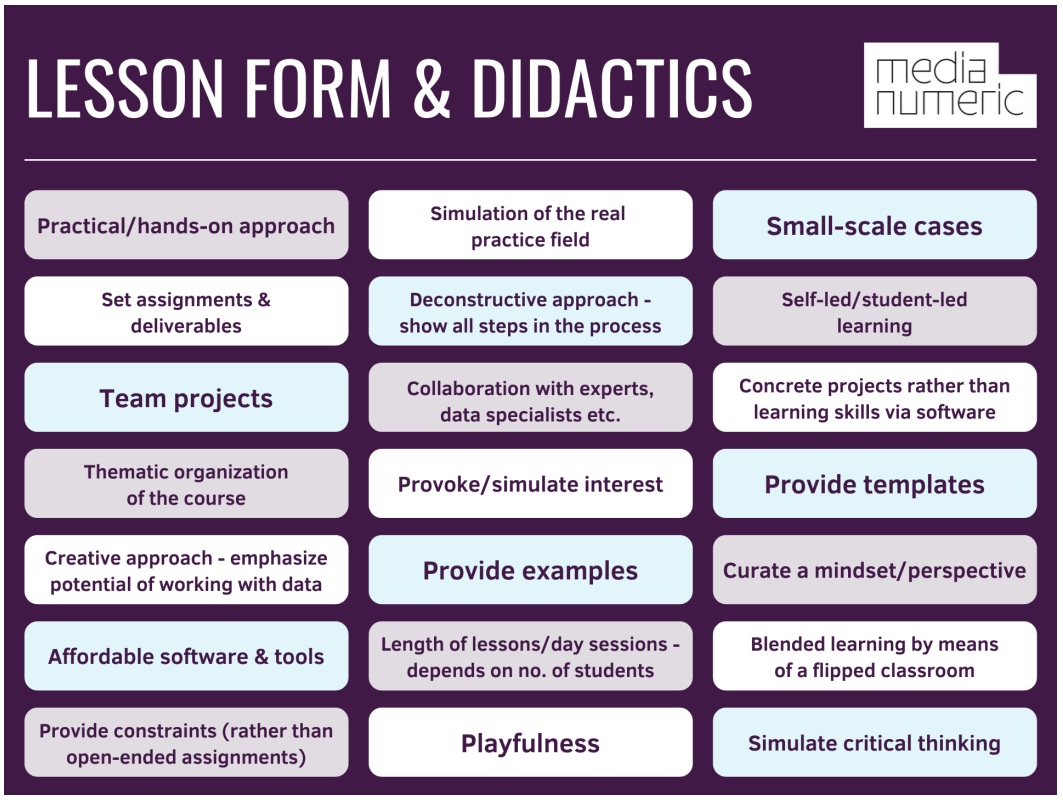


Figure 5. *Summary of experts' views on lesson form & didactics that MediaNumeric should focus on/employ.*

4.3. Target Groups

The question of who the course's target group is was regularly asked by the experts. They wanted to know who we were planning to teach. Yet, some indirect suggestions were given regarding specific groups, meaning (future) professionals, who will benefit from the course that we have in mind. Namely, the following types of professionals:

- Journalists & creative professionals who want to work with data;
- Data analysts who will work in the media/ creative industry;
- Managers in media and creative business.

After the series of consultations, all partners agreed upon the following profile of the target group: third to fourth-year BA students and first-year MA students in media, journalism and creative industries curricula.

4.4. Topics that Need to be Included in the HE Curricula

Before the partners of our alliance will be able to answer the question, “what should we teach?” a clear definition of *whom we are going to teach* should be given. The relevance of the topics discussed in the interviews would depend on which target group the training will be addressed to: beginners vs advanced; creative professionals (incl. journalists) vs managers; level of computer and data savviness etc.

Analysis of the whole corpus of the interview transcripts shows that there is no obvious consensus among the experts about the exact topic list for our teaching programme. Yet, a few topics were emphasized by most of the respondents:

- Basic, "universal", non-data- or computer-driven skills in storytelling (also for visualisation). Students should be shown how these skills are applicable to data-driven storytelling.
- Ability to collaborate with professionals from other fields/establish “common grounds.”
- Understanding principles of human-computer interaction/computational thinking.
- Basic statistics.

As we already mentioned, collaboration should become one of the most important elements of didactics. At the same time, collaboration as a topic should also be included in the course that we are aiming to develop. There are different types of cooperation:

- Inside a project team: how to divide roles and collaborate.
- Outside a project team: communication with field specialists.
- Global collaboration.

These are statements from the experts interviewed that illustrate this:

> *“In general, it is very rare for a data journalist to be able to combine all the skills on their own, and they are instead spread over several people (journalist, designer, developer).”* (Karen Bastien, WeDoData, Co-founder, FR)

Or with regards to outside expertise:

> *“There are some topics that I don’t touch by myself because I think it really needs a domain expert.”* (Shirley Wu, independent creator of data visualisations, USA)

A young data journalists who regularly collaborates with less data-skilled colleagues explained the necessity of 'common grounds':

> *"[...] you also need to be able to present [data] in an understandable way to your fellow co-workers who are less experienced with this. So they can basically understand what you want. Usually, I get questions from the other journalists about what they want in terms of data.* (Tom Claessens, data journalist, Follow the money, NL)

The same thought was echoed by the other experts:

> *"I think that it sounds very old-fashioned, but I think it's really, really important to focus on language."* (Daniela Kraus, Presseclub Concordia, Secretary-General, Journalist, AT)

> *"[...] You have to know how to 'talk about it', how to make it accessible to as many people as possible.* (Karen Bastien, WeDoData, Co-founder, FR)

The red line of all interviews is the emphasis on the priority of basic storytelling skills. This demand will comprise a challenge for the developers of our course. Shall we include themes such as 'specifics of an audience', 'how to formulate research question & hypothesis', 'critical thinking' and 'deontology' in a programme? If yes, to what extent should we spend time on the *classical component?*

While teaching data collecting skills, the experts advised us to pay special attention to good preparation work. Namely, the students should learn to choose and formulate an angle of their future story. After the right angle is found and understood, the students could use suitable capturing tools such as API and scrape most effectively.

The experts did not agree upon the answer to whether or not we should teach programming skills. The need for coding skills was advocated less often than advice to teach students how to manage data without programming.[37] However, all interviewees insisted on the necessity of knowledge in statistics. The name Excel, Microsoft's well-known spreadsheet programme, was mentioned by almost all our informants.

Although the experts mentioned Excel often during the alliance meeting in Paris in October 2021, the members of our consortium decided to work with Google sheets for the MediaNumeric training programme. Free access, integration with other Google products, creating charts and other advantages of this application make it more useful for the students.

[37] For the case that we'll choose to teach programming, the preferred languages should be Python or R. SQL could be advised for those with more advanced computational skills.

In terms of statistics, the emphasis should be on understanding *what* statistics should be used for, not just *how* to work with them.

> *"The basic knowledge of statistics, statistic reporting would be good for journalists in order to better understand data manipulation, to avoid data manipulation themselves."* (Bartłomej Balcerzak, Polish- Japanese Academy of Information Technology, PhD in computer science with a major in natural language processing, PL)

The following competencies related to data collecting were considered by the experts as most valuable for future story creators:

- ability to read, understand and interpret statistics;
- ability to explain numbers to a specific audience;
- manage data capturing tools.

> *"First, there is a story, then you go into the data to see whether the data can support the story, or what is the most important information that can be used for the stories."* (Tom Claessens, data journalist, Follow the money, NL)

> *"It's good to have an introduction to statistics and big data and an introduction to programming."* (Marco Rendina, Istituto Luce - Cinecittà, IT)

Next to data collecting skills, attention has to be paid to an ability to authenticate and verify data, which experts described as a combination of 'forensic work' and computational tools.

4.4.1. Media & Data Literacy

In the framework of the MediaNumeric programme, it seems useful to formulate our common definitions of media and data literacy. Although the experts didn't use these terms explicitly, they often mentioned topics that could be assigned to these kinds of literacies. For practical course building reasons, it will be good to operationalize and introduce them in our project. For example:

- Know trustworthy sources
- Source assessment
- Distinguish mis- and disinforma
- Recognise manipulation in mec
- How to query
- Data retrieval
- Data exploration
- Ask questions

- Know debunking tools/platfor
- Open data sources & navigatio archives)
- Data manipulation
- Data law

"Determining the ways in which something bears an element of truth but is misleading or is completely false, is often very hard."

- Peter Cunliffe-Jones, University of Westminster & International Fact-Checking Network

Understanding that data is not the truth, and essentially that data sets come with pre-inscribed biases and motivations also falls under critical thinking.

4.4.2. Data Visualisation

Data visualisation as a potential topic of our future programme was also discussed during the interviews. Based on the answers, we can conclude that, unlike other data manipulation knowledge and skills, data visualisation more often require an in-depth, special (self-) education. Still, for our target group, working knowledge of data visualisation would be of great value. Special attention should go to the understanding of:

- Audience
- Potentials of visualisation
- The power of visualisation, and how it can be misleading
- How visualisation can provide clarity/new perspectives
- Learn to avoid over-visualisation; accessibility is key
- Different types of visualisation, such as network analysis
- Tools

"For me, the key for a good project would be a combination between text and visualisation. Sometimes, for DataViz, I'm afraid that we are more attracted, as journalists, to make something that is very beautiful, to the detriment of the quality of information. First, what matters is the quality of the data, then the quality of the text and only after the quality of the data visualisation." (Andrea Abellan, European Journalism Centre, NL)

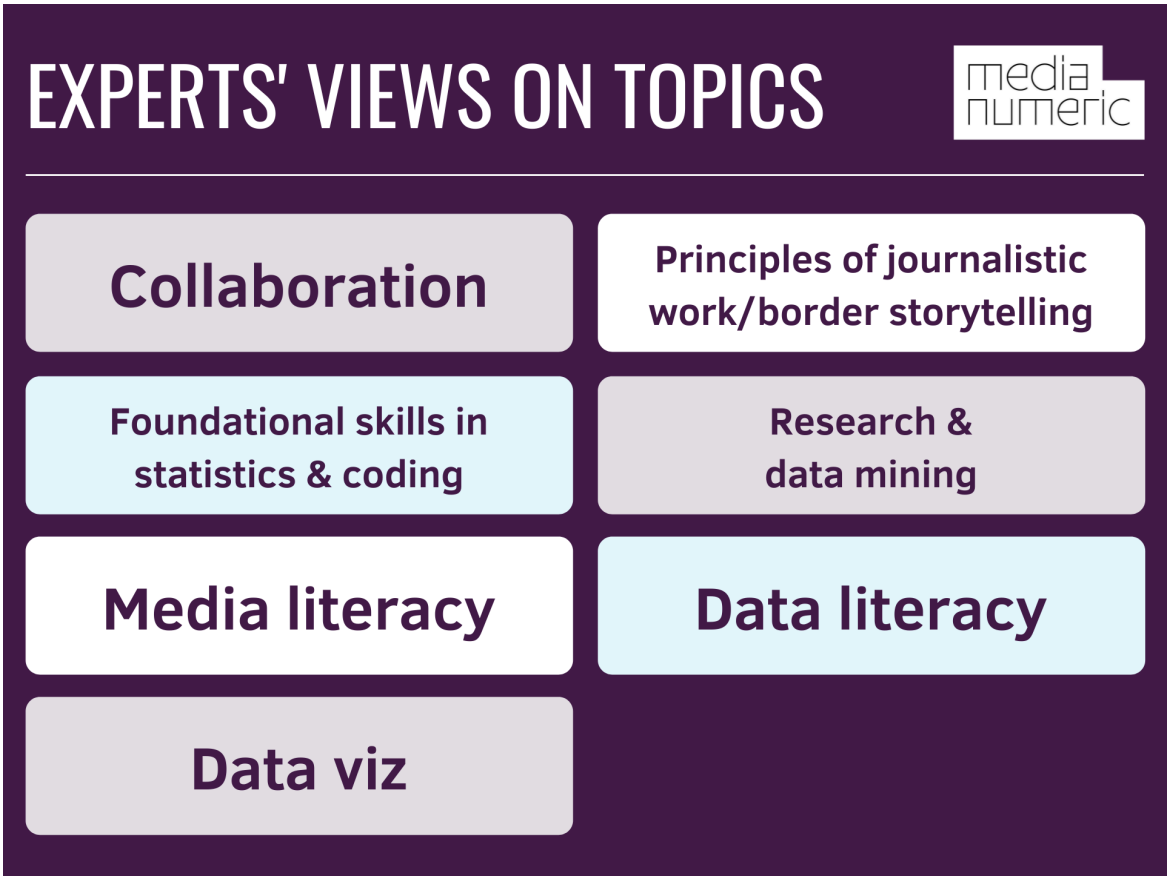


Figure 6. *Summary of experts' views on topics for MediaNumeric.*

4.5. Local Specificity

Analysis of the interviews and the given benchmark examples show that different national contexts require space to 'own' the teaching modules. This will increase the chances of it being used by people in different contexts. The following ideas are relevant for the course developers:

- Different understanding of transparency and accountability in different countries;
- Access to information in different countries;
- Quality & reliability of the information;
- Position of data-journalism in newsrooms;
- A local repertoire of tools and applications (for data manipulation & visualisation; fact-checking): Local Focus (NL).

4.6. The "Shopping List": Tools & Platforms

In the examples given, experts tended to contradict themselves; they both offered countless examples of software packages and of programming languages, only to then insist that these

are all less important than having the right mindset. However, in Figure 7 there is an attempt to systemise advice given on tools and platforms that can be useful while working on a data-driven story. The extended list of tools is offered in *Appendix IV*.

Figure 7. *Summary of mentioned sources and tools for data-work*

4.7. Benchmark Examples

The interviews resulted in an extensive list of benchmark examples. The respondents did not necessarily agree on specific examples, but there are some themes and recurring sources observed. The benchmarks are categorized in the working document:

- Data visualisation
- Data story
- Data investigation
- Collaboration
- (Sources for) fact-checking and debunking

These are not exclusive; An example can be part of multiple categories.

Because of the international nature of this interim report, the examples in this summary will mostly be available in English.

General Observations

Most examples stem from the journalistic field, specifically from news outlets (physical newspapers and online publications). Most commonly, *The New York Times*.

The language of the examples is mostly English, with a bit of French, Polish, Spanish and Russian in the mix. A side note could be made that the interviews were mostly executed in English and all respondents were aware of the international approach of the project. These factors could have influenced some of the answers but, according to Alexandre Léchenet, local journalistic traditions could in part explain the lack of local examples:

> *"There are many interesting examples of this in Anglo-Saxon countries. (…) In France, there are fewer examples of this because there is less of an investigation culture here than in Anglo-Saxon countries."* (Alexandre Léchenet)

4.7.1. Data Visualisation

Graphs and Charts

The bulk of examples regarding data visualisation depict graphs, maps and charts. One notable series is *What's going on in this graph*[38] by *The New York Times*, displaying several different (standard) graphs, maps, and chart types that have appeared in the newspaper. This series is made for students to learn how to look at graphs, what questions to ask of them, and what one can do with the information. More creative and impactful graph examples are The New York Times' front page[39] of March 27th 2020, which depicts an off the charts bar chart, and "Iraq's Bloody Toll"[40] by Simon Scarr in the *South China Morning Post*, which flips a chart to look like dripping blood. One of the most referenced and familiar examples of invoking graphs is the description "flattening the curve", a description for a technique used worldwide aimed at controlling the spread of the coronavirus by levelling off COVID-19 infections.

Numbers without Graphs

Besides graphs and charts depicting number-driven data, there are also alternative ways to make the reader understand through visualisation. For example "A Room, a Bar and a

[38] https://www.nytimes.com/column/whats-going-on-in-this-graph
[39] https://static01.nyt.com/images/2020/03/27/nytfrontpage/scan.pdf
[40] https://www.scmp.com/infographics/article/1284683/iraqs-bloody-toll

Classroom"[41] (also available in French, and Spanish) by El País depicts different scenarios and the chance of becoming infected with the coronavirus.

But what *Is* Data? Data without Numbers

Besides quantitative or statistical data, some examples depict other forms of data. The Twitter Presidency[42], for example, shows recent history through the social media platform Twitter. A beautiful example of a different way to look at data and the question of what data is can be found in One Angry Bird[43], a visual analysis of facial expressions by former U.S. presidents during their inaugural address. The expressions are depicted in a feather-like pattern.

State of the Art Data Visualisation

Besides static visualisations, there are also some examples of moving, interactive, and/or personalized data visualisations. A well-known example is the U.S. interactive election maps[44] with which the audience can visualize the different election scenarios. Out of sight, out of mind[45] is a high-impact, moving visual story regarding drone attacks in Pakistan and their corresponding number of victims. What's my place in the world population?[46] is a webpage depicting the journey of *your* life in numbers. It is completely personalized by the reader by inserting the date of birth, country of birth and sex at birth. How much warmer is your city?[47] operates in the same vein. The audience can learn about the story of global warming using the example of their own city.

4.7.2. Data Investigation

In this category, the respondents would mostly point out specific news outlets and platforms known for their investigative or slow journalism. Besides internationally known newspapers like *The New York Time*s, *The Guardian* and *The Washington Post*, there were also a lot of local examples for France, Poland, Russia, and the Netherlands. They will be included in the extensive list. Between the more specific examples mentioned there seems to be a theme of high-profile cases where journalists have unravelled secrets and scandals by looking at (mostly open-source) databases and recognizing trends and patterns. These are not limited to statistical databases; they use a lot of governmental databases, for example, this French

[41] https://english.elpais.com/society/2020-10-28/a-room-a-bar-and-a-class-how-the-coronavirus-is-spread-through-the-air.html

[42] https://www.nytimes.com/interactive/2019/11/02/us/politics/trump-twitter-presidency.html

[43] https://emotions.periscopic.com/inauguration/

[44] https://abcnews.go.com/Politics/2020-Electoral-Interactive-Map

[45] https://drones.pitchinteractive.com/

[46] https://population.io/

[47] https://www.bbc.co.uk/news/resources/idt-985b9374-596e-4ae6-aa04-7fbcae4cb7ee

article[48] that investigated the political labels of beneficiaries of subsidies; they use social media platforms like Inside the capitol riot: An exclusive video investigation[49]; or they map out public documents, like the Spotlight Investigation: Abuse in the Catholic Church[50], where they examined the history of parish placement of suspected abusers in the Catholic Church to imply complicity.

4.7.3. Data Story

The Data Stories category has a lot of overlap with data investigation and data visualisation. They were categorized differently as the respondents gave them as an example related to storytelling or presentation. These data stories could be used to inspire, and to show the different possibilities of presenting a data-driven story. Two of the examples given are articles by Reuters Investigates (Life in the Camps[51] and Tracking China's Muslim Gulag[52]), in which both are presented with a focus on visual storytelling. A new and exciting way of storytelling is presented in the example of Lost in Europe[53], a project investigating the disappearances of child migrants in Europe. On their "dialogues" page they present interviews with migrant children by way of chatbot. Instead of reading the personal stories in an article, the audience can engage by asking questions themselves, as if they were the interviewer. There also is an example of stories "written" by AI: presented by The Pudding[54], this AI analyses the Twitter messages of members of Congress (U.S.) and presents the leading topics and discourse.

4.7.4. Collaboration

This category also overlaps with data investigation but these examples were named because of their collaborative effort. The #DutchArms[55] investigation is an example of a boot camp in which journalists from different news outlets participated to track Dutch weapons via an open-source investigation of visual data. Furthermore, Spotdata[56] and AFP Medialab[57] were named as examples of multidisciplinary teams.

---

[48] https://www.mediapart.fr/journal/france/100321/argent-public-comment-laurent-wauquiez-arrose-les-siens

[49] https://www.nytimes.com/2021/06/30/us/jan-6-capitol-attack-takeaways.html

[50] https://www.bostonglobe.com/news/special-reports/2002/01/06/church-allowed-abuse-priest-for-years/cSHfGkTIrAT25qKGvBuDNM/story.html

[51] http://fingfx.thomsonreuters.com/gfx/rngs/MYANMAR-ROHINGYA/010051VB46G/index.html

[52] https://www.reuters.com/investigates/special-report/muslims-camps-china/

[53] https://lostineurope.eu/dialogues

[54] https://congress.pudding.cool/

[55] https://medium.com/lighthouse-reports/writeup-dutcharms-investigation-87ed682447cb

[56] https://spotdata.pl/?lang=en

[57] https://www.afp.com/en/agency/medialab

4.7.5. (Sources for) Fact-Checking and Debunking

This category is mostly made up of a plethora of fact-checking platforms, research papers regarding specific themes where misinformation is prevalent, efforts to formalize debunking efforts, books regarding fake news and journalism, and some examples of deep-fakes, like this video[58] "depicting" former US president Barack Obama.

5. Conclusion

This report is a result of the collective effort of data collectors from all MediaNumeric partner institutions and researchers from Inholland. Based on the systematic analysis of the field data (56 in-depth expert interviews) and desk research (review of open information about HE-curricula collected through the web) we can formulate our answer to the research question: *What knowledge and (teachable) skills in storytelling are needed to embolden students of journalism, media, and creative industries to take on the opportunities that data-driven innovations bring?* For reasons of clarity, the main research question was divided into four sub-questions. We will follow this logic for the conclusions as well.

The first two sub-questions were:

- *What are the relevant educational options offered by European higher education institutions (HEI) and commercial enterprises?*
- *How do the experts value the relevant higher education (HE) offerings in their countries?*

The regular HE offers within the programmes in Creative Business, Communication and Journalism in France, the Netherlands and Poland primarily serve the interest of students who want to specialize in a particular (narrow) application of data in media and creative practices. Knowledge and skills related to storytelling with data are scattered throughout courses that are usually focused on specific aspects of digital technologies in a specific (narrow) domain. There is no universal course oriented to all creative students that address the demand for basic knowledge and skills in telling stories with data. The experts were critical when asked to share their opinion about the relevant regular HE offer (they noted poor quality, 'amateurish', fragmented content and lack of data-related courses in the curricula). The shared idea was that a data-related course needs to be developed, and better sooner than later.

[58] https://www.youtube.com/watch?v=cQ54GDm1eL0&ab_channel=BuzzFeedVideo

Besides the educational options in data-driven storytelling offered by conventional HE institutions, there is a plethora of commercial training courses, workshops, etc. offered by regular universities, schools, and private bureaus. Usually, these programmes have a broad postgraduate professional group as their target audience.

In addition to these, extensive data- and storytelling-related online commercial courses are available, as well as free Massive Open Online Courses (MOOCs). For those who don't want to pay or do not have the means, there are free courses in data and its applications in creative work. Also, large media enterprises such as Dutch broadcasters RTL and NOS, Britain's *Daily Telegraph*, national broadcaster the BBC, and many other mass-media companies provide their employees with up-to-date professional on-the-job training, in which data and its use are becoming more relevant.

The third sub-question was:

- *What are the essential knowledge, skills and tools (for the search and exploration of data - storytelling - checking and debunking misinformation) in their professional fields, according to experts?*

As the MediaNumeric project description formulated its major modules in some detail, the interview questions and discussion with the experts were modelled to reflect the same three topics: 1) search and exploration of data; 2) tracking and debunking misinformation and 3) telling stories with data.

Regarding the content of **Module I** (search & exploration data), the experts agreed on the following focus points for such a course:

- Basic journalistic/storytelling skills, principles etc. are not new but still essential in data journalism and data-driven storytelling;
- Data literacy: overcoming the fear of numbers & technology; basic statistics; data manipulation skills; create common ground with tech professionals;
- Tools & databases: basic principles and choice of apps.

While developing **Module II** (tracking & debunking misinformation) we, the MediaNumeric project, were advised to consider:

- Making critical thinking an essential part of the programme. Teach/let students learn to ask questions. What is ‘truth’?;
- Confirming deontology/ethics as the basis of the journalistic profession;
- Using open sources and initiatives such as FactCheck.org, Poynter.org, Snopes.com

For the creators of **Module III** (telling stories with data), the following ideas expressed by the experts are worth keeping in mind:

- Basic skills are important. Like the other steps of the storytelling process, the presentation and reporting (writing, telling etc.) phase demand the mastering of ABC knowledge about the audience, genres, psychology of text, colour and image perception;
- Numbers & statistics always need explaining and contextualization (in line with academic research: Corner, 2021);
- Collaboration between different content creators is needed – this request is applicable for all units concerned with storytelling with data, also in the last stage of journalistic/creative work, i.e. when creating the final product.

In every interview, the question of **tools** was raised and discussed. The experts emphasized the importance of having simple and *versatile* tools. In this regard, the following should be kept in mind:

- There is currently a lack of tools in universities of applied sciences;
- Priority must be given to tools that do not require coding skills;
- Students must (learn to) test new tools and question themselves constantly;
- Tools must be easy and fun to use.

According to our informants, the proposed MediaNumeric course should focus on breaking down essential components of data retrieval – that is, a *deconstructive approach*. Emphasis was also placed on *being creative in your approach to data* – tailor the course to the creative skills of the students and *tie it to storytelling.*

Analysis of the whole corpus of the interview transcripts shows that there is no obvious consensus among the experts about the exact topic list for our teaching programme. Yet, a few components were emphasized by most of the respondents:

- Basic, "universal", non-data- or computer-driven skills in storytelling (also for visualisation). Students should be shown how these skills apply to data-driven storytelling;
- Ability to collaborate with professionals from other fields/establish "common ground";
- Understand principles of human-computer interaction/computational thinking;
- Basic statistics.

The course should be built upon the following key elements: *respect for basic storytelling skills*; *collaboration*; *critical thinking*; *real-life, small-scale projects*; *clear frameworks* (time, goals, templates etc.); *tools*: Google sheets or other spreadsheets. The rest of the choices should be made after defining the target group of students for the course.

The following competencies related to data collecting were considered by the experts as most valuable for future story creators: the ability to *read, understand and interpret statistics*; the ability to *explain numbers to a specific audience*; the ability to *manage data capturing tools*.

In sub-question four we asked:

- *What is the best way to teach (learn) these skills in a higher education setting, according to experts?*

Were we to sum up the experts' ideas about a 'dream format' for our course, it would be a list with the following recommendations:

- Use a multidisciplinary approach to the content, assignments and the teaching team (coaches, guest lecturers from different fields);
- Start from basic storytelling skills (first the basis should be in order – knowledge and skills in finding the story, collecting data through "old school" investigation: interviews, searching for documents, etc.; verification and presentation);
- Organise a course around real-life cases, use small-scale data sets, let students collaborate with professionals, show them the logic of working with data as a process and include a playful element in the course structure.

Experts shared with us a few specific recommendations regarding the didactic side of the planned course in storytelling with data. The most prevailing advice was to be *practice-oriented.* In sum, this means: to use real-world, small-scale cases and questions; work on projects/concrete examples related to news or students' daily life; learning by doing; study visits; simulate the work field by attending to the following four pillars: 1. practical/hands-on projects, 2. set in real and tangible contexts/topics, 3. fast production (like in a newsroom), 4. work in (interdisciplinary) collaborations (no one can do it all). We were advised to teach students to deal with time pressure.

The practical approach is not so much about learning every skill but about *becoming familiar with the process*. According to the experts, this will make it easier to imagine possible approaches/solutions and collaborate with other disciplines. In short, this will help students to know where to look, and to understand what is possible and what is not. In line with the 'real world' approach, the experts insisted on strong "best practice" examples, which should

be incorporated into all elements of the training program. Different experts mentioned that the ‘play’ element incorporated in lessons and assignments would be effective.

The experts insisted on *collaboration* as an essential requirement for data work in a storytelling business. This is a big theme across all interviews. The experts said that no one can do it all alone, you need to collaborate with people, especially if you want to make new innovative formats. But you need to learn *how* to work together with people from other disciplines. That is, you need to learn what to ask for, how to speak their language (understand the need for common ground), and how to manage a collaborative project.

Working on collaborative projects can lead to *reducing fear of numbers and technology* (this fear is often named as one of the main reasons that many creative professionals are reluctant to engage in data work).

The experts were concerned about the time restrictions that we will have for the MediaNumeric training – and that we need to think about how we will work through the extended topics list in one week's time. Choices should reflect the profile of the target group. After the series of consultations, all partners agreed upon the following profile of the target group: the 3rd - 4th- year BA students and first-year MA students in media, journalism and creative industries curricula.

To sum up the needs analysis results, we would like to note that the decision to conduct expert interviews as the methodology of this research was well formed. The insights gathered through the interviews defined a solid basis for the direction of the training programme’s content, format and approach. In addition to forming the basis of the course, the conversations and connections made with these interviewee experts have also been a source of inspiration, continuing to enthusiastically illustrate the importance and need for such a course in data-driven journalism, creative storytelling and fact-checking.

6. References

Corner, J. (2021) Figuring Covid. A note on media flows and 'statistically'. *Media Theory.* http://journalcontent.mediatheoryjournal.org/index.php/mt/article/view/131

Barker, C. (2004). *The Sage dictionary of cultural studies*. Sage.

Hogan, P. C. (2016). *Affect Studies*. In Oxford Research Encyclopedia of Literature.

Glaser, B. G., & Strauss, A. (1967). L.(1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The discovery of grounded theory; strategies for qualitative research. *Nursing Research*, 17(4), 364.

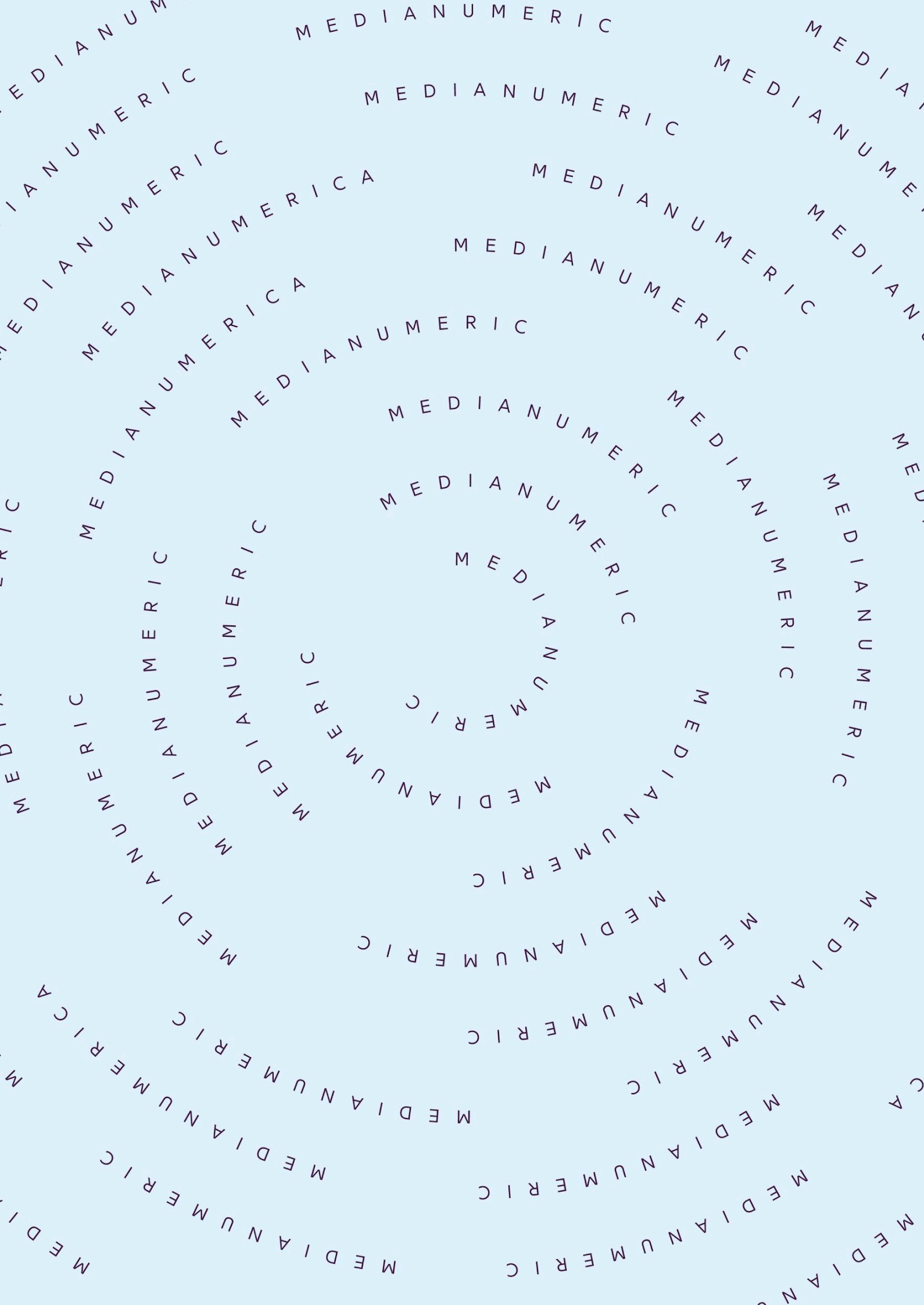
Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, 2(163-194), 105.

Lofland, J. & Lofland, L.H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont, CA: Wadsworth.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*. Sage publications.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Sage Publications, Inc.

Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research and applications,* 5, 147-158.


MEDIANUMERIC