The research leading to these results has received funding from the European Commission under grant agreement No. 621610-EPP-1-2020-1-NL-EPPKA2-KA.

D3.1 - First Version of Teaching Modules

Version 1.0 (*Final*)

June 15, 2022

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

Grant Agreement number:	621610-EPP-1-2020-1-NL-EPPKA2-KA
Project acronym:	MediaNumeric
Project title:	MediaNumeric: Alliance on Multimedia Data-Driven Journalism & Media Production
Funding Scheme:	ERASMUS+ KA2: Cooperation for Innovation and the Exchange of Good Practices - Knowledge Alliances
Project coordinator name, Title and Organisation:	Johan Oomen, The Netherlands Institute for Sound and Vision
E-mail:	joomen@beeldengeluid.nl
Project website address:	http://medianumeric.eu/

Revisions

Version	Date	Author	Reviewer(s)	Changes
0.1	28.04.2022	Kuba Piwowar	Natalia Berger	
1	09.06.2022	Kuba Piwowar	Jacqueline Pietsch	First review. Adds details of study visit
1	15.06.2022		Rachel Somers Miles	Final review for publishing

Statement of Originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1. Executive Summary

This deliverable describes the activities carried out during the first iteration of a Winter School of the MediaNumeric programme held in Paris in February 2022, Work Package 3 (WP3) by the eight consortium partners and describes the results achieved by this work package. It outlines how the programme's structure reflects both the Needs Analysis and State of the Art reports' findings, refers to all materials used during the Paris programme and the idea behind the case study. The case study was the backbone of the entire programme: during the Winter School the students worked on the topic of migration, using knowledge and tools they were provided during classes. As a result, each group created their own 'pitch' presented to the consortium representatives and fellow students, based on a topic of their choice that referred to a broader theme of migration. Each lecture and workshop was designed in a way that provided the students with skills necessary to complete all tasks, starting from data gathering and analysis, through storytelling and visualisation, and finishing with debunking disinformation exercises.

2. Introduction

This deliverable consists of a detailed description of the Winter School that was held in Paris in February 2022. It entails a full syllabus for all three modules of the course in data journalism and digital verification, descriptions of all lectures and workshops, links to accompanying presentations, data sets, handouts, and additional materials that were provided to students during the six-day programme. It also includes the three quizzes that were designed to verify knowledge gained by the students throughout the course, as well as links to the final presentations of the case studies that students put together during the week. The document also outlines the study visit to Agence France-Presse (AFP). The final section details the consortium's plan for recording the classes with video and outlines options being discussed about the digital format of this training course. The report concludes with detailed notes taken during a debriefing among consortium partners on the final day of the Winter School. It consists of guidelines for further iterations of the learning programme.

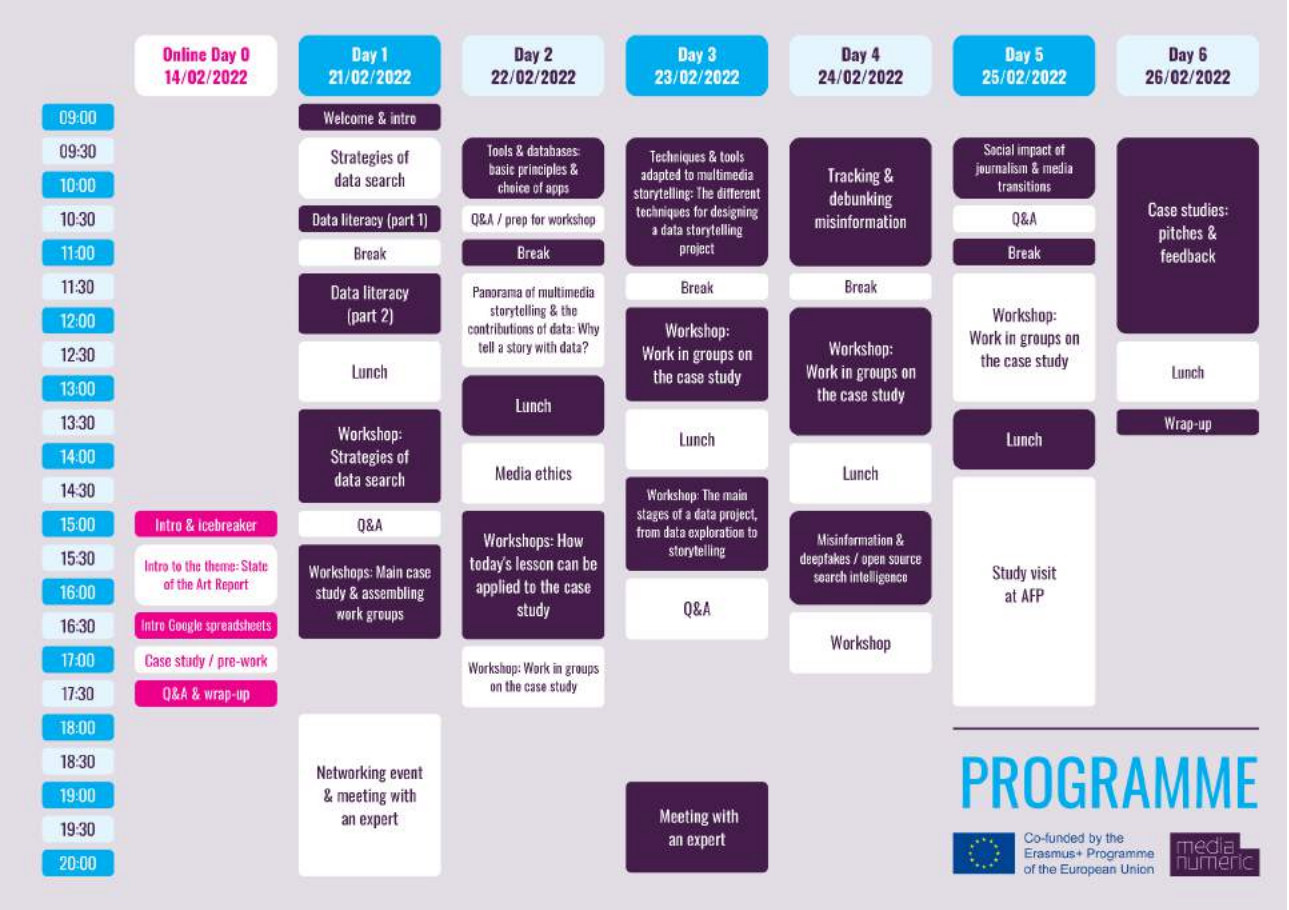


Figure 1. Detailed programme of the Winter School

3. Detailed Syllabus

The final syllabus was constructed as an iterative process, started in mid-2020 (in M6) by making an assessment of which parts of the entire curriculum will be useful and how they can be incorporated into the programme. Crucial parts of the syllabus creation were also the Needs Analysis and the State of the Art reports. The programme was based on findings provided by extensive desk and field research. Hence, assumptions made in the project description about the topics were not entirely supported by the needs assessments' results. Among the decisions made after these analyses, we decided to target students with no or basic data skills, as well as that we searched for teachers who felt comfortable teaching in such a setting. The initial programme description was juxtaposed with these findings, resulting in a programme reflecting these findings. As a result, a set of three syllabi was created, one for each module, and they were then merged into one coherent document. The below section covers all of the contents of the syllabus, but if you would like to see the actual syllabus document, you can view it here.

3.1. Brief Description of the Course

The aim of this course is to provide the students with the basic ways in which data can be used in journalism and creative storytelling, with concrete examples of tools used in data gathering, analysis and visualisation. Since part of the process is making sure no misinformation is spread, part of the curriculum will be devoted to strategies of combating it. The programme is divided into three modules.

The first module ‘Search and exploration of web data’ revolves around real-life small-scale data set cases, and explains the process of finding, interpreting and using data. Its goal is not only to introduce students to working with data but to inspire them to continue to do so. It is based on an assumption that basic journalistic/storytelling skills and principles are essential to data journalism and data-driven storytelling. The course provides the basis for acquiring basic data literacy skills and helps overcome the concerns created by numbers and technology.

The training module ‘Telling stories with multimedia data’ consists of courses providing the participants with the basic skills needed to use data sets as part of stories used in journalism, as well as show how different tools can help in data visualisation. This module combines theoretical contributions in the field of data storytelling with practical exercises.

The third module, ‘Tracking and debunking misinformation’ aims to provide the students with an overview of how misinformation and disinformation fit into a broader reflection of media, with examples of methods used in debunking them and accompanying exercises to practice these techniques, and work on a framework that can be used in the students’ further projects.

The programme is closely intertwined with a case study that the students are required to work on in small groups in order to put the ideas, tools and skills presented in the lectures into practice. The course includes a study visit to a consortium partner's premises.

3.1.1. Objectives of the Course and Learning Goals

The following outlines the objectives and learning goals of the MediaNumeric training course:

- Students will be able to apply basic skills in finding a story and information to data-driven storytelling;
- Students will be able to find relevant data sets and extract the relevant information for a story;
- Students will be able to collaborate with professionals from other fields and establish 'common grounds' while creating a data-driven story;
- Students will understand the basic principles of human-computer interaction/computational thinking;
- Students will be able to find and extract relevant data from digital databases (data capturing skills);
- Students will understand the concept of storytelling and its evolution in the digital age;
- Students will understand the implication of data in storytelling and its recent uses;
- Students will be able to read, understand and interpret basic statistics, as well as use basic statistical tools (ie. Google Sheets);
- Students will understand the potential of the most common data visualisation tools;
- Students will be able to implement basic data storytelling projects;
- Students will have general knowledge of the contemporary media ecosystem, its role for democracy and challenges it faces;
- Students will have specific knowledge on misinformation in media, including different kinds of fake news, media specificity, and emergence of new technologies and tools to disseminate them;
- Students will have specific knowledge on the social obligations of the journalism profession, its responsibilities and ethics;
- Students will become acquainted with the contemporary journalist's toolbox and procedures to organise their work;
- Students will be able to spot misinformation and distinguish it from facts;
- Students will be able to use a framework to debunk fake news;
- Students will develop a more critical attitude, become better informed and more responsible professionals and citizens.

3.1.2. Prerequisites, Admission Requirements & Procedures

Students involved in the MediaNumeric training course are expected to meet the following criteria:

- English proficiency at least at B2 level with capacity to read texts in English and participate in group discussions;
- Complete pre-work and reading assignments prior to the course (the students will be given the full list of materials to get acquainted with during the Day 0 introduction);
- Participate in the Day 0 introductory session.

3.1.3. Class Rules, Assessment Criteria & Methods

Students involved in the MediaNumeric training course are expected to adhere to the following:

- The students will be asked to work in small groups (4-5 people) around a case study presented during the Day 0 session. The task is to work on the case and present the results to the entire cohort and tutors at the end of the week;
- The students are expected to attend all classes and activities in person and bring their own laptops (if possible);
- There will also be a set of small analytical tasks to check the students' understanding of the contents covered throughout the course. At the end of each module, participants will be assessed to ensure that the knowledge has been assimilated. These evaluations will take place in the form of a quiz with 30 questions (1 point per question), to be completed over a period of 30 minutes. Participants who obtain an average of 50% correct answers on the quiz - or 45 points out of the 90 points in total - will pass the training course and obtain a certification.

3.2. Course Contents

This part consists of detailed information covering each of the three Modules, outlining details of each course.

3.2.1. Day 0 Online Introductory Session - Pre Winter School Class

Day 0 took place on 14th February 2022, a week before the Winter School in Paris. The goal of this online session was for the students to get to know the MediaNumeric team and each other, as well as get acquainted with the basic premise of the learning week. The students had a chance to listen to the results of the State of the Art report, participated in a short workshop in using Google Sheets, and heard about the idea behind the programme, including the case study, which was introduced during the meeting.

The three hour long meeting consisted of the following sections:

1. Introduction to the MediaNumeric Programme
2. Introduction to the topic of data journalism: State of the Art report
3. Google Sheets hands-on

4. Introduction to the case study

3.2.2. Winter School in Paris

The Winter School took place between 21st and 26th February 2022. The MediaNumeric team, accompanied by the teachers, led the programme throughout six days, detailed below.

3.2.2.1. Module 1: Search & Exploration of Multimedia Data

The below details the topics and related content covered across Module 1: Search & Exploration of Multimedia Data.

Strategies of Data Search

- Short history of data-driven publications
- Why working with data matters
- How to find data: basic knowledge of France's legislation on the issue
- How to find data: how to use Open Data portals
- How to find data: how to scrape the web
- How to treat data: basic knowledge of spreadsheets
- Visualisation: the main types of data visualisations
- How to choose a relevant type of chart

Data literacy: People as Data Points

- What is data literacy
- Artificial Intelligence & Machine Learning
- Machine Learning & NLP: an experiment
- Big Data and social media
- Social media monitoring: the echo-chambers

Data literacy: How to Deal with Numbers. Google Sheets

- Computational thinking (the essence of the human-computer interaction; what can you expect from a machine and what not?)
- Data exploration, basic statistics, basic maths
- Data manipulation skills, filtering & cleaning
- Understanding potential biases

Stages of a Data Project

- From ideas to data
    - Motivation
    - Resources
    - Collect
    - Aim + Context = Research Question

- From data to stories
    - Tidy up
    - Understand
    - Communicate
- Data workflow: programming vs. clicking ⇒ advantages and disadvantages of both approaches
- Science vs. journalism

Getting Data through Web Crawling

- Definition of web crawling
- An example of a crawling process
- Using this example to explain the two possible drawbacks of crawling
- An example of a crawling process with the use of the Seealsology tool

Getting Data through Web Scraping

- Definition of web scraping
- Scraping data automatically from the source code i.e. Desmog
- Google bookmarklets tool
- Minet twitter scrape tool
- The limits of web scraping
    - Problem of copyright
    - Problem of collecting personal data (GDPR regulation)
    - Problem of the scraping difficulties implemented by websites that do not want to be scraped

3.2.2.2. Module 2: Telling Stories with Data

The below details the topics and related content covered across Module 2: Telling Stories with Data.

Panorama of Multimedia Storytelling & the Contributions of Data: State of the Art of Storytelling Multimedia. Why Tell a Story with Data?

- Why and how do we use data in telling a story?
- Introduction: why tell a story with data? Data and data visualisation as a rhetorical tool to demonstrate and illustrate a point.
- What is storytelling with data? Basis of storytelling = why and how

Techniques & Tools Adapted to Multimedia Storytelling: The Different Techniques for Designing a Data Storytelling Project

- The actors involved: journalist/copywriter/author, data analyst, data designer, designer/illustrator, photographer, developer, motion designer, producer, etc.
- The tools involved: pen and paper, Word, Excel, coding for data-analysis (R, Python, D3, etc.), design software, visualisation software, coding for web (Python, HTML, js, etc.), animation software, etc.
- Data visualisation tools: free data visualisation software (Flourish, rawgraphs, Kartis, Tableau, etc.)
- Introduction to Flourish

3.2.2.3. Module 3: Tracking & Debunking Misinformation

The below details the topics and related content covered across Module 3: Tracking & Debunking Misinformation.

Media Ethics

- Introduction to the ethical challenges of journalism
- Legal regulations, professional ethics and codes of conduct
- Problematic cases and examples showcasing different ways in which media ethics guidelines were breached
- New forms of unethical practices connected to digital media and data manipulation
- Ethical decision-making models
- Big data, data visualisation and its risks

Confronting the Misinformation Universe

- The misinformation universe: definitions, origins, dynamics and environment. Anatomy and itinerary of false information: case study
- Data and science in a fast-changing world: popular examples of misinterpretation of data
- Why do we fall for false or misleading information
- Basic rules of verification: the journalist's 5W toolkit
- Out of context: how to debunk photos and videos shared online

Fake News & Digital Manipulation/Open Source Intelligence

- Digital manipulation: how to discover and verify a photomontage, a doctored video, or tinkered sound in social media content
- Geolocation: how to find any place on Google Earth and why is it important in fact-checking (Suncalc, Google Earth)
- Advanced Search: Google + Twitter (Crowdtangle)
- Archiving as an accountability tool: how to fight cowardice and prevent link rot
- Deepfakes phenomenon: are they a future of disinformation or are they irrelevant? Tools and materials, Deep Fake Detection Challenge overview (source 1 + source 2)

- Monitoring: creating lists of the usual suspects (Crowdtangle, BuzzSumo)
- Underbelly of the Internet: dummy profiles, infiltrating closed communities and learning the language of the 'other' tribe (Closed FB groups, Telegram, Whatsapp)
- Other tools for Fighting Disinformation: InVID plugin, MapChecking, other open-data sources

Social Impact of Journalism & Media Transitions

- Consequences of the internet's platformisation and new business models on the media ecosystem and democracy
- Democratisation of the access to the audience, especially social media: new media practices and experiences (media formats mix)
- Internet fraud, social antagonisms, depreciation, false positive beliefs, algorithmic inclusion/exclusion
- Automated journalism (impact on the news)
- Social consequences of new media models
- User generated content and its influence on media streams

3.3. Readings

The students were asked to read two texts listed below - and these were required for the course. Other reading materials, listed under the "Suggested reading" section, were meant to support the classes and were suggested to the students who wished to further their knowledge.

3.3.1. Pre-Reading

- Dario Compagno, 2016, Families of practices. A bottom-up approach to differentiate how French candidates made use of Twitter during the 2014 European Campaign, in: "Tweets from the Campaign Trail: Researching Candidates' Use of Twitter During the European Parliamentary Elections" (Alex Frame et al., ed), pp. 33-52.
- Jacqueline Pietsch and Daniel Sorabji, 2021, State of the Art report in storytelling with data and misinformation.

3.3.2. Suggested Reading

Suggested reading in Module 1: Search & Exploration of Data

- Ashley Fell, 2017, Why storytelling is so powerful in the digital era, in: YouTube/TED.
- Lisa Raehsler, 2013, What people search for – most popular keywords, in: SearchEngineWatch.com.
- Jacquelyn Bulao, 2022, How Much Data Is Created Every Day in 2022?, in: Techjury.net.
- Brahim Zarouali, 2020, Persuasion Effects of Psychometric Targeting and Chatbots, in: Social Media & Politics.
- Chantel Ridsdale, James Rothwell and Mike Smit, 2015, Strategies and best practices for data literacy education: Knowledge synthesis report.

- Guillaume Plique, 2020, Empowering social scientists with web mining tools, in: FOSDEM 2020.
- Facebook offers a distorted view of American news, in: The Economist.
- Facebook's Top 10, a Twitter account.

Additional reading in Module 1: Search & Exploration of Data

- C. Dawson, C., 2020, A–Z of Digital Research Methods. Routledge, Chapter 23.
- N. H. Riche, C.Hurter, N. Diakopoulos and S. Carpendale (Eds.), 2018, Data-driven storytelling. CRC Press, pp. 59-85.
- A. Khan and H. Din, 2021, New Age Journalism and Big Data (Understanding Big Data and Its Influence on Journalism). Big Data Analytics for Internet of Things, pp. 333-349.
- S. Knowles, 2018, Narrative by Numbers: How to Tell Powerful & Purposeful Stories with Data. Routledge. pp. 1-67.
- I. Foster, R. Ghani, R. S. Jarmin, 2020, Big Data and Social Science, 2nd Edition. Routledge. pp. 23-34.
- B. Franklin, S. Eldridge II (Eds.), 2016, The Routledge companion to digital journalism studies. Taylor & Francis, (Part V)**.**
- R. Rogers, 2019, Doing digital methods. Sage, pp. 1 - 41.

Suggested reading in Module 2: Telling Stories with Data

Visual Story-Telling

- Life in the Camps and Tracking China's Muslim Gulag, by Reuters Investigates.
- Lost in Europe.
- The Pudding.

Dataviz

- Graphs & tabs
  - What's going on in this graph and The New York Times' frontpage of march 27th 2020, by the New York Times.
  - Iraq's Bloody Toll, by South China Morning Post.
- Numbers without graphs
  - The Twitter Presidency, by the NY Times
  - One Angry Bird, by Periscopic.
- Data Visualisation state of the art
  - The U.S. interactive election maps, by ABC news.
  - Out of sight, out of mind.
  - What's my place in the world population?
  - How much warmer is your city?, by the BBC.
- Data Visualisation
  - Can Visualization Elicit Empathy? Our Experiments with "Anthropographics".
  - Connecting with the Dots.
- Pragmatism
  - Collected Papers 5.9, by Charles Sanders Peirce.

Suggested reading in Module 3: Tracking & Debunking Misinformation

- S. Zuboff, 2018, , The Age of Surveillance Capitalism, Profile Books.
- J. van Dijck, T. Poell, M. de Waal, 2018, The Platform Society: Public Values in a Connective World, Oxford University Press.
- T. Gillespie, 2014, The Relevance of Algorithms, in: Media Technologies, ed. T. Gillespie, P. Boczkowski, K. Foot, MIT Press.
- W. Chun, 2008, On "Sourcery", or Code as Fetish, "Configurations", vol. 16, no. 3.
- W. Chun, 2021, Discriminating Data, MIT Press.
- K. Crawford, 2021, Atlas of AI, Yale University Press.
- Jacob Soll, 2016, The Long and Brutal History of Fake News, in: Politico.
- Francesca Tripodi, 2018, Searching for Alternative Facts. Analyzing Scriptural Inference in Conservative News Practices, in: Data & Society, pp. 18 - 48.
- Bruce Schneier, Allie Wong, Samantha North and Mick West, 2021, The Battle for Truth: Disinformation, Misinformation, & Conspiracies, in: The CyberWire.
- The commitments of the code of principle, in: IFCN Code of Principles.
- Kristen Panthagani, 2022, 10 logical fallacies used in vaccine arguments, in: You can know things.
- M. West, 2020, Escaping the Rabbit Hole, How to Debunk Conspiracy Theories Using Facts, Logic, and Respect, Chapter 4.

Additional reading in Module 3: Tracking & Debunking Misinformation

- Niam Yaraghi, 2019, How should social media platforms combat misinformation and hate speech?.
- Craig Silverman (ed.), Verification Handbook A Definitive Guide To Verifying Digital Content For Emergency Coverage, in: DataJournalism.com.
- Tommy Carl-Gustav Linden, 2017, Algorithms for journalism: The future of news work, in: The Journal of Media Innovations.
- Beate Josephi, 2016, Digital Journalism and Democracy, in: Digital Journalism, Sage, pp. 9-24.
- A. K. Schapals, A. Bruns and B McNair (eds.), 2019, Digitising Democracy, Routledge.
- Reuters' report on deep fakes, in: Reuters.
- Jonathan Albright, 2017, FakeTube: AI-Generated News on YouTube, in: Medium.
- Riana Pfefferkorn, 2020, Deepfakes in the Courtroom, in: Boston University Public Interest Law Journal. Vol. 29 Issue 2, pp. 245-276.
- Oscar Schwartz, 2018, You thought fake news was bad? Deep fakes are where truth goes to die, in: The Guardian.
- Lane Wilkinson, 2016, Post-truth, propaganda, and bullshit: a glossary, Sense and Reference, in: Sense and Reference.
- Giselle Rampersad and Turki Althiyabi, 2019, Fake news: Acceptance by demographics and culture on social media, Journal of Information Technology & Politics, 17:1, pp. 1-11.
- Cristian Vaccari and Andrew Chadwick, 2020, Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News, in: Social Media + Society. January 2020.
- Julia Angwin (ed.), 2022, The Markup Splitscreen project, in: The Markup.

- Kevin Roose, Rabbit Hole podcast, in: The New York Times.
- Julia Bayer and Ruben Bouwmeester, 2022, Fact check: How do I spot a deep fake?, in: Deutsche Welle.
- AFP, AFP Code of Ethical Standards, in: AFP.
- Reuters, Reuters Standards and Values, in: Reuters.
- Brian Walski, 2003, Bronx Documentary Center, Altered Images, in: Altered Images BDC.
- Council of Europe Resolution on the Ethics of Journalism (1994), in: Accountable Journalism.
- Jean-François Furnémont and Tanja Kerševan Smokvina, 2017, European Co-Regulation Practices in the Media, in: The Council of Europe.
- Ethical journalism practices on migrants and refugees, in: European Journalists.
- Georgia Wells, Jeff Horwitz and Deepa Seetharaman, 2021, Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show, in: The Wall Street Journal.
- Damien Gayle, 2021, Facebook aware of Instagram's harmful effect on teenage girls, leak reveals, in: The Guardian.
- The Open University, 2017, The Potter Box, in: OpenEdu.
- Danah Boyd and Kate Crawford, 2011, Six Provocations for Big Data, in: A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society.
- Catherine D'Ignazio and Lauren F. Klein, 2020, Data Feminism, The MIT Press.
- Michael Schermann, 2019, A Reader on Data Visualisation, in; Github.
- Peter Haferl, 2019, The Ethics of Media Visualisation, in: Medium.
- Vladan Joler, 2021, New Extractivism, in: OpenSecret.

4. Course Materials

Below provides a list of all materials used during the Winter School classes. Each graphic is an interactive link that leads to the file of the presentation online.

4.1. Background Materials

4.1.1. Master Deck: Day 0

This document is a presentation used during an introductory meeting held online on 14th February 2022. The meeting's aim was to introduce the students to the MediaNumeric team, as well as presenting them with details of the Winter School's programme.

Figure 2: Master deck Day 0

4.1.2. Master Deck: Paris Winter School

The master deck consists of slides used during the Winter School as back-up material, with all necessary information, ie. the programme for a given day or planned breaks.

Figure 3: Master deck Paris Winter School

4.2. Lectures

4.2.1. Strategies of the Data Search (Lecturer: Gary Dagorn)

This course offers a quick introduction on how to find, process and use online data when writing a data-driven story. It includes learning basic knowledge of search techniques, Open Data and the legislation surrounding it, as well as some examples of how to format/process/visualise them to add informative value to your story.

Figure 4: Slide 1 Strategies of the Data Search

4.2.2. Data Literacy (Lecturer: Gijs van Beek)

This lecture covers topics such as what is data literacy, artificial intelligence and machine learning. It also covers tools like Textgain and an experiment in machine learning and natural language processing using big data and social media. Parts of the lecture will cover social media monitoring and, in particular, echo-chambers.

⇒ Handout for the lecture here

Figure 5: Slide 1 Data Literacy

4.2.3. Tools & Databases: Basic Principles + Choice of Apps (Lecturer: Héloïse Théro)

The internet is a powerful medium from which all kinds of data can be extracted. The class takes on answering the question how web crawling can be used to visualise the relationships between different websites, and thus build a 'web of knowledge' on a specific topic. The students see that we can extract specific data from web pages by using web scraping. For example, you can easily download the results of a Google search, or all the tweets containing a specific hashtag or keyword. Finally, the students get to know how different databases can be accessed on the web, such as the metadata of YouTube videos. There are no coding skills necessary to follow the class, the goal is only to acquire a general culture about web data and how it can be extracted.

⇒ Lecture outline here
⇒ Handout for the lecture here

Figure 6: Slide 1 Tools & Databases: Basic Principles + Choice of Apps

4.2.4. Panorama of Multimedia Storytelling and the Contributions of Data. State of the Art of Storytelling Multimedia. Why Tell a Story with Data? (Lecturer: Julie Brunet)

Why and how do we use data in telling a story? These are the two main questions we'll try to answer in this lecture by focusing on a quick panorama of multimedia storytelling and by discussing examples of successful and innovative uses of data in the newsroom.

⇒ Lecture outline here
⇒ Handout for the lecture with additional materials here
⇒ Datasets used during the class here

Figure 7: Slide 1 Telling Stories with Data

4.2.5. Workshop: The Main Stages of a Data Project, from Data Exploration to Storytelling (Lecturer: Dario Compagno)

In order to produce a data analysis one has to move through several steps. Firstly, one needs to understand the context of the analysis: what are the needs grounding it? How can these needs be translated into research questions, and then into data collection? This requires intuition and clear ideas about our aims and means. Once we have this sorted out, the practical work begins. Data has to be treated for visualisation and modelling. It's here that software knowledge becomes central. In the end, our work has to be brought to the relevant public, so graphical and writing skills should now lead the way. This module will try to identify the most important points to check at each of these steps.

Figure 8: Slide 1 Strategies of the Data Search

4.2.6. Techniques and Tools Adapted to Multimedia Storytelling. The Different Techniques for Designing a Data Storytelling Project (Lecturer: Julie Brunet)

In this lecture, the students consider all of the actors involved in a data storytelling project and their mutual tools before focusing on Flourish, a free online data visualisation tool. The students get the chance to put their knowledge into practice using the aforementioned tool and data sets they started gathering to complete their case study assignment.

⇒ Lecture outline here

Figure 9: Slide 1 Storytelling with Data: Tools

4.2.7. Media Ethics (Lecturer: Mirosław Filiciak)

The Media Ethics course is devoted to ethical dilemmas related to journalistic work - with tools to solve them and examples of good practices. The focus of the class is aimed at challenges that have arisen with the transformations of the media ecosystem, ranging from the functioning of classic problems in the environment of blurred categories, and ending with specific issues related to Big Data and data visualisation.

Figure 10: Slide 1 Media Ethics

4.2.8. Confronting the Misinformation Universe (Lecturer: Marion Dautry)

What is disinformation and misinformation? How do you recognise it? Where does it come from? How does it circulate, particularly on social media? Students will gain an understanding of a trend that has been identified as one of the biggest threats to democracy. They will learn where to start the process of verification and the basic tools to debunk manipulated content.

Figure 11: Slide 1 Confronting the Misinformation Universe

4.2.9. Social Impact of Journalism and Media Transitions (Lecturer: Mirosław Filiciak)

The Social impact of Journalism and Media Transitions course systematises knowledge about the social consequences of Internet proliferation. Its theme is the emergence of new media structures and forms, but also the appearance of new intermediaries in the distribution of information, with growing importance of the technological component, which changed the way the public sphere functions.

Figure 12: Slide 1 Social Impact of Journalism and Media Tranistions

4.2.10. Debunking Dis/Misinformation and Digital Manipulation (Lecturer: Robert Barca)

A deep dive into open-source tools that can be used to investigate, track, and debunk dis/misinformation and digital manipulation. Students will learn advanced search criteria, geolocation, image analysis tools and other techniques to identify manipulated content. They will also learn the importance of accountability and how to record their investigation so that their work stands up to scrutiny. During the class the students will have a chance to use the inVid tool that will help them in their debunking work, including the one that refers to their case study projects.

⇒ Handout for the lecture here
⇒ Fact-checking exercises here

Figure 13: Slide 1 Debunking Dis/Misinformation and Digital Manipulation

4.3. The Case Study and the Pitching Session Materials

4.3.1. The Case Study

Below is the detailed description of the case study that the student groups worked on. It was presented during the Day 0 online meeting, and was a reference point for the students in their work on their pitches. The case study document is available here.

Syrian refugee children play beside tents in Domiz refugee camp. Source: © UNHCR/B.Sokol

Context

Among a variety of important topics that fill the public discourse, but also a great social, political, cultural and ethical challenge is the global migration crisis. As 'Doctors without Borders' claim, there are now 82 million forcibly displaced people[1], whereas the UN estimates that in 2020 there were around 281 million people globally considered as international migrants[2].

Reporting on the issue poses a plethora of challenges: defining a migrant and a refugee, gathering and reporting data on the topic, its complexity and multi-dimensionality. Modern data journalism offers a handful of solutions that will inform the general public about the migrant crisis, without the hassle of cutting through the vast data sets. Being trained in statistics, critical analysis methods, but also ways in which people digest information around and about data, they are the link between the vast, blurry, unstructured and messy datasets and the information that needs to be conveyed, to make the public aware of and care about the problem of migration, but also to allow for public policy changes. It's especially hard to achieve in times when misinformation is unrecognisable from facts, and selected facts are being treated as a single source of truth, when they are far away from it.

[1] https://www.doctorswithoutborders.org/refugees
[2] https://worldmigrationreport.iom.int/wmr-2020-interactive/

The aim of this case study is to approach the migration crisis topic from the data journalism perspective, and provide the public opinion (a.k.a. the MediaNumeric team and the school's cohort) with a robust, compelling story about migrants.

Visualisation of the Resettlement Process. Source: Resettlement Data Finder (2022)

Choosing Your Datasets

For this case study challenge you are encouraged to focus on one migration crisis topic, such as the Syrian refugee crisis, the Rohingya Refugee Crisis, or Migration from Europe in the 21st century. Try to think like a journalist and select a focus that is relevant, recent and interesting to you. You are free to use any data set that your group agrees on. You can scrape the data from the internet, download them from any source, or use links to data sets and resources provided by the MediaNumeric team (here).

Tasks

As a group, you are expected to deliver the following tasks during the winter school experience between February 21-26, 2022:

- **Create a key question** your group wishes to work on - it's up to you how you approach the case, which datasets you will use and how you want to craft and convey the message;
- **Gather data and include rationale** standing behind picking particular data sets;
- **Analyse** the chosen data sets and draw conclusions;
- **Visualise** the data (if relevant to your chosen story medium);
- **Create a story** (1-2 pages, 1800 - 3600 characters with spaces) for a medium of your choice;
- **Share** your work with the MediaNumeric team by Friday 25thFebruary by noon (prior to the AFP Study Visit) via Participants' Google Drive. Please check the Deliverables list for what we expect to be uploaded on the drive;
- **Deliver the final data journal story** to an audience (Saturday morning, 26th February): you can present it in front of the group or you can record a video, podcast etc.

Deliverables

Materials to Upload to the Drive

When sharing your final story on the Google Drive, please upload the following:

- A PDF version of your final data journalism story (or a relevant link to listen/watch your story with);
- A folder with the dataset(s) used for the story;
- A bibliography of sources used to create your story;

Pitch Rules & Requirements

The aim of the pitching session is to encourage you to learn how to sell your idea, in a limited amount of time. The final pitch should only focus on the final data journalism story, **not** the background information (i.e. the process you took to arrive at this final story). The latter will come up during the post-pitch discussion. You should expect questions about your data gathering process, decision making, group work dynamics, etc.

Rules & Format:

- Each group pitches in front of the audience (panel of experts & classmates);
- Each group has **7 minutes** for their presentation;
- Each presentation is followed by a **10-minute feedback session** with a Q&A from the panel of experts and a **5-minute open Q&A** from experts & classmates;
- Each group is free to use its own slide template if you wish, or any format of the presentation is acceptable;
- There´s no minimum / maximum number of slides that can be used;
- You are free to organise the structure of your pitch as you wish.

Things to Consider Including in Your Pitch:

- Catchy title/headline
- Key question/problem you're tackling
- (A) data point(s) with explanation
- A visualisation or other media format (optional)
- So What? Why is this important? Why should we care?

This is not a competition :-)

Additional Materials

Suggestions of Data Sources

In this section we have provided a few data sources that can be used for this case study. Please be aware that this list is not exhaustive and that you are encouraged to also find your own data sets online as well to support your selected case study.

UNHCR (United Nations High Commissioner for Refugees):

- About Asylum and Migration: https://www.unhcr.org/asylum-and-migration.html
- UNHCR Global Trends on Refugees: https://www.unhcr.org/data.html
- UNHCR Refugee Statistics: https://www.unhcr.org/refugee-statistics/
- UNHCR 2021 Refugee Trends: https://www.unhcr.org/mid-year-trends
- UNHCR Resettlement Data Finder (RDF): https://rsq.unhcr.org/

Eurostat:

- Home Page: https://ec.europa.eu/eurostat
- Overview on Migration and Asylum: https://ec.europa.eu/eurostat/web/migration-asylum
- Statistics on Migration to Europe: https://ec.europa.eu/info/strategy/priorities-2019-2024/promoting-our-european-way-life/statistics-migration-europe_en
- Migration Data Portal (Europe): https://www.migrationdataportal.org/regional-data-overview/europe

IOM (International Organization for Migration):

- IOM Data Analysis Center: https://gmdac.iom.int/

Lighthouse Reports:

- Home page: https://www.lighthousereports.nl/
- Listen to episode 40 of this podcast: https://datajournalism.com/listen

Harvard Dataverse:

- Home page: https://dataverse.harvard.edu/
- Syrian Refugee Life Study (UC Berkeley): https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/GWK4NT

The Migration Observatory (University of Oxford):

- Home page: https://migrationobservatory.ox.ac.uk/
- Data Guide: https://dataguide.migrationobservatory.ox.ac.uk/

Data Scraping Tips and Tools

In this section, we have provided a couple of tools that might help you to collect and analyse your datasets in a cleaner/deeper fashion:

- **Dataminer** (a web scraping tool that operates as a Google Chrome extension): https://dataminer.io/
- **Open Refine** (A tool for cleaning messy data): https://openrefine.org/

Data Visualisation Tips and Tools

In this section, we have provided a few data visualisation tools and examples that might help you to visualise your own data:

- **Reuters Graphics** (Migration visualisation): http://graphics.thomsonreuters.com/15/migrants/index.html
- **Migration Data Portal** (Example Visualisations): https://www.migrationdataportal.org/blog/10-coolest-visualizations-migration-data
- **Tableau Public** (Data visualisation tool): https://public.tableau.com/s/
    - Check out the **Tableau Gallery** as well for inspiration: https://public.tableau.com/app/discover/viz-of-the-day
- **Flourish** (Beautiful data storytelling platform): https://flourish.studio/
- **Infogram** (browser-based infographics generator): https://infogram.com/

4.3.2. Pre-Work and Pre-Read Document

The document linked here consists of all information important for the students prior to attending the Winter School, including required reading, technical prerequisites and basic information about the case study preparation work.

Pre-work

- Please ensure that you have the latest versions of Google Chrome or Firefox installed on your computers;
- Add the InVid Verification plugin to your browser. For Chrome, click here. For Firefox, download directly from the InVid page;
- Add the Wayback Machine plugin to your browser. For Chrome, click here. For Firefox, click here;
- Add the CrowdTangle plugin to Chrome by clicking here. This plugin does not exist on Firefox;
- If possible, watch Spotlight, the 2015 film by Tom McCarthy about the true story of how the Boston Globe newspaper uncovered the massive scandal of child sex abuse within the local Roman Catholic Archdiocese, shaking the entire church to its core.

Case Study Pre-Work

1. Please write a short answer to the following question: How would you go about finding data about migration? Where would you look for the data? Why?
2. Make use of your outline above and select 2-3 data sources, with a rationale for selecting them. Please have these data sources ready for Day 1 of our Winter School;
3. Find an example of a visualisation of the topic of migration, from any angle you think is interesting, and describe things you like, and don't like about it. Have it ready for our Day 1 meeting;
4. Listen to episode 40 of the Conversations with Data podcast.

Pre-Reading

1. Dario Compagno, 2016, Families of practices. A bottom-up approach to differentiate how French candidates made use of Twitter during the 2014 European Campaign, in: "Tweets from the Campaign Trail: Researching Candidates' Use of Twitter During the European Parliamentary Elections" (Alex Frame et al., ed), pp. 33-52.
2. Jacqueline Pietsch and Daniel Sorabji, 2021, State of the Art report in storytelling with data and misinformation.

4.3.3. Pitching Session's Support Deck

The deck below consists of the pitching session rules and was presented as a visual background for the session during which the students presented their work. The rules were:

- Each group will pitch on the stage in front of audience;
- Each group will have 7 minutes for their presentation;
- Each presentation will be followed by a 10 minute feedback session with Q&As from the panel of experts and a 5 minute open Q&A;
- Each group is free to use its own slides template - but you also do not have to - any format of the presentation is acceptable;
- There´s no minimum/maximum number of slides that can be used;
- You are free to organise the structure of your pitch as you wish.

Figure 14: Slide deck pitching session

4.3.4. Student Presentations

On the last day of the Winter School, the students presented their work to the other groups and to the MediaNumeric consortium representatives. There were four groups, assembled prior to the Winter School by the MediaNumeric team. The groups were created in a way to assure gender balance, but also a mix of students from different institutions and countries. Below are the descriptions of the stories created by the students. The presentations themselves are not shared to make sure the students' anonymity is preserved.

Group 1 shared their presentation entitled "A broken asylum system: Europe is failing asylum seekers who live in uncertainty that is impacting their life dramatically" that aimed to understand the situation of asylum seekers in Europe, as presented in the publicly available data sets.

Group 2 shared their presentation entitled "1 vs. 47,329. You've heard the story of the few, but not of the many" that provided a perspective on how journalists report on migration crises through a lense of individual cases (ie. Alan Kurdi's death), and how it's connected to the lack of possibility to give proper attention to each of the victims of the migration crisis.

Group 3 created a short movie that aimed to understand how Eric Zemmour's presidential campaign in France used data and how journalists can debunk some of the false or misleading data points used in the public discourse.

Group 4 created a presentation and a video that aimed to understand to what extent migrants in Europe contribute to the economy and enrich local communities. They showed a perspective of a situation in which migrants are missing - and potential effects of this void.

5. Study Visit

As part of the MediaNumeric programme, a study visit is planned to the offices of one of the consortium partners during each of the three training sessions. The aim is to demonstrate to the students how the skills learnt in the classroom can be applied in a real-life professional setting.

For the Paris school, the study visit was scheduled at Agence France-Presse (AFP). AFP is one of the largest three global news agencies in the world, with 2,400 members of staff of 100 different nationalities spread across more than 150 countries. AFP's core business is factual storytelling. Its journalists report on world events 24 hours a day in six languages, delivering the news in video, text, pictures, multimedia, and graphics. AFP also has the world's biggest fact-checking network with more than 100 journalists reporting on misinformation and disinformation in 24 languages.

5.1. Identifying the Goals of the Visit

In order to make the visit as relevant as possible to the course, the MediaNumeric project manager for AFP consulted with consortium members on the students' needs and expectations. It was decided to show the students the different stages of creating a story: from the morning editorial conference to the instructions given to the various journalism teams and disciplines, to reporting in the field, editing and publication. Once the needs were identified, the MediaNumeric project manager for AFP identified the relevant people within the agency who could best represent and relay these processes.

5.2. Preparing for the Study Visit

With social distancing restrictions still in place to limit the spread of Covid-19, special dispensation was sought from the agency's security team in order to permit a group of 25 people to enter AFP’s building.

Three key people were identified, approached and secured to lead the presentations for the students: the global deputy news editor, the deputy head of the data and graphics, and the head of the digital verification team.

5.2.1. Unexpected Challenges

In the weeks leading up to the Winter School in Paris, tensions grew between Ukraine and Russia culminating in the invasion of Ukraine on the day of the study visit. With a major news event unfolding, people who had previously been scheduled to present to the students were now solely focussed on organising the news coverage and the security of the teams on the ground. The study visit had to be reorganised at the last minute.

The director for Europe, Christine Buhagiar, stepped in to replace the deputy global news editor, and the deputy head of data and graphics, Simon Malfatto, agreed to present remotely while the

head of the Europe digital investigation team, Bronwen Roberts, stepped in to replace the head of the digital verification service.

Taking the news of Russia's invasion of Ukraine as an example, each of the representatives explained to the students how AFP approaches an event such as this. These included:

- How reporting teams are put in place;
- How specialists from different disciplines (text, photo, video, graphics, digital investigators) work together both on the ground and in bureaus around the world to ensure quality of reporting on a fast-moving news story, including an exchange of information and facts;
- Gathering reliable data on maps, troop movements, military hardware at speed, and organising the information into a readable and explanatory graphics;
- Identifying and debunking misinformation and propaganda that surrounds any conflict.

One of the goals of the study visit was to create a connection between the lectures and workshops that the students had worked on during classes, as well as to show how applied knowledge and experience look in practice. With a focus on interdisciplinarity of the team in the newsroom, as well as emphasising the importance of teamwork, the study visit demonstrated how complicated, but also responsible a journalists' work is, and that fact-checking, debunking and data storytelling is a joint effort.

6. Quizzes and knowledge verification

In order for students to pass the MediaNumeric training course they were required to sit for quizzes that covered all lectures taught spanning the three main modules of the course. The students were required to take the quizzes in two parts. The first part consisted of quizzes covering Modules I and II, the second quiz referred to the material covered in Module III. The students were given 90 minutes in total. Below are all three quizzes listed with students' answers and a complete list of questions. Wording of the quiz' questions align with the quizzes that the students took.

⇒ Module 1 quiz
⇒ Module 2 quiz
⇒ Module 3 quiz

Quizzes are part of a broader evaluation process carried out mainly by INA, with support from other partners. The 'Evaluation Methodology' report refers to three stages of the evaluation:

1. "**Assessment of the level of knowledge and skills of the participants before the training course**: verification of the understanding of the contents covered throughout the course;
2. **Assessment of the knowledge and skills acquired by the participants at the end of the training course**: alignment of the results of the training programme with the pedagogical goals;
3. **Assessment of participants' satisfaction with the training course**: course content, pedagogical approach, formats and methodology, speakers, etc."

The quizzes are part of the second point in the above list which aims at assessing the students' knowledge and skills acquired during the Winter School and were designed to measure participants' progress in all three modules: "the examination of learning outcomes of the onsite training session consists of end-of-training validation tests (...). The MediaNumeric programme awards academic credits (5 ECTS for on-site training programme (...), detailed in a transcript of records and skills. (...) The examination of learning outcomes for each on-site training programme will be conducted through end-of-training tests designed in the form of quizzes. The quizzes will be based on the teaching modules (...), the needs analysis (...), and the EQF (European Qualifications Framework)."

Based on these assumptions, the following quizzes were created and used as an evaluation tool during the Winter School.

6.1. Module 1 Quiz: Search and Exploration of Data

⇒ Correct answers are in bold

Strategies of Data Search

1. How would you define data-driven journalism?

"A type of journalism practice aimed at analysing datasets to tell a story and give context or useful / insightful explanations to better understand the news."

2. What is Open Data?
   a) The idea that Science data should be open access
   b) Spreadsheet software
   c) A Data scientist tournament
   d) **A practice of openly publishing government's data**

3. What is CSV?
   a) A mathematical software
   b) An American data journalism outlet
   c) **A data file format**
   d) A forum specialised in Data Queries

4. If you were to visualise the voting intentions from one poll, would you use...?
   a) A Line Chart
   b) **A Column Chart**
   c) A Pie Chart
   d) A Scatter Plot

5. What is the purpose of a scatter plot?
   a) Compare data points between them
   b) **Represent the distribution of a series**
   c) Show relationships between items
   d) Visualise the evolution in time of a dataset

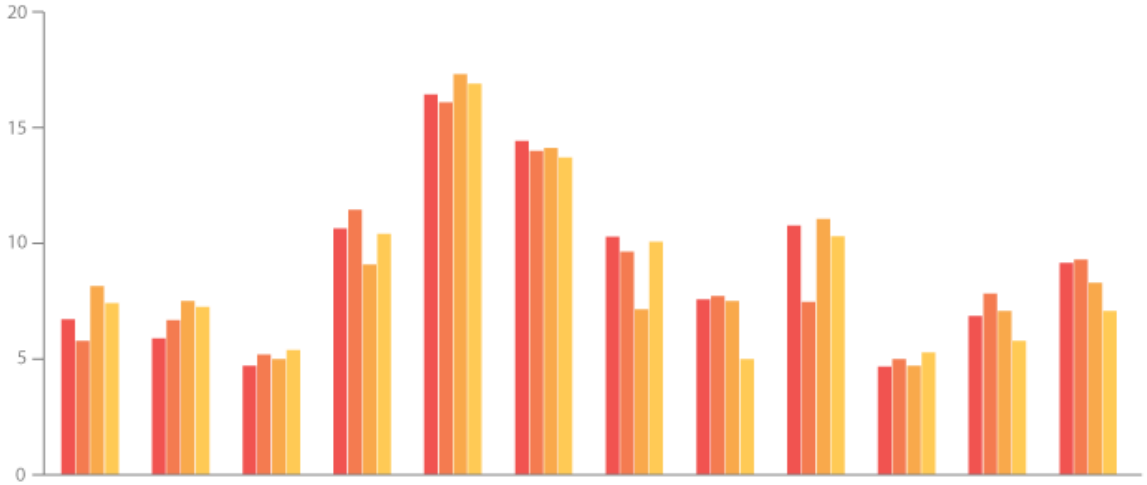
6. In order to fairly compare territories, you would need...?
   a) A Bubble Map
   b) A Treemap
   c) **A Choropleth Map**
   d) A Flow Map

7. What is a Pivot Table?
   a) **A tool to easily reshape your spreadsheet's data**
   b) A feature that cleans your spreadsheet's data
   c) A filter that allows you to select the data you are looking for
   d) A cell's function

8. How would you find data on Spanish Baccalaureate graduates?

"By finding the Spanish Government's Open Data Portal (ideally by using the "Open Data Inception" website) and then searching the dataset on it"

9. Is this chart...?



a) An Area Chart
b) A Stacked Column Chart
c) **A Grouped Column Chart**
d) A Bar Chart

10. What is the name of the American Law that ensures access to public records?
    a) Freedom of Transparency Act (FOTA)
    b) Liberty of Information Act (LOIA)
    c) Freedom of Data Act (FODA)
    d) **Freedom of Information Act (FOIA)**

11. What Search Engine operator would you use to search for a specific file extension?
    a) Extension
    b) Fileformat
    c) Format
    d) **Filetype**

12. If you were to count how many white federal judges were appointed by Barack Obama, what would you use?

Last Name	First Name	Commission Date	Birth State	Gender	Race or Ethnicity	U.S. District Court for . . .	Appointing President
Birotte	Andre	2014	NJ	Male	African American	the Central District of California	Barack Obama
Brodie	Margo	2012	Antigua	Female	African American	the Eastern District of New York	William J. Clinton
Childs	Julianna	2010	MI	Female	African American	the District of South Carolina	Barack Obama
Blakey	John		IN	Male	White	the Northern District of Illinois	Barack Obama
Brann	Matthew	2012	NY	Male	White	the Middle District of Pennsylvania	Barack Obama
Bruce	Colin	2013	IL	Male	White	the Central District of Illinois	Barack Obama
Feinerman	Gary	2010	IL	Male	White	the Northern District of Illinois	Barack Obama
Mazzant	Amos	2014	PA	Male	White	the Eastern District of Texas	Barack Obama
Jordan	Sean	2019	NY	Male	White	the Eastern District of Texas	Donald J. Trump
Dow	Robert	2007	WI	Male	White	the Northern District of Illinois	George W. Bush
Hardiman	Thomas	2003	MA	Male		the Western District of Pennsylvania	George W. Bush
Laplante	Joseph	2007	NH	Male	White	the District of New Hampshire	George W. Bush
Crane	Randy	2002	TX	Male	Hispanic	the Southern District of Texas	George W. Bush
Gelpi	Gustavo	2006	Puerto Rico	Male	Hispanic	the District of Puerto Rico	George W. Bush
Howard	Marcia	2007	FL	Female	Hispanic	the Middle District of Florida	George W. Bush
Nguyen	Jacqueline	2009	Vietnam	Female	Asian American	the Central District of California	Barack Obama
Bennett	Alfred	2015	TX	Male	African American	the Southern District of Texas	Barack Obama
Bolden	Victor	2014	NY	Male	African American	the District of Connecticut	Barack Obama
Hanks	George	2015	LA	Male	African American	the Southern District of Texas	Barack Obama
Logan	Steven	2014	WA	Male	African American	the District of Arizona	Barack Obama
Humetewa	Diane	2014	AZ	Female	American Indian	the District of Arizona	Barack Obama

a) A Tilter
b) **A Pivot Table**
c) A Geocoding tool
d) The Median function

Data Literacy

13. How would you organise collaboration with your colleagues while working on a project? Check all data-driven options.
    a) **I would make a map on my desktop to store all documents related to the project there.**
    b) **I will make a WhatsApp group for communication with colleagues.**
    c) **I will arrange the project space in one of the online applications such as Microsoft Teams, Miro, Trello etc.**
    d) I would just follow my intuition and make notes in my notebook.

14. What steps does systematical data collection include? Check two most adequate answers.
    a) Look what is available via Google
    b) **Explore what data is available in specialised databases**
    c) Use everything that has to do with my research question(s)
    d) **Critically assess sources of data for trustworthiness**
    e) Other

15. What artificial intelligence is not? Check at least 3 correct answers.
    a) **Machines learn by themselves**
    b) **Machines are objective**
    c) Ability of computer to iteratively improve themselves based on the information they collect
    d) **AI is the same thing as machine learning**
    e) Ability of a computer or robot to perform tasks commonly associated with intelligent beings
    f) **AI is smarter than people**

16. How could big data technologies be applied in social media use by companies? Check at least 3 possible answers.
    a) **Personalisation of content**
    b) **Identify trends**
    c) Solving problems that are not well-defined
    d) **Measure the effectiveness of campaign**
    e) Recommend content on demand
    f) **Work on creative tasks**

17. What applications would facilitate your project collaboration and management? Choose 3 correct answers
    a) **Miro**
    b) **Padlet**
    c) TikTok
    d) **Microsoft Teams**
    e) Zoom
    f) LinkedIn

18. What skills are part of data literacy? Choose all correct answers
    a) **Ability to evaluate data**
    b) Ability to read
    c) **Ability to cooperate in a data project**
    d) Ability to count
    e) **Ability to visualise data**
    f) Ability to write

19. What is sentiment analysis?
    a) “The process of computationally identifying of a persons’ feeling”
    b) **“The process of computationally identifying and categorising opinions expressed in a piece of text”**
    c) “The process of computationally analysis of emoticons”

Tools & Databases: Basic Principles + Choice of Apps

20. A web crawler is:
    a) **a piece of software going from hyperlink to hyperlink on the web while collecting data**
    b) automatically extracting structured data from a webpage
    c) a new species of insect that can only lives on servers connected to the web
    d) somebody who gets all your personal data to sell it to personalised advertisers

21. Web crawling can be used to:
    a) Extract a specific column from a table.
    b) Understand how a web page has been coded.
    c) Hack and destroy a website.
    d) **Understand which websites are quoting other websites, or build a web of knowledge on a specific topic.**

22. When we start to web crawl (i.e. to follow the links found on a specific web page), we often fall back into well-known websites (such as Facebook, Wikipedia, Twitter, …) called the "high layer" of the web. Why do we often fall back on those websites?
    a) **Because most web pages have a link that points to this high layer.**
    b) Because Facebook, Wikipedia and Twitter have a high number of users.
    c) Because Facebook, Wikipedia and Twitter mostly have European or American users.
    d) Because all websites are forced to have at least one link that points to this high layer.

23. The sealsology tool shown in class is a web crawling tool that allows you to:
    a) Give you a list of places to visit when you enter the link of a web page.
    b) See the touristic landscapes associated with a specific place around the world.
    c) **Extract the links from the "see also" section of a Wikipedia page to build the web of knowledge around one or many Wikipedia pages.**
    d) Build a graph of knowledge of the topics of discussion of a specific Facebook group.

24. Scraping a web page can be defined as :
    a) Copying and pasting some content of a web page.
    b) **Automatically extracting data from the HTML code of a web page.**
    c) Collecting the links to other web pages to build a network of pages quoting each other.
    d) Logging in a specific database.

25. The HTML code of a web page (that allows web scraping) can be accessed by:
    a) Double-clicking on the web page.
    b) **Right-clicking on the web page and selecting "View page source".**
    c) Logging into a web page using a special admin access.
    d) The HTML code of a web page can never be accessed.

26. What is one advantage of web scraping?
    a) It is very easy to learn, even without any notion of coding.
    b) The web scraping tools are often generic and can easily adapt to any website you wish to extract data from.
    c) **It is easier to extract data using web scraping than to copy-paste the content of many web pages.**
    d) It often takes a few seconds to develop a web scraping code.

27. What is a limit of web scraping?
    a) Only very experienced developers (20+ years) are able to do it.
    b) **Some websites prevent scraping by blocking visitors with unusual behaviour (such as visiting too many web pages in a limited period of time).**
    c) You need a specific scraping licence only given by the government to a handful of developers.
    d) It is forbidden by the GDPR (General Data Protection Regulation).

28. Web scraping creates many legal problems. What is the one problem that do not arise?
    a) you may scrape and store personal or sensitive data, which is an issue with the GDPR regulation
    b) you may scrape article or book content, which creates a problem with copyright if you reuse this content
    c) you may scrape websites that have explicitly stated in their terms of service that scraping is forbidden
    d) **while scraping a website, you may put classified documents online by accident and be accused of spying.**

29. What is a web API?
    a) **an interface between somebody asking for data and a database**
    b) a Google spreadsheet
    c) a piece of software going from hyperlink to hyperlink
    d) a farmer raising bees on the web

30. About data that we can found on the web:
    a) The only way to collect data from the web is through APIs.
    b) **There are many formats for web data, and some websites can offer to download data using an API.**
    c) The only way to collect data from the web is by using code.
    d) You cannot download data from the web.

6.2. Module 2 Quiz: Telling Stories with Data

⇒ Correct answers are in bold

Panorama of multimedia storytelling and the contributions of data. State of the art of storytelling multimedia. Why tell a story with data?

1. What phenomenon is called « infobesity »?
   a) a modern tendency of over-producing data
   b) a modern tendency of over-consuming data
   c) an ecological issue of storing all the data we create
   d) a personal cognitive incapacity of dealing with all the data we create

2. What main reason makes that a bare graph is usually not enough to engage an audience?
   a) it's too complicate
   b) it over-simplifies reality
   c) **it's not engaging enough**
   d) it's too long to understand

3. What main reason makes that a bare graph is usually not enough to create empathy?
   a) we have too much biases
   b) **we have issues understanding big amount of numbers, we need exemples**
   c) we're too much solicited in modern society
   d) we have trouble understanding it and its issue

4. What phenomenon is called « statistical numbing »?
   a) **the difficulty we have to process statistics and their implications in real life**
   b) the difficulty we have to remember statistics
   c) the difficulty we have to visualize statistics
   d) the difficulty we have to create statistics on certain events

5. In scrolly-telling, what do we call the trigger?
   a) the action of opening the internet explorer
   b) **the action of scrolling**
   c) the action of creating emotions through the story
   d) the action of sharing the story on social media

6. What is the most basic way to create a beginning of storytelling within a chart?
   a) adding a caption
   b) adding an illustration
   c) adding an explanatory title
   d) **adding an exploratory title**

7. What is an “outlier” in data?
   a) an external variable
   b) **a value extremely different from the rest of the dataset**
   c) an unexpected trend
   d) an added variable that changes the dataset

8. What is the main reason why design is a powerful tool in storytelling with data?
   a) **it convey emotions in an immediate way**
   b) it’s understandable by everyone
   c) it take less time to read
   d) it’s better memorise by people

9. What does Jake Harris means by we must « connect to the dots » in datavisualisation?
   a) we must understand how data are connected to each other
   b) **we must see beyond data to connect with the real people behind it**
   c) we must connect to the audience through storytelling
   d) we must rely on data to tell stories

10. What steps or tools can you use to introduce more storytelling with data? (open-ended question)
Having exploratory title, choosing the right chart for the right mental representation, using an incarnation of your data to start from, using scrollytelling, using interactivity, etc

Techniques and Tools Adapted to Multimedia Storytelling. The Different Techniques for Designing a Data Storytelling Project

11. Which one of these positions is usually NOT involved in storytelling?
   a) data analyst
   b) illustrator
   c) **data engineer**
   d) journalist

12. Which one of these tools is usually NOT involved in storytelling?
   a) D3
   b) Datawrapper
   c) Khartis
   d) **PowerPoint**

13. Which one of these softwares is NOT used to create data visualisation from a dataset?
   a) Rawgraphs
   b) Tableau
   c) **InDesign**
   d) Illustrator

14. In which format export a data visualisation created on an online platform to be able to customise it fully on a vector-based software?
    a) jpg
    b) **svg**
    c) png
    d) csv

15. What is the main downside of using a free online platform?
    a) you can't collaborate with other person
    b) **your data becomes publicly available**
    c) you can't save your work
    d) you can only use a limited amount of data

16. What do we call « sensitive data »?
    a) data that concerns a controversial subject
    b) data about under age persons
    c) data collected without consent
    d) **data that you can't share publicly**

17. In order to choose the good visualisation, you need to focus first on :
    a) **the type and relationship in your data**
    b) the size of your dataset
    c) the source of your data
    d) the liability of your data

18. In Flourish, which one of this format is NOT read by the platform to upload your data ?
    a) csv
    b) json
    c) geojson
    d) **png**

19. In Flourish, what is the basic step to visualise data?
    a) **defining the columns in the table to visualise**
    b) defining the type of data to visualise
    c) defining the ID of the data to visualise
    d) defining the numbers of data to visualise

20. In Flourish, can you upload more than one dataset in a project?
    a) No you can't, you have to create a new project
    b) No you can't, you have to erase the previous one before
    c) **Yes you can, but the second dataset must share at least one column with the previous one**
    d) Yes you can, but the second dataset must contain the same number of rows that the previous one

The Main Stages of a Data Project, from Data Exploration to Storytelling

21. Data collection deeply affects the results of a data project. A good data collection would better be described as a matter of external validity or of internal validity?
    a) **External validity**
    b) Internal validity

22. A data analysis project is made of different stages. Which is the last one? (open-ended question)

Communicate

23. It is well known that some stages of every data analysis project take much more time than the others. Which ones are they? (open-ended question)

Import data and tidy them up

24. Data science is often thought of as a combination of three skills. Which are these three skills? (open-ended question)

Statistical knowledge, coding skills, substantive expertise in the concerned domain

25. Which combination of these three skills has been called the “danger zone”?
    a) **Coding skills together with substantive expertise, but without statistical knowledge**
    b) Coding skills together with statistical knowledge, but without substantive experience
    c) Statistical knowledge together with substantive experience, but without coding skills

26. Programming languages for data analysis are the most powerful software available. Because of this reason, they are expensive and can only be afforded by teams with considerable economic resources.
    a) True
    b) **False**

27. Which are the two most commonly used programming languages for data analysis?
    a) Java & R
    b) Python & Ruby
    c) Ruby & Java
    d) **R & Python**

28. Which is one of the main disadvantages of using point-and-click software for data analysis projects?
    a) It takes a lot of time and motivation to start learning it
    b) The graphics produced out-of-the-box are often frustrating
    c) **Each software is specialized in a limited scope of functions**
    d) You need to figure out every little detail of the functions you are using

29. For which reason journalism has a hard time reporting about science?

Because journalists produce well-formed stories with a clear beginning and a clear end, in which the good and the bad can be easily inferred by readers. Science instead describes phenomena in tentative terms, pointing to the limitations of its claims, and refraining from offering value judgements as to what would be better and should be done.)

30. Expressions like “There is no evidence supporting this claim” are ambiguous and can be understood in at least two contrasting ways. Please describe in a few words such ambiguity and produce two examples, one for each contrasting interpretation.

“No evidence” may refer to the temporary lack of confirmation for a reasonable hypothesis, for example about some side-effects of a drug, or to the status of a useless fantasy, such as a conspiracy theory.)

6.3. Module 3 Quiz: Tracking & Debunking Misinformation

⇒ Correct answers are in bold

Media Ethics

1. When you discover a mistake in the published material
   a) **if you can correct it (eg on the website), do so - and make a note of the changes**
   b) if you can correct it, do so – but do it discreetly
   c) even if you can correct it, you shouldn't – what's published is published!
   d) create a new material, with information correcting the mistakes of the previous one

2. When working with data visualisation, take into account:
   a) A technical correctness
   b) Avoiding of potentially misleading composition techniques (eg axis manipulations, aspect ratio, data selection)
   c) **Both of the above**
   d) None of the above

3. "Potter Box" method helps making ethical decisions by:
   a) funny references to the works of J.K. Rowling
   b) consideration of pros and cons
   c) **analysis using four dimensions of dilemma – facts, values, principles and loyalties**
   d) analysis using four dimensions of dilemma – legal, economic, social and cultural

Tracking & Debunking Misinformation

4. What is the Dunning–Kruger effect?
   a) **When people with low knowledge of a topic overestimate their understanding of it**
   b) When by mentioning a topic someone trying to deflect attention ends up attracting more
   c) When someone only believes what confirms their opinion and discredits what doesn't
   d) When someone is overly influenced by the first information they received on a topic

5. I am interested in a viral post I see on social media but I am not sure if the photo in it is real. I…
   a) Share it and hope someone will tell me
   b) Ask for a source in the comments
   c) Run a reverse search to see if the photo was posted by other people
   d) **Run a reverse search to find the author and context of the photo**

6. What is the difference between misinformation, disinformation, and fake news?
   a) **Misinformation is false or misleading information, disinformation is misinformation shared on purpose and fake news are fabricated information given the appearance of news**
   b) Misinformation is unintentional, disinformation is intentional and they are both fake news
   c) Misinformation is slightly false information, disinformation is completely false, and fake news are fabricated information
   d) They are all the same

7. Which of these search engines would be the most appropriate to search a picture allegedly filmed in Russia?
   a) **Yandex**
   b) Google
   c) Baidu
   d) Tineye

8. Where can you get results in chronological order for an image reverse search?
   a) Yandex
   b) Google
   c) Baidu
   d) **Tineye**

9. What is the term for “too many true, false and misleading information circulating during a disease outbreak”?
   a) **Infodemic**
   b) Epidinfo
   c) Pandinfo
   d) Infostorm

10. What is the “Continued-influence effect”?
   a) **When someone keeps believing things even after they’ve been corrected**
   b) When someone cannot stop listening to the same source
   c) When someone cannot get out of their own information bubble
   d) When someone believes something because a lot of people believe it

11. What is Google Lens?
   a) A text recognition and translation tool
   b) A tool to download videos
   c) **A mobile and browser application to perform Image reverse search**
   d) An archiving tool

12. What are the questions you need to answer when researching suspicious images?
   a) What sort of content is it, who posted it on social media, when, where and why was it posted
   b) **What is the content and is it original, who is the author of the content, when was it taken/made, where was it taken and why was the content made and published**
   c) What sort of content is it, when was it posted, why is it important, who is the author and where was it taken
   d) What can you see on the images, where and when were they taken, who posted them on social media and why

13. What is Baidu useful for?
   a) Identifying and translating text from a video in Chinese
   b) Downloading a video from Weibo
   c) Archiving links from Chinese governmental websites
   d) **Running a reverse image search on images seemingly coming from China**

14. Which of these is NOT a reason we believe in false information?
    a) **Low level of education**
    b) It confirms our opinion
    c) It explains a complicated or random event
    d) It comes from an authoritative source

15. How would you search on the CDC website for information on vaccines side-effects but without mention of Covid-19 vaccines?
    a) site: cdc.gov vaccines side-effect -Covid
    b) inurl:vaccines AND side-effect -Covid
    c) cdc AND vaccines AND side-effects -Covid
    d) **site:cdc.gov side-effects AND vaccines -Covid**

16. Say you want to search for PDF files about JFK marked confidential on the FBI's website. Which of the following search queries will give you the best results?
    a) file:pdf site:fbi.gov confidential JFK
    b) filetype:pdf site:https://www.fbi.gov/ confidential JFK
    c) filetype:pdf site:fbi.gov JFK John Fitzgerald Kennedy
    d) **filetype:pdf site:fbi.gov JFK OR "John Fitzgerald Kennedy" OR "John F. Kennedy"**

17. You want to search for PDF files about "climate change" on any website run by the Australian government, published between 2015 and 2021. Which of the following search queries will give you the best results?
    a) site:australia.gov.au filetype:pdf "climate change" (and then go to "Tools --> Any Time --> Custom range --> from: 2/10/2015)
    b) inurl:gov.au typefile:pdf "Climate change" 2015..2021
    c) **inurl:gov.au filetype:pdf "climate change" (and then go to "Tools --> Any Time --> Custom range --> from: 1/1/2015 to 31/12/2021)**
    d) site:gov.au typefile:pdf climate change (and then go to "Tools --> Any Time --> Custom range --> from: 1/1/2015 to 31/12/2021)

18. What are metadata?
    a) **Information about the making of a an image or video such as when, where and on which device it was created**
    b) Proof that an image was manipulated
    c) Edit history of a video
    d) Information about the owner of a website

19. Where can you get metadata from?
    a) A Facebook video
    b) A Whatsapp video
    c) A Twitter video
    d) **The original video sent by email or transfer**

20. Which of these is NOT a reason to archive
    a) To prevent link rot
    b) **To obtain metadata on a photo or video**
    c) To keep reliable traces of posts on social media
    d) To keep track of changes made on websites or articles

21. What is the first step when investigating a video?
    a) Contacting the user who posted it
    b) Doing an image reverse search on screenshots with Invid
    c) Watching it carefully with sound for visual and audio clues
    d) **Downloading it in case it is later deleted and archiving the link**

22. What is the best place to check for street level photographs when regular Streetview is not available?
    a) Satellites.pro
    b) Deepware Scanner
    c) InVID
    d) **Mapillary**

23. What is one social network with the least amount of content moderation?
    a) Instagram
    b) Twitter
    c) **Telegram**
    d) Facebook

24. Which of these is a recognized global international network of fact-checking organisations?
    a) **IFCN**
    b) Bellingcat
    c) Reporters sans frontières
    d) Ruptly

Social Impact of Journalism and Media Transitions

25. What has the internet changed for journalism? Find a statement that is not true.
    a) New intermediaries appeared
    b) New business models appeared
    c) The definition of journalism has blurred
    d) **Information flow regulations have been strengthened**

26. The new logic of internet companies is changing the role of the media - alongside communication tools, making them
    a) **means of data extraction**
    b) means of self-expression
    c) marketing tools
    d) manipulation tools

27. Calculated publics, the term coined by Tarleton Gillespie, refers to dangerous process, in which
    a) **self-perception of the public is based on its algorithmic presentation**
    b) the public only absorbs information that contains figures
    c) the public sphere produces data
    d) people only talk to avatars

28. Algorithms are objective, because…
    a) they are based on mathematical calculations
    b) they are technical creations
    c) they are intelligent
    d) **It is a myth - they are created by people who set specific goals for them, so they are not objective at all!**

29. According to the extractivist paradigm, data mining is like mining coal, because
    a) it is only profitable if you disregard the social costs
    b) it is the basic tool of progress today
    c) building and maintaining its infrastructure is destroying the planet
    d) **a and b**

30. What has the Cambridge Analytica scandal revealed? Please point WRONG answer:
    a) that mass personalization of the manipulated message is possible
    b) that the new media system gives politicians new tools for manipulation
    c) that manipulation practices based on illegally obtained data also took place outside the US
    d) **that cyber corporations respond to such violations on their own initiative**

7. Conclusion

7.1. Video Recording Plan and Initial Guidelines for the Publication of the Training Materials

The initial plan for the programme was to record each session and use recordings as materials for the online part of the course. Throughout the week we used equipment provided by Storytek - a high-quality recording device with both video and sound recording - as well as Zoom recording and streaming that was used by the consortium partners who were not present in the classroom, but wanted to keep track of what was going on in the room. The initial plan was to use these recordings after analysis to create the online course but we realised that the format and quality of these recording were not appropriate for an online learning session (teachers talking to students and not to camera, difficulty to follow the exercises, audio, Q&A from the room not appropriate to online setting, etc.). Lessons were learned from this experience to review lesson recording in the future and review the online course format.

Some other challenges that we spotted were as follows:

- The teachers paid more attention to the students, rather than to the recording device, hence quality of some materials is poor, with teachers being within or outside of the camera's reach;
- Questions asked by the students were inaudible;
- Some technical issues (ie. lack of internet connectivity or not having the microphone properly attached to the teacher's garment);
- Necessity for someone to be in the room at all times to make sure all teachers were aware of the recording, as well as all necessary steps required to record the session were taken.

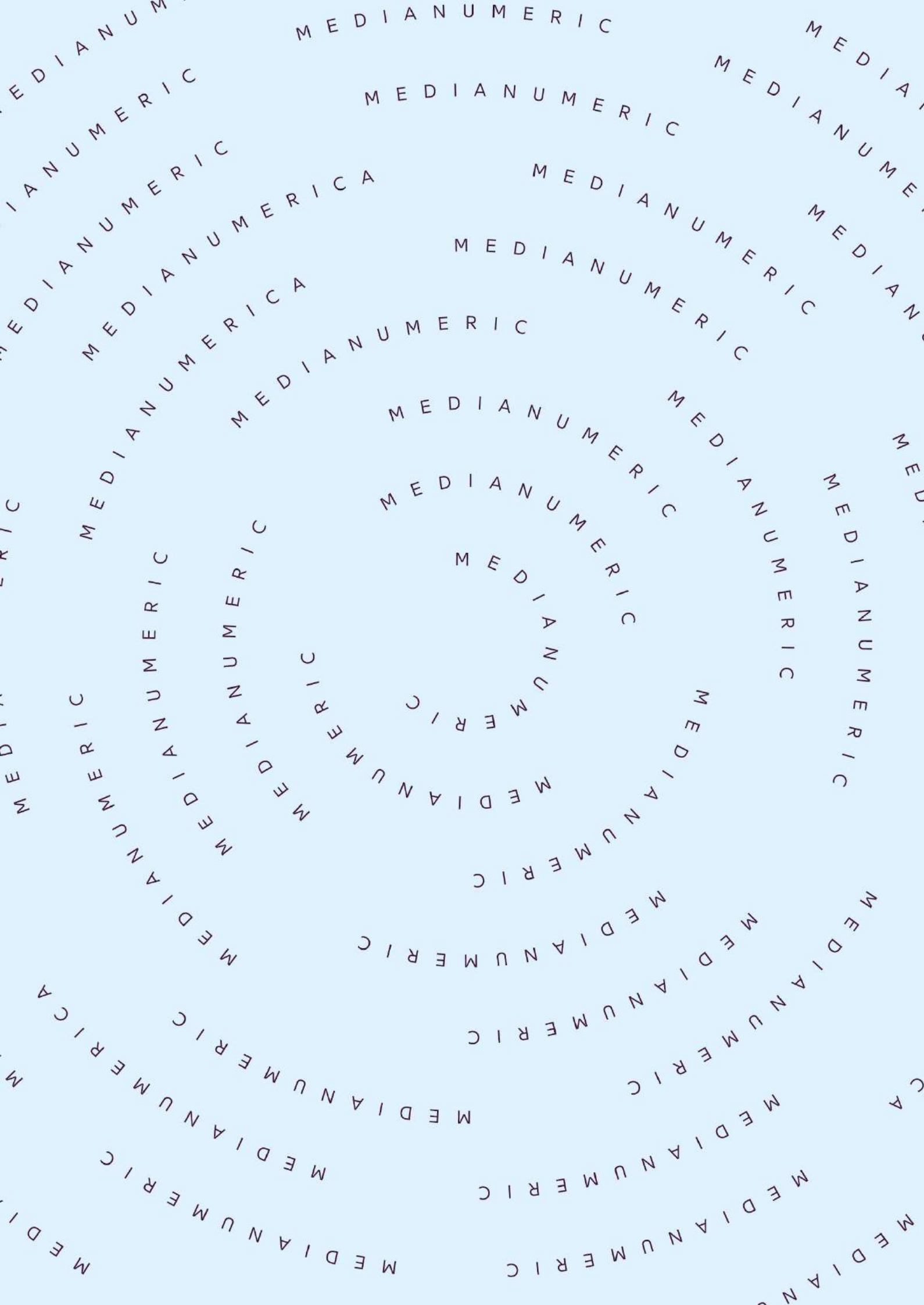
Based on these observations, as well as a few discussions we had, we came to a conclusion that:

- Recording of the lectures should happen outside of the classroom - we should ask teachers to create a 1:1 version of their course identical or very similar to the one presented in the classroom, based on the materials used there, either during next iterations of the school or in between the second and third training weeks (June/July 2022 and February 2023);
- We need to supplement the material with online-specific content that will adhere to a different way of presenting teaching materials in a self-paced environment.

7.2. Conclusion

Based on several discussions during and after the training programme in Paris, the group created a *post mortem* document, outlining thoughts and remarks based on the experience gained during the Winter School. Aside from the fact that - for teachers, students and the MediaNumeric partners participating in the event - the entire experience was a high velocity and intense week of learning, we came to the following conclusions:

- While preparing and delivering the programme, we should keep the end goal of the online training course in mind, meaning that eventually some materials will not be fit to use in the online course; that means proper preparation of additional materials, briefing the teachers and potential additional costs;
- More emphasis should be put on workshops and group work;
- Throughout the course there should be more smaller hands-on tasks and more interaction with less lectures;
- Students should have more time to work on the case study;
- The students should work on Google Presentations or Microsoft PowerPoint when preparing their pitches. One of the groups in Paris used a free version of Prezi that does not allow exporting the content - hence their deliverable is missing in the report;
- There should be more time to implement and practise the tools shown during the classes;
- Proper technical preparation of the classes is crucial - a technical person should be available at all times to support students' work;
- We should make sure the teachers are aware of the time they're given and that students will probably be asking quite a few questions (so they need to think about their lectures to include time for that);
- We should work on a better structure of providing all lecture materials and handouts to the students via an accessible Google Drive before the classes start;
- We should make sure the teachers know they will have time to show only one tool, the rest should be mentioned in the handouts (links to additional tools, links to the tutorials, etc.);
- We should make sure that the students know what's expected of them regarding the case study - not only mention that during Day 0, but also repeat that during the first day of the school and throughout the week;
- Lectures on Media Ethics and Social Impact of Journalism should take the form of a keynote lecture kicking off the course;
- The students should have some time off, ie. on Wednesday: finish around 2-3pm and give them time to do whatever they'd like (ie. network, work);
- The programme should allocate lots of time for practical exercises in the afternoon;
- The programme should allocate time and space for revision of the materials prior to the quizzes;
- The quizzes covering past classes should take place at least one day after the last lecture that the test will cover to give the students time to revise the materials;
- We should have each case study session start with a small task/exercise;
- We should make sure that all materials are kept consistent and that the teachers use templates provided by the MediaNumeric team;
- We should make sure that all students have access to, and have explored, the Google Drive before classes start;
- Direct live communication with students (Basecamp, Slack or WhatsSpp group) - to be considered.


MEDIANUMERIC