RESEARCH PROTOCOL DIAGRAMS AS DIDACTIC TOOLS TO ACT CRITICALLY IN DATASET DESIGN PROCESSES

Michele Mauri, Ángeles Briones, Beatrice Gobbo, Gabriele Colombo

Politecnico di Milano (ITALY)

Abstract

In this paper we present the design of research protocol diagrams as a didactic tool in communication design education. The goal is to teach critical aspects of dataset design processes to design students. The context is a five-month course of Data Visualization addressed to master students in Communication Design. For the last five years, the course has been structured in three phases that gradually introduce students to the demanding issues deriving from communication with data in complex situations. In the second phase, students learn how to map and present - through an interactive report -, a controversial issue based on data coming from online sources. The process is question-driven: each group starts with a set of research questions, defines a protocol for data collection and analysis, and produces research findings using data visualizations. In this phase, students are asked to critically reflect on data collection as a design activity and to visually communicate the dataset design process creating protocol diagrams.

Dataset design is a critical aspect of the entire research process, raising technical and ethical considerations. From our point of view, designing a dataset means dealing with data as an artifact, as the result of a series of steps leading to the construction of the dataset that will be encoded through data visualization. Even if the protocol diagram represents all the steps of the research, from the question to the final visualization, the most important part is the visual representation of the dataset design process.

Visual representation of processes by diagrams is an established practice dating back more than one hundred years. While research protocol diagrams have been widely used in other scientific disciplines to represent processes, they can be repurposed for the field of Communication Design for educational purposes.

In this context, protocol diagrams have two functions at different times: before the final delivery, when the diagram is continually updated as the research goes on, is a *negotiation tool* which helps students in previewing the possible path they should follow to complete the analysis, in disentangling and correlating individual actions, in examining, doubting and legitimizing the arbitrary choices they are led to make in the process of dataset design and finally is a tool for discussing within the group work and with teachers during the weekly reviews.

Once students complete their research, they submit a final report that includes the protocols, visualizations, and main findings. Hence, the protocol diagram becomes a *dissemination tool* addressed both to researchers or future students that want to replicate the same process. Students are free to choose the shape of the diagram they prefer, and they are not provided with international standards to follow. Usually, results are hybrids forms of process charts and activity diagrams attributable to the visual archetype of the flow-chart. The protocol diagram teaches students about the non-objective, situated and interpretative nature of any data-driven study and pushes them to exploit their communication skills to transform the research protocol, which is traditionally textual, into something exclusively visual.

Keywords: protocols diagrams, dataset design, data visualization, communication design.

1 CONTEXT: THE DENSITYDESIGN STUDIO-COURSE

In this paper we present the design of research protocol diagrams as a didactic tool in communication design education. The goal is to teach critical aspects of dataset design processes to students. The didactical context is the DensityDesign course: a studio course of the master's degree in Communication Design held at Politecnico di Milano [1]. Since 2004, the DensityDesign course focuses on the visual communication of complex phenomena. By working with data and its visualization, students conduct research on different issues and present their results with visual communication artifacts, reflecting on the role of information design in the study and communication of socially relevant issues.

The course brings together about 50 students of the last year of the master's degree in communication design. For five months, students work in groups studying topics proposed by the faculty, which address complex issues that impact society today (e.g. in the 2019 edition students covered topics such as facial recognition, deepfake technology, scams and fake profiles in dating apps, and face modification apps). Topics are broad enough for each group to find a more personal angle. The course combines practical workshop lessons in which students present their work and receive feedback from the faculty, and traditional classroom lectures about the design of data communication devices. Covered topics range from data visualisation to data collection and analysis, to strategies for public engagement through data-driven artifacts. Furthermore, lessons from statistics and semiotics are integrated. Throughout the course, working with data is presented to the students as an interdisciplinary endeavour that needs to be tackled with different and complementary skills.

The course is structured in three main phases. Each phase approaches the process of working with data from different perspectives: data as material, data as artifact, and data publics [2]. In the first phase, data is presented to students *as a material* for the design process, focusing on its limits and possibilities: students work with existing dataset to design a data-driven communication artifact. In the second phase, *data as artifact*, students are asked to map a controversial issue by collecting, analysing, and visualising data from online sources (e.g. search engines and social media platforms) using digital methods [3]. The last phase of the course, *data publics*, focus on the communication of data for the publics [4] that exist or could exist around the issues addressed in the previous phases. Borrowing from discursive design [5], students are asked to design a communication artifact that invites audiences to critically reflect and act upon the issues under study.

The context of the didactical tool presented in this paper is the second phase of the course, *data as artifact*. Looking at the web as a source of data, students follow a question-driven process: from research questions, they design protocols for data collection and analysis, and produce research findings using data visualizations. At the same time, students are also asked to document their process with the design of protocol diagrams. Phase two concludes with the delivery of an interactive report that is presented to the class and published online. In this phase, data visualization is presented as a tool supporting the iterative research process, and students are confronted with the technical and ethical aspects of data design.

2 DIDACTICAL AIM: RESEARCH PROTOCOL DIAGRAMS AS TOOLS TO ACT CRITICALLY IN DATASET DESIGN PROCESSES

In the second phase of the course, *data as artifact*, students perform analyses with digital methods [3,6]: they collect, analyse and visualise data from online platforms and publish their research methods and findings in an interactive report. During this phase, students are also asked to trace and communicate their research process by designing protocol diagrams. In this context, protocol diagrams are useful tools for the students to advance in their design process, used to remember each step of the research path and communicate it to the other group members. Furthermore, the design of protocol diagrams is a didactical tool, used to nudge students to critically reflect upon the issues related to the design of data analyses. Specifically, the design of protocol diagrams help students to reflect on three main critical aspects of data-driven analyses:

1 **Fragmentation**. Moving from research questions to research findings, students face a very fragmented process, made of several different actions, ranging from micro-actions (e.g. making a query on Google) to macro-actions (e.g. renaming with the same logic thousands of images). Furthermore, it is not always easy to isolate individual actions: where does an action start, and where does it end?

2 **Nonlinearity**. Performing analysis with online data is often thought of as a linear process made of three main phases: data collection, data analysis and data visualization. In practice, when students start to work on their analysis, they follow a non-linear path: often multiple processes develop in a parallel way, and some of them can reach a dead end. Sometimes parts of the process must be reworked due to errors or changes in the approach.

3 **Uncertainty**. The process contains several degrees of uncertainty. Uncertainty can be related to the data itself, the methods used to collect it, or the sources providing information. In all these cases external factors may influence the results of the analysis. Furthermore, the different authorial choices made by the students add a layer of uncertainty to the process.

2.1 Expected learning outcomes

The design of protocol diagrams helps students before, during and after the analysis. Before starting the analysis, the requirement of representing the process in a protocol diagram nudges students to have a more structured workflow, as they are forced to lay out a plan of what they wish to achieve, and of the actions they need to perform to reach their goal. During the analysis, protocol diagrams can help students to keep a record of what they are doing and what they have done. Furthermore, protocol diagrams can be used as a baseline for discussion among group members and, during revision sessions, to present and explain design choices to the teachers. Finally, at the end of the analysis, protocol diagrams become dissemination tools, used to clearly communicate the research path to other users.

Since during the process students are really focused on the output of their analysis, the risk is that they somehow would tend to not reflect or ignore the critical issues related to data-based analysis listed above. From a didactical point of view, the risk is that students will perform several actions in a mechanical way without fully acknowledging the different choices they made throughout the process. With the task of designing protocol diagrams, teachers can assure that students will reflect upon the choices they made during the analysis. Specifically, the task of designing protocol diagrams seek to achieve the following expected learning outcomes:

1 **To preview**: Students will be able to elicit in advance the possible path they should follow to complete the analysis;

2 **To disentangle:** Students will be able to isolate and describe individual actions in the analysis;

3 **To correlate:** Students will be able to identify how actions are related among each others;

4 **To legitimize/empower:** Students will be able to identify which actions were driven by their authorial choices;

5 **To disseminate:** Students will be able to communicate the research process they have followed.

3 DIDACTIC METHOD

The design of protocol diagrams is introduced as a specific task in the second phase of the course. Besides taking care of data collection, data visualisation and data analysis, students are asked to design a protocol diagram for each analysis they choose to perform. In order to make protocol diagram design really work as a learning moment, we adopt a specific teaching methodology, which leaves room for students to learn by doing. Specifically, we offer a wide definition of protocol diagrams, accompanied by the presentation of examples of well-designed diagrams. In addition, to help students in the design process, we provide them with two macro groups of indications: a series of intended functions that the diagram must satisfy, and a collection of good practices to keep in mind during the design phase.

3.1 A broad definition of protocol diagrams

Protocol diagrams are introduced with a deliberately broad definition, which loosely identifies the perimeter within which students can move in the design process. First, the concept of research protocol is introduced. A research protocol is described to the students as the result of the operationalisation of a research question. Borrowing from the field of computational literary criticism, in which the term operationalizing refers to the translation of concepts into measurements [7], here research protocols are presented as the operations (or actions) needed to move from research questions to research findings. In this sense, protocols are described as the steps required to proceed from research questions to the other parts of the analytical process: from data collection to data analysis with data visualisation. Second, protocol diagrams are introduced as the visual representation of research protocols. Here, a specifically broad definition is used: protocol diagrams are simply described as “flowcharts” that should present “research questions and research steps in a compact visual form” [8]. Mentioning the visual archetype of the flow diagram helps design students to recall a repertoire of known references that they can use in the design process. At the same time, describing the diagram as a compact visual artifact poses a stimulating design challenge to the students, in that they can appreciate the need to condense a very complex process into a limited amount of space.

In order to further support students in their design process, after a broad definition, two aspects of protocol diagrams are introduced. First, we present their intended functions (what protocol diagrams are for); second, we give some indications on good practices to follow (or simply things to keep in mind) when designing a protocol diagram.

3.2 Clarifying intended functions of protocol diagrams

Diagrams are presented to the students for their dual function, performed in two different stages of the process: as negotiation tools and as dissemination devices.

During the initial stages of the research, the protocol diagram is a *negotiation tool* used to keep track of the research path (including failures and trials). Here it should be seen as a living document [8, 9], constantly updated as new directions are explored, or some others are abandoned. The main functions of the protocol diagram in this phase are to keep track of the complexity of the process, to encourage discussion among students and to guide them in their analytical decisions.

Once students complete their research, they deliver an interactive report presenting their research. In this phase the protocol diagram becomes a *dissemination device,* addressed to those who are interested in learning about the research path followed by the students. In this stage, the main functions of the protocol diagram are to communicate outside of the student group, to expose the research process making it inspectable from others who can replicate it, hence building trust among researcher and reader.

3.3 Presenting good practices in the design of protocol diagrams

In addition to presenting the functions of protocol diagrams, we provide students with a set of good practices to follow in the design process, discussed with the help of examples of well designed protocol diagrams. The following good practices are drawn from the observation of protocol diagrams designed by students in the previous years of the course:

- **Self explanatory and self sustainable**. Protocol diagrams should present the research process in a clear way, without the need of additional materials. The research process should be intelligible just by looking at the diagram, without the need of other resources.
- **Visuals over texts**. As the protocol diagram should offer a quick and fast rendition of the research process, it is better to prefer visual elements (icons and schemes) over long textual representations, which require more effort to be decoded from the user.
- **Visual hierarchy**. As often protocol diagrams are very complex and dense, it is helpful to build a visual hierarchy to help the user approach the diagram. For example, by creating different layers of reading between macro actions (e.g. data collection, data analysis, data visualisation) and micro actions (e.g. data scraping, data transformation processes, content analysis).
- **Consistent visual grammar.** As with any other data visualisation, protocol diagrams should adopt a consistent visual grammar. Using a consistent visual language to differentiate among the various elements included (used tools, performed actions, obtained results) reduces the effort required to decode the diagram, making it easier to understand.

4 RESULTS

The task of designing protocol diagrams is assigned to students of the course since six editions. Since 2014 — the 10th edition of the course — a total of 272 students, divided in 48 working groups, designed 334 protocols diagrams (see table 1). To make an analysis of the results, we took as example one protocol diagram for each working group, across the six editions (fig 1). In this section we present some preliminary results stemming from the observation of students' projects over the years. In general, we have analysed students' results with a focus on the five learning outcomes listed above. By looking at the evolution of students' visual choices over the years, it is possible to observe the overall evolution of protocol diagrams and how the review of previous work helps students to identify, understand and tune best practices identified by previous colleagues..

Table 1. Number of protocol diagrams designed by students in the last 6 editions of the course.

Course edition	N° of groups	N° of students	N° of protocol diagrams
10	8	41	35
11	5	36	20
12	8	35	71
13	7	37	66
14	10	63	90
15	10	60	52
TOT	**48**	**272**	**334**

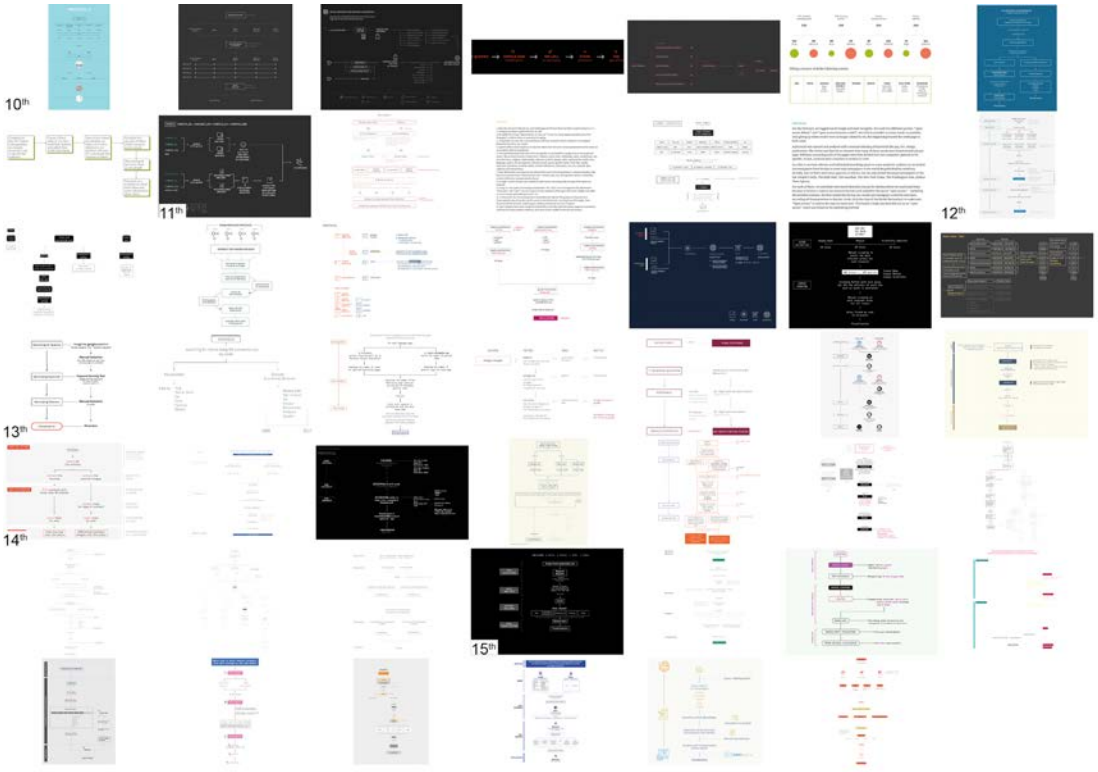


Figure 1. 48 protocol diagrams, from the 10th to the 15th edition of the course.

4.1 Preferring visuals over text

Looking at the full sample of protocol diagrams, we can say that most of the working groups have achieved the goal of designing *self-explanatory* and *self-sustainable* protocol diagrams, choosing, year after year, visual elements over textual ones. During the 11th edition, two working groups designed only textual protocol diagrams which turned out to be very difficult to read and evaluate, unlike the visual ones (See Figure 2).

The increasing choice of representing each action of the protocol in a visual form may be read as a signal of a higher awareness by the students of the specificity of each individual action of the process. The exercise of choosing the right visual element (icons, arrows and other diagrammatic representations) (see Figure 3) may in fact stimulate students to give “an identity” to each action (see learning outcome 2), to define groups of similar actions, and to reflect on how these actions relate to each other (see learning outcome 3).

4.2 Archetype of the flow diagram largely used

Especially in recent years, the archetype of the flow diagram [10] is well recognizable (boxes of different shape and size connected by arrows and lines), and horizontal diagrams have rarefied, leaving room

for the vertical diagram form (Figure 4). The increasing use of the flow diagram structure suggests that the design of protocol diagrams stimulates students to conceive the research process as a set of individual actions connected in various ways (see learning outcomes 3). It is also possible to say that splitting the process into a series of micro actions, represented by the boxes in the flow diagram, allows students to design the research process before starting it, therefore helping them in the elicitation of their research path beforehand (see learning outcome 1).

4.3 Linking visually the same type of actions

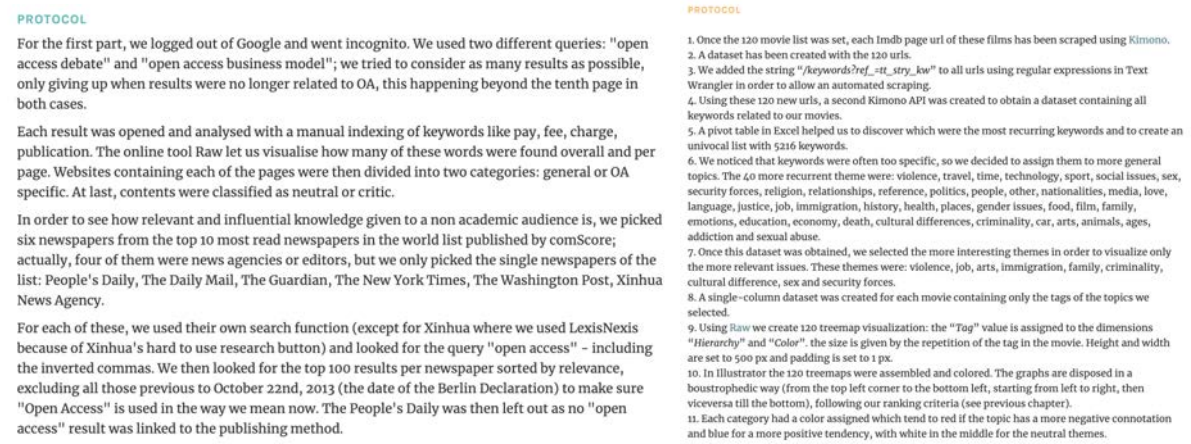
In general, most working groups prefer a white or light background, with dark texts and colored elements, using contrast to allow for straightforward decoding by the reader, able to understand the research process that is being followed. The main color is usually chosen according to the visual style of the interactive report and, for that reason, can be any color of the chromatic spectrum. Regardless of the different visual styles adopted by each group, a consistent visual language to differentiate among the various elements included in the diagram has progressively been adopted, enabling a better decoding of the represented processes. Various strategies have been used over time by the students to visually link different classes of actions in the diagrams (e.g. use of icons, same-shape blocks, visual clustering). For instance, in the diagrams from the last editions of the course, the same type of action is usually associated with a color or shape, whereas in the older editions of the course, diagrams make less use of colors and shapes to create visual hierarchies (Figure 5)

The increasing use of a complex visual grammar to distinguish different types of actions, suggests how the design of protocol diagrams also teach students to critically reflect on the different types of actions that make up the entire research process (see learning outcomes 2 and 3), including their personal authorial choices (see learning outcome 4).

4.4 Repurposing effective visual tricks from previous years' works

While in the beginning the shape and the structure of the diagrams were not particularly homogeneous, over time each working group applied the four good practices mentioned above also inspired by the protocol diagrams designed in past editions by former students. Specifically, by observing the sample, it is evident that some visual tricks used effectively during an edition become recurrent in subsequent years. For instance, the tendency to regroup actions of the process in macro phases, which started during the *11th* edition, has become recurrent in the following years. Instead, other visual experiments, such as the integration of zoomed visualization were not emulated in the following years.

The adoption of winning solutions and the abandonment of less effective ones shows that, by reading protocols made in the previous years, students test themselves the efficiency of different protocol design. This activity informs them on how to better communicate their results (see learning outcome 5). For example, visually grouping actions of the same type (data collection, data analysis, data visualization) is a strategy adopted to communicate in a simple way the complexity of the research process. (See Figure 6)



PROTOCOL

For the first part, we logged out of Google and went incognito. We used two different queries: "open access debate" and "open access business model"; we tried to consider as many results as possible, only giving up when results were no longer related to OA, this happening beyond the tenth page in both cases.

Each result was opened and analysed with a manual indexing of keywords like pay, fee, charge, publication. The online tool Raw let us visualise how many of these words were found overall and per page. Websites containing each of the pages were then divided into two categories: general or OA specific. At last, contents were classified as neutral or critic.

In order to see how relevant and influential knowledge given to a non academic audience is, we picked six newspapers from the top 10 most read newspapers in the world list published by comScore; actually, four of them were news agencies or editors, but we only picked the single newspapers of the list: People's Daily, The Daily Mail, The Guardian, The New York Times, The Washington Post, Xinhua News Agency.

For each of these, we used their own search function (except for Xinhua where we used LexisNexis because of Xinhua's hard to use research button) and looked for the query "open access" - including the inverted commas. We then looked for the top 100 results per newspaper sorted by relevance, excluding all those previous to October 22nd, 2013 (the date of the Berlin Declaration) to make sure "Open Access" is used in the way we mean now. The People's Daily was then left out as no "open access" result was linked to the publishing method.

PROTOCOL

1. Once the 120 movie list was set, each Imdb page url of these films has been scraped using Kimono.
2. A dataset has been created with the 120 urls.
3. We added the string "*/keywords?ref_=tt_stry_kw*" to all urls using regular expressions in Text Wrangler in order to allow an automated scraping.
4. Using these 120 new urls, a second Kimono API was created to obtain a dataset containing all keywords related to our movies.
5. A pivot table in Excel helped us to discover which were the most recurring keywords and to create an univocal list with 5216 keywords.
6. We noticed that keywords were often too specific, so we decided to assign them to more general topics. The 40 more recurrent theme were: violence, travel, time, technology, sport, social issues, sex, security forces, religion, relationships, reference, politics, people, other, nationalities, media, love, language, justice, job, immigration, history, health, places, gender issues, food, film, family, emotions, education, economy, death, cultural differences, criminality, car, arts, animals, ages, addiction and sexual abuse.
7. Once this dataset was obtained, we selected the more interesting themes in order to visualize only the more relevant issues. These themes were: violence, job, arts, immigration, family, criminality, cultural difference, sex and security forces.
8. A single-column dataset was created for each movie containing only the tags of the topics we selected.
9. Using Raw we create 120 treemap visualization: the "*Tag*" value is assigned to the dimensions "*Hierarchy*" and "*Color*". the size is given by the repetition of the tag in the movie. Height and width are set to 500 px and padding is set to 1 px.
10. In Illustrator the 120 treemaps were assembled and colored. The graphs are disposed in a boustrophedic way (from the top left corner to the bottom left, starting from left to right, then viceversa till the bottom), following our ranking criteria (see previous chapter).
11. Each category had a color assigned which tend to red if the topic has a more negative connotation and blue for a more positive tendency, with white in the middle for the neutral themes.

Figure 2. Two textual protocol diagrams from the 11th edition. On the left a protocol diagram by Mauro Abbattista, Piero Barbieri, Maria Elena Besana, Chiara Cirella, Manuel Impellizzeri, Andrea Lacavalla. On the right a protocol diagram by Camila Borrero, Pietro Cedone, Lucia Cosma, Sebastián Forero, Giovanni Rabuffetti, Ziqiao Xie.

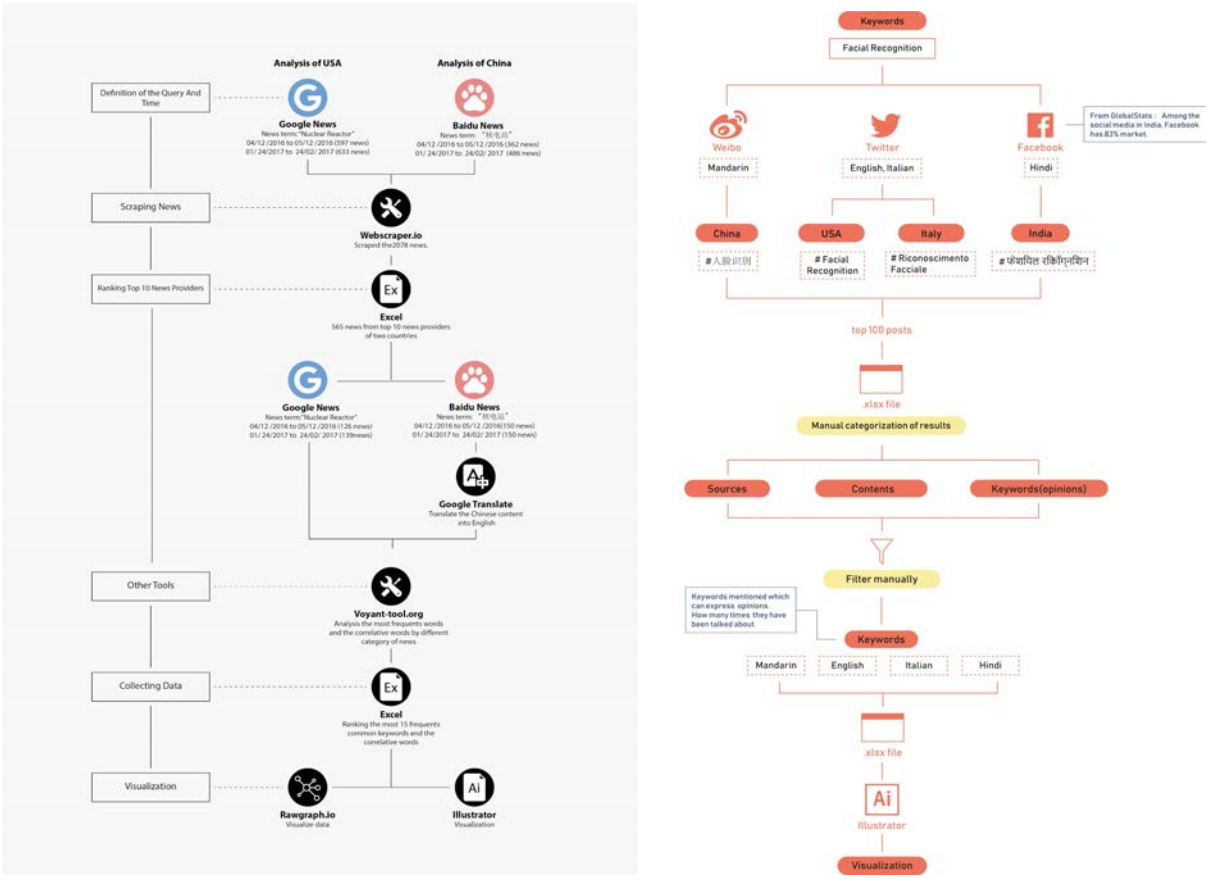


Figure 3. Diagrams including ìcons and logos have increased over the years. On the left, a protocol diagram from the 11th edition by Manli Zhu, Simone Casartelli, Xiaoxi Huang, Xuechun Zhao, Yue Qiu. On the right, a protocol diagram from the 15th edition.

Figure 4. Horizontal protocol diagrams are more common in the first editions of the course. A horizontal protocol diagram from the 10th edition by Marianna Caserta, Francesca Casnati, Chiara Cingottini, Giulia Sonzogni, Ilaria Tedoldi

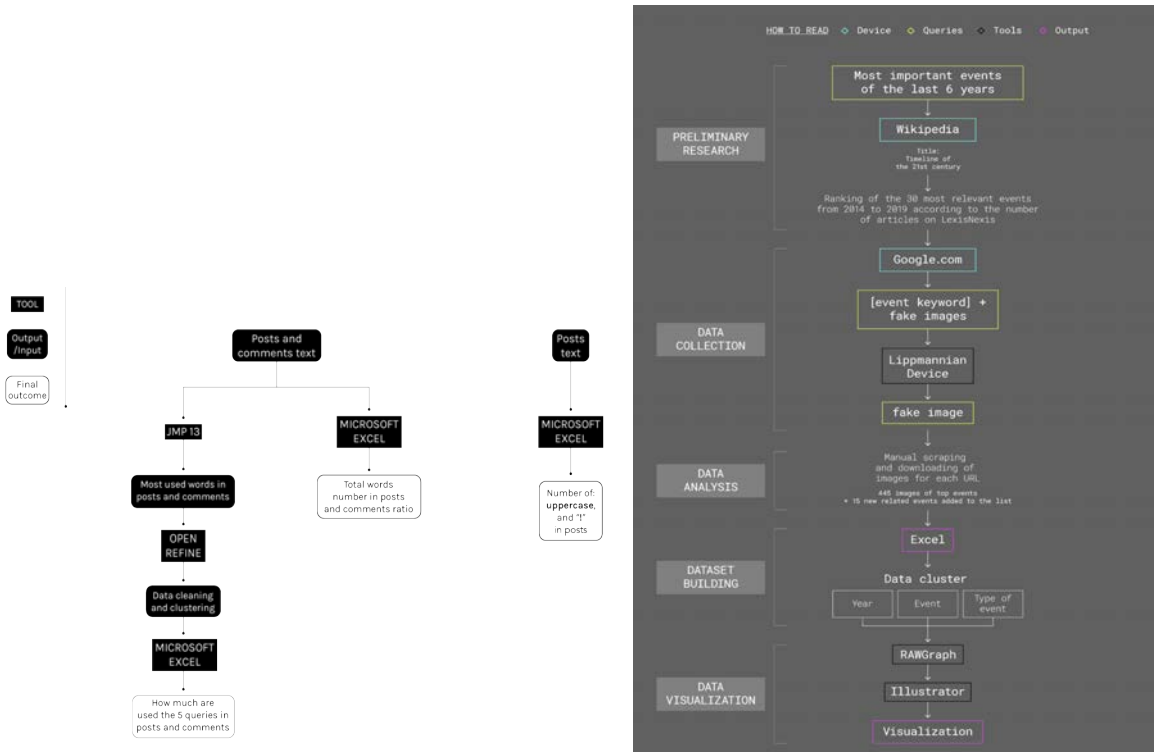


Figure 5. On the left a protocol diagram from the 12th edition by Vincenzo Bisceglia, Valeria Brienza, Nicola Cerioli, Matteo Maria Tartaglia. They differentiate actions using different shapes. On the right a protocol diagram from the 15th edition by Gabriele Della Pepa, Andrea Elena Febres Medina, Caterina Ghio, Francesca Granzotto, Paola Rondi, Elena Stefani. They differentiate actions by using different colors.

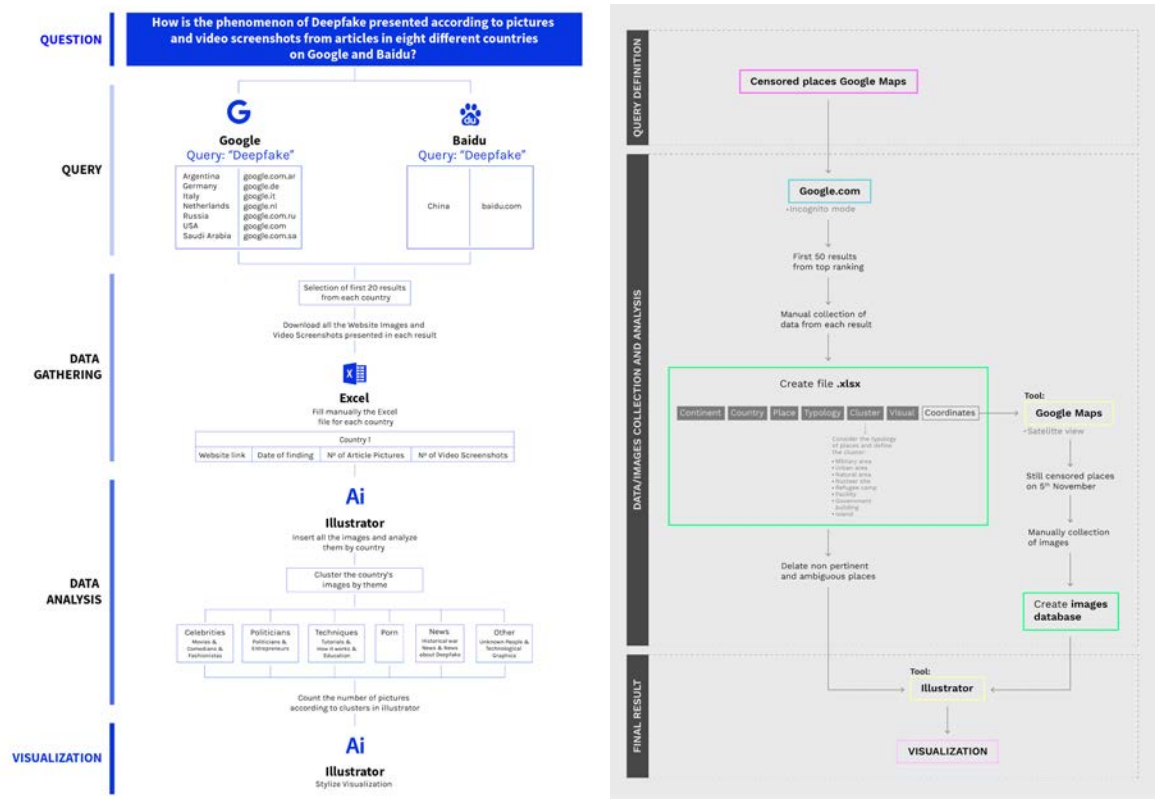


Figure 6. Two protocol diagrams from the 15th edition, in which similar actions are visually grouped together. On the left a protocol diagram by Andrej Cattaneo, Ivana Riva, Noura Sammoura, Maria del Pilar Suarez Anzorena, Arthur van der Werf, Yueling Wu. On the right a protocol diagram by Antonella Autuori, Agnese Bartolucci, Elena Catani, Mattia Cittadino, Anna Gazza and Barbara Vanoli.

5 CONCLUSIONS

In this paper, we presented how the design of research protocol diagrams has been adapted as a didactic tool for teaching critical aspects of the dataset design process to communication design students. We have first introduced the didactic context: a five-month data visualisation course of the design faculty of Politecnico di Milano that gradually introduces students to the communication of data in complex situations. In the context of the phase dedicated to the collection, visualization and analysis of data from online sources, we have introduced the didactic innovation presented in this paper: the task of designing protocol diagrams as a tool to nudge students to critically reflect on three critical aspects of the dataset design process (fragmentation, non-linearity, uncertainty). We have then argued that the design of protocol diagrams is an effective didactic tool with respect to five expected learning outcomes: preview, disentangle, correlate, legitimize, and disseminate. Then, we have described the didactic method used to introduce the design of protocol diagrams to students: a specially broad definition of protocol diagrams, coupled with the presentation of protocol diagrams main functions and a set of good practices discussed with examples. In the final part of the paper we have discussed some preliminary results stemming from the observation of students' projects over the years, with a focus on the five listed learning outcomes.

Observing the evolution in the design of protocol diagrams over the years, it is possible to empirically evaluate the effectiveness of this didactical initiative as a means to teach critical aspects of the dataset design process. Specifically, designing protocol diagrams 1) helps students in previewing their research path by splitting the process into a series of actions represented by boxes in the diagram; 2) improves students' awareness of the role of each individual action of the process, pushing them to give an "identity" to each action with a consistent and well-defined visual grammar; 3) stimulates students to conceive the research process as a set of distinguished actions and to reflect on how these actions relate to each other; 4) teaches students to critically reflect on their arbitrary choices; 5) forces students to identify the most effective visual solutions for disseminating the research process, by considering the final users and reviewing visual solutions successfully adopted in the past.

ACKNOWLEDGEMENTS

We want to acknowledge all the teachers and assistants that participated in the course throughout the years, in particular the ones that participated in the latest edition: Salvatore Zingale, Simone Vantini, Andrea Benedetti, Alessandra Facchin, Lea Fabiano and Tommaso Elli.

REFERENCES

[1] DensityDesign, Mediazioni Algoritmiche. Designing in a data-intensive society, *Final Synthesis Design Studio 2019/2020*, 2020. Retrieved from https://densitydesign.github.io/teaching-dd15/

[2] M. Mauri, G. Colombo, Á. Briones, P. Ciuccarelli, "Teaching the critical role of designers in the data society: the DensityDesign approach", *Fifth International Conference for Design Education Researchers. METU Department of Industrial Design,* p. 183-195, 2019. *http://dx.doi.org/10.21606/learnxdesign.2019.12046*

[3] R. Rogers, *Digital Methods,* Cambridge: MIT press, 2013.

[4] C. DiSalvo, "Design and the Construction of Publics." *Design issues 25.1,* p. 48-63, 2009. https://doi.org/10.1162/desi.2009.25.1.48

[5] B. M. Tharp & S. M. Tharp, *Discursive design: critical, speculative, and alternative things,* Cambridge: MIT press, 2019.

[6] R. Rogers, *Doing Digital Methods,* London: SAGE Publications, 2019.

[7] F. Moretti, "Operationalizing": or, the function of measurement in modern literary theory, *New Left Review*, 84 (November-December), p. 1-13, 2013.

[8] S. Niederer & G. Colombo. "Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-platform Analysis, and Public Participation", *Diseña,* 14, p. 40-67, 2019. https://doi.org/10.7764/disena.14.40-67

[9] Mauri, M., & Ciuccarelli, P. “Designing diagrams for social issues” In *Proceedings of DRS2016: Design + Research + Society - Future-Focused Thinking* (Vol. 3, pp. 1–17), 2016. https://doi.org/10.21606/drs.2016.185

[10] Harris, R. L. *Information Graphics: A Comprehensive Illustrated Reference*. New York, NY, USA: Oxford University Press, p. 156, 2000.