Revista de Comunicación Digital

Staging and storing data sprint-based research results: a communication design approach

Organización y almacenamiento de los resultados de la investigación basada en *data sprints*: un enfoque desde el diseño de la comunicación

Beatrice Gobbo
beatrice.gobbo@polimi.it
Density Design Lab, Politecnico di Milano

Elena Aversa
elena.aversa@polimi.it
Density Design Lab, Politecnico di Milano

Andrea Benedetti
andrea.benedetti@polimi.it
Density Design Lab, Politecnico di Milano

Gabriele Colombo
gcolombo@iuav.it
King's College London

María de los Ángeles Briones Rojas
mariade.briones@polimi.it
Density Design Lab, Politecnico di Milano

Tommaso Elli
tommaso.elli@polimi.it
Density Design Lab, Politecnico di Milano

Michele Mauri
michele.mauri@polimi.it
Density Design Lab, Politecnico di Milano

Received: 14/01/2022
Accepted: 30/05/2022

ABSTRACT

Data sprints are multidisciplinary, time-limited, practice-oriented, and group activities that explore complex issues through data gathering, visualization, and analysis. Data sprints usually conclude with final presentations where participants showcase the research process they followed and the achieved results. Research centers conducting data sprint activities store research results repurposing existing formats (i.e., Wikis) not directly tailored to organizing, archiving, and disseminating online data sprints' output. Starting from the experience of a data sprint on Digital Methods, the paper describes the design process of a digital format which condenses and reinterprets good practices already in use in current formats for disseminating data sprints results. The design process of the format is structured in two steps: (i) designing and (ii) staging the format. The former aims to dissect and recompose data sprint results in hierarchical block structures filled with variable contents; the latter describes outcomes that emerged during a preliminary evaluation phase with experts and a stress-test activity. The format is a container that participants can fill at the end of a data sprint to organize, archive, and present research results. Evaluation activities confirmed the efficiency of the format, which is distinguished by its nested structure and flexibility with different content.

KEYWORDS

digital methods, communication design, scientific dissemination, data sprint

How to cite this article:

Gobbo, B.; Briones Rojas, M. A.; Aversa, E.; Elli, T.; Benedetti, A.; Mauri, M. & Colombo, G. (2022). Staging and storing data sprint-based research results: a communication design approach. *Dígitos. Revista de Comunicación Digital,* 8: 87-108. DOI: 10.7203/drdcd.v1i8.241

RESUMEN

Los *data sprints* son actividades multidisciplinares, de duración limitada, orientadas a la práctica y en grupo, que exploran cuestiones complejas mediante la recopilación, visualización y análisis de datos. Los *data sprints* suelen concluir con presentaciones finales en las que los participantes muestran el proceso de investigación que han seguido y los resultados obtenidos. Los centros de investigación que llevan a cabo actividades de data sprint almacenan los resultados de la investigación reutilizando formatos existentes (por ejemplo, Wikis) que no están directamente adaptados para organizar, archivar y difundir los resultados de los *data sprints* en línea. Partiendo de la experiencia de un sprint de datos sobre *Digital Methods*, el artículo describe el proceso de diseño de un formato digital que condensa y reinterpreta las buenas prácticas ya utilizadas en los formatos actuales para difundir los resultados de los *data sprints*. El proceso de diseño del formato se estructura en dos pasos: (i) el diseño y (ii) la puesta en escena del formato. La primera tiene como objetivo diseccionar y recomponer los resultados del *data sprint* en estructuras jerárquicas de bloques llenas de contenidos variables; la segunda describe los resultados que surgieron durante una fase de evaluación preliminar con expertos y una actividad de prueba de esfuerzo. El formato es un contenedor que los participantes pueden llenar al final de un sprint de datos para organizar, archivar y presentar los resultados de la investigación. Las actividades de evaluación confirmaron la eficacia del formato, que se distingue por su estructura anidada y su flexibilidad con diferentes contenidos.

PALABRAS CLAVE

métodos digitales, diseño de comunicación, difusión científica, *data sprint*

Staging and storing data sprint-based research results: a communication design approach

1. Introduction: disseminating data sprint-based research

Although the format of the sprint was born in the computer science context, over the years sprints have been applied in different areas, from service design to the more recent academic context for data gathering and analysis (Mauri et al., 2019, p. 162). Within the academic context, data sprints are multidisciplinary events where teams of researchers, doctoral and master students from various disciplines, such as Science and Technology Studies, Future Studies, Sociology, Media Studies and Digital Humanities (Bounegru et al., 2018; Ciuccarelli and Elli, 2019) come together to best achieve a given goal through practice-based research. As stated by Mauri in 2019, data sprints usually address wicked problems (Mauri et al., 2019, p. 163), namely situations where it is not easy to define an initial design brief and where different professionals must agree on the type of result they want to realize (Rittel and Webber, 1973). In this regard, communication design and data visualization skills are considered crucial when dealing with controversy mapping and prototyping activities (Mauri and Ciuccarelli, 2016). Moreover, data sprints are time-limited, practice-oriented, and driven by case study activities where communication design researchers play a pivotal role in *informing* and *tracking* the research process while guiding and tutoring visualization activities within the teamwork (Mauri et al., 2019, pp. 161–180). Data sprints usually conclude with a

final presentation, in which the teams show the results achieved during the days of intensive work.

However, results produced during data sprints are fragmented and scattered across disparate media (slides, notebooks, repositories, shared folders), thus they are not always easily accessible and consistently organized, and the risk that some parts might be lost is frequent. Existing dissemination formats employed during data sprints (Wikis, web reports, slides) require users to define a structure for presenting outputs and often have to make compromises to include different specific features (i.e., slides do not allow insert interactive visualizations, Wikis do not generate a visual index). Hence, it seems appropriate to design a platform that inherits from existing formats the flexibility of content publication, while integrating their specificities (i.e., hypertexts from Wikis, visual content from presentations) into a single system. In this regard, the paper delineates the role of communication designers in the context of data sprints in the ideation of a dissemination format, namely a digital *place* that allows data sprints' results to be stored, staged, and accessed after the research activities.

1.1. The *Algocount* data sprint

The starting point of the paper is the work carried out during the *Algocount* data sprint held online from January 18th to 22nd, 2021.1. Emulating the successful approaches of the Digital Methods Initiative Schools in Amsterdam and the Smart Data Sprint in Lisbon, all activities in the data sprint were conducted using Digital Methods (DM), a collection of web-native techniques for research into cultural change and societal conditions (Rogers, 2013). The presented data sprint is one of the main activities conducted within a broader multidisciplinary project that seeks to expand the current understanding of the role of algorithms in society and raise awareness about them, with a specific focus on public opinion in Italy. The data sprint was framed within the *infodemic* theme[1, 2], and it exploited the discussions that emerged on social networks during the pandemic to study the influence of algorithms in the formation of Italian public opinion on social media platforms (Facebook, Twitter, Instagram[3], TikTok, Reddit, and YouTube). The participants were divided into five sub-groups and, as a result, the data sprint produced a total of seven "experiments", namely single investigations addressing a research question. Indeed, each group carried on at least one experiment during the data sprint. Materials of the seven experiments have been used as resources for designing the dissemination format presented in the paper which is digital, web-based, and ought to preserve features of data sprint results (i.e., datasets, static and interactive visualization, video animations, GIFs, and texts). Indeed, although the seven experiments had common features, research processes were not always generalizable or inscribable into definite categories, thus it was important to design a format that on the one hand was flexible, but on the other left room for users to be thorough.

In Section 2 authors present a review of formats for dissemination of data sprint-based research, reflecting on opportunities for innovating existing solutions; Section 3 describes in detail the design process from the collection of data sprint results to the

preliminary evaluation of the format. Section 4 shows the strengths and weaknesses of the format, using one of the experiments carried on during the data sprint as a case of application. Finally, the authors envision future research opportunities and usage scenarios.

2. Existing dissemination formats for data sprint-based research

When it comes to disseminating the outcomes of data sprints, every research team is interested in disclosing the technical details of the conducted experiments, and in conveying discovered findings. Authors identified four types of formats to disseminate sprints-based research: final presentations (1), written reports (2), web platforms (3), and field guides (4). These four dissemination formats describe current practices in disseminating research results in data sprints, and they can inform how the design of a new format could integrate their specificities to build on consolidated habits of documentation.

(1) **Final presentations** organize process and findings in slides that mostly include visual contents, to encourage verbal exposition and discussion in front of other teams of researchers[4]. Presentations can be recorded and shared on streaming platforms to preserve the discussions[5]. Researchers can insert hyperlinks to resources like datasets and high-quality pictures in slides, but the format is not suitable for storing them. Researchers are used to simultaneously preparing slides using cloud-based software. Research runs fast and heuristically, therefore they hardly have the time to organize the produced materials as the work proceeds, involving the risk, after the sprint ends, of losing resources as they are moved or deleted from shared folders. Consequently, the body of work needs to be recollected to be preserved.

(2) A part of the research team is usually assigned the task of writing a **comprehensive report** after presentations and research activities are concluded. Reports include written and detailed descriptions that are not included in presentations and may link to external resources. However, they usually do not include interactive features, nor store materials[6].

(3) Another example of dissemination format is a custom **web platform**[7]: a website designed to host the outcomes of multiple data sprint experiments, which share the same context or topic because part of the same overarching research project (Venturini, 2014). A web platform resonates with other attempts aimed at renovating dissemination formats (Lupo et al., 2021). It provides comprehensive descriptions and gives space to visual and interactive contents. Additionally, it stores and gives access to materials like datasets, source files (such as networks in Gephi format, or vector-based

4. Example of a data sprint final presentation https://bit.ly/3JbB43e (accessed: 21/12/2021)

5. Example of a record of data sprint presentation https://bit.ly/3mk68UD (accessed: 21/12/2021)

6. Examples of written reports can be accessed from https://bit.ly/3yOPxgI or https://bit.ly/3J1kjYy (accessed: 21/12/2021)

7. Climaps is an exemplary case of a platform that collects sprint-based research. http://climaps.eu/ (accessed: 21/12/2021)

projects files), high-quality images, and code snippets. Web platforms proved capable of appropriately displaying contents and surviving the passing of time, however, they require considerable efforts to be designed, implemented, and maintained.

(4) **Field guides** are a further dissemination format that is aimed at reflecting on the research process conducted during sprints (Bounegru et al., 2018). Guides stress the methodological approach behind sprint-based research and aim at packaging reusable procedures to allow the adoption by other practitioners. As for web platforms, field guides require significant efforts in that researchers must consistently analyze and reproduce previous research.

While presentations and reports are produced by individual research teams during or immediately after sprints, web platforms and field guides are the results of a post-sprint re-elaboration. In particular, the design and implementation of platforms requires significant resources, and we can't assume that all data sprints result in a dedicated website.

3. Design Process

Building on the previous considerations, authors identified the need for a format designed to simplify dissemination and preservation of sprint-based research. The proposed format requires researchers to give structure to their works, it allows for collaboration between multiple users, and, thanks to a dedicated technical infrastructure, it produces a functional website that can be hosted on any server. The design process of the presented dissemination platform includes two main phases. The first phase (Figure 1), finalized to design the dissemination format interface (3.1), involves: the design of a data sprint report structure, the reorganization of all the reports into a matrix, and the iterative design of a format mock-up. The second phase (3.2) is focused on staging the format and involves focus groups and stress-test activities.

3.1. Shaping the format: restructuring data sprints results into a dissemination format

As mentioned above, results of data sprints can have different dissemination formats, often like compositions of semi-finished visuals and text in the form of slides or reports with variable structures. Hence, all research carried out during the data sprint this paper refers to have in common that they are based (1) on data gathered from social networks, (2) deal with addressing ill-defined problems, (3) they are carried out with DM, (4) are driven by the main research question and eventual sub-research questions, (5) and they include a significant number of visual outputs (i.e., visualizations, videos, GIFs).

Designing the report structure — The 20 participants of the presented data sprint — a team of researchers from the social sciences and a communication design team — worked in five groups, each studying one or two platforms and experimenting with DM auditing and reverse engineering techniques (Ziewitz, 2016) combined with data visualization. Right after the sprint, each group stored slides containing the data sprint results, and afterwards the communication design team drafted a report structure to be filled by each group.

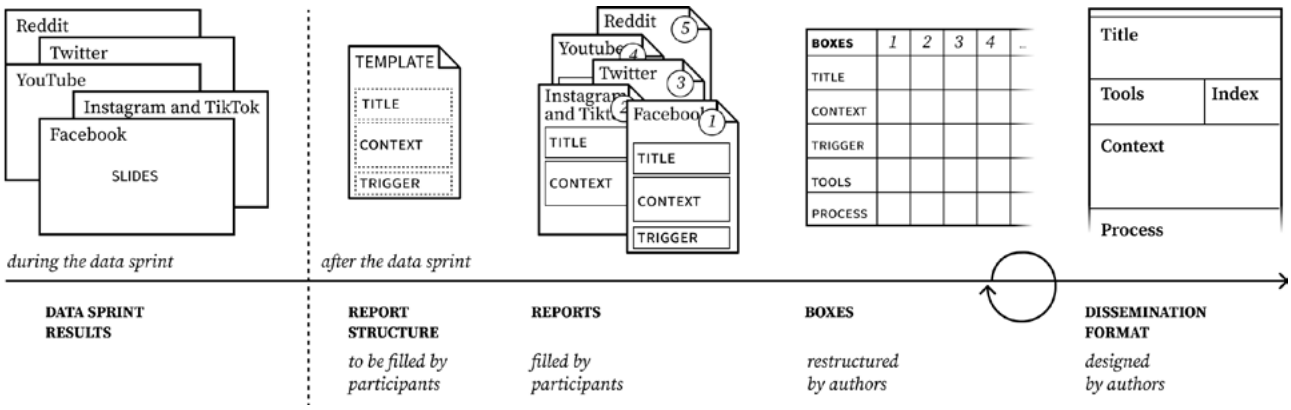


Figure 1. Diagram depicting the whole design process started from the Algocount data sprint. On the left, the starting point of the process, namely the collections of slides designed by participants. Post-data sprint activities included: the design of a report template to be filled by participants, the dissection of the filled reports, the re-organization of the gathered material into boxes and the design of the dissemination format.

The report structure was intended to collect processed information from all the groups consistently to have an arrangement that would make the experiments comparable to each other. Inspired by existing DM reports, the format required researchers to complete five main sections:

- **Introduction***:* section dedicated to describing the research object (i.e., the Facebook Newsfeed) and the research problem triggered by an inexplicable algorithmic situation.

- **Experiment(s):** section meant to collect the experiment(s) carried on during the data sprint. While some working groups produced just one experiment, others were able to carry out more.

  Each experiment includes:

  - the *research question*, specifying the research object and the addressed wicked problem.
  - the *experiment setting*, describing the digital environment in which the data collection took place, with which devices, over what time range.
  - the *outline* of the process, depicting the step-by-step procedure for carrying out the data-driven experiment.
  - the *dataset(s)* output, containing the list of gathered data.
  - the discursive *description* of the experiment.
  - the main *findings* of the experiment, emerging from visualizations.

- **General findings:** if the research is comprised of multiple research questions, this space includes a general finding that summarizes the data sprint-based research outcomes.

- **New research questions**, such as future research opportunities arising from the outcomes and outputs of the data sprint.

- **Envisioning for a dissemination format:** here, participants are invited to imagine a format for sharing their work to gather ideas that will be useful in the next phase, dedicated to the dismantling and reassembling of the works.

At the end of this process, once that all the participants have filled the report structure, authors collected five reports (one for each group), respectively including at least one experiment.

Packing into boxes: from reports to content blocks — Once the reports were completed, authors reviewed and dissected them having an overview of each and all of them at the same time. From five reports authors identified seven experiments. Given the structure of the reports previously described, it was easy to identify similar aspects in the work processes of each group. However, the diversity with which each experiment used DM also emerged. Through an activity that the authors called "*packing into boxes*", they recomposed excerpts from the reports into a matrix (Figure 2. screenshot of the "packing into boxes" output: a matrix that reconfigured the information of the experiments.). Rows of the matrix represented boxes (i.e., platform, title, tools, teaser, contextualization), while each column was dedicated to a report (Facebook, TikTok, YouTube, Twitter, Reddit, Instagram). The goal was to get an overview of the five reports and the diverse experiments within them: *what are the similarities and differences in content, and what are the existing information hierarchies*?

Thus, the matrix enabled authors to organize results according to shared components and highlight the peculiarities of each experiment. For example, the authors identified tactics and tools that have been used in more than one experiment, such as *RAWGraphs* (Mauri et al., 2017) — a data visualization tool, and Fake Name Generator — a website

	1	2	3	4
	https://docs.google.com/document/d/1kkIq9D1Y-eXdFwa6W9SLlpJe9YT3fJJZSMaA8ssdflw/edit?usp=sharing	https://drive.google.com/file/d/10enxJMZJk7NuAJVEihoUm6qRp1_BEYz0/view?usp=sharing	https://drive.google.com/file/d/10z4cDtBRBSV-HsaeUGLptOY_Qw5_uReb/view?usp=sharing	https://drive.google.com/file/d/1HNPgly1t54uEyHHUwTPiVbe/view?usp=sharing
Experiment type	AUDIT			
Platform	Facebook	TikTok	Youtube	Twitter
Title [trigger + platform + studied digital object]	Why I am seeing this infodemic on my Facebook's newsfeed?	The spread of polarization through opposite infodemic # on my TikTok's For you page	Recommendation system on YouTube platform and its specific functioning.	Twitter: hashtags and Who to follow
Tool tag	- OBS - Apify - RAWGraphs - ~~Illustrator~~ Inkscape - Premiere - Fake name generator it.fakenamegenerator.com	- RAWGraphs - ~~Illustrator~~ Inkscape - Fake name generator it.fakenamegenerator.com	- YouTube Data Tools suite - Google Spreadsheet/Excel - Gephi - RAWGraphs - ~~Illustrator~~/ Inkscape	- Python (Spyder environment) - Google Spreadsheet/Excel - R (RStudio environment) - Gephi - RAWGraphs - ~~Illustrator~~/ Inkscape
Teaser	The project focuses on: - exploring the day by day development of a newborn facebook News Feed interested in no mask related issues in Italy; - collecting data about news feed and use activity; - visualize and tell the radicalization process following platform affordance.	The project focuses on: - Exploring how the For You page becomes polarised for a user in favour of the Covid-19 vaccine and for a user against it; - Understanding how the TikTok algorithm can contribute to the circulation of Covid-19 related content;	In this project we will put the algorithmic radicalisation logic through a "stress test", using the topic of the Covid-19 in Italy.	The project focuses on: - How infodemics structure themselves o ~~- How researchers can map and visualis~~ ~~infodemic in order to be read and analys~~ - How Twitter's affordances influence the developments of infodemics
Contextualization [trigger + platform + studied digital object] > provide all the information for understanding the relevance of the studied digital object and the trigger.	Newsfeed description > "each of this buttons" from https://www.dummies.com/social-media/facebook/facebook-news-feed/ News Feed is the centerpiece of your Home page. When you log in to Facebook, you see the familiar blue bar on top and the left sidebar, but mostly you see News Feed. It's a constantly updating list of stories by and about your friends. Stories here refers to actions that your friends have taken on Facebook — things like writing a status update, sharing a photo or link, or becoming friends. You may also see stories from Pages that you follow. Facebook sometimes refers to stories as posts. More accurately, News Feed is an algorithm. It doesn't show you everything from your friends; instead, it tries to show you things it thinks you will find interesting.	[TikTok description based on https://www.nbcnews.com/better/lifestyle/what-tiktok-guide-clueless-parents-ncna1066466] TikTok is a social media app and platform you can use to create 15 second videos with hashtags, sound, filters, usually by lip-synching or dancing along with top songs. You can share the videos you create with friends, or with a wider circle of TikTok users. It was created by a chinese company in 2018. You can also follow other creators on the app without posting your own content. People can become virally famous on TikTok — many minor TikTok celebrities have 100,000-plus followers, and the biggest stars break 10 million and more. It is widely used by the "Gen Z". TikTok works through 3 main componentes	YouTube is an American online video sharing and social media platform launched in 2005. Around the world, its users watch more than one billion hours of videos each day. YouTube content creators, popularly referred to as YouTubers, upload over 100 hours of content per minute.	Twitter introduces itself as 'where people out what's happening in the world right n invites new users to 'come to Twitter to s people are talking about and join the con Through the use of hashtags, Twitter use immediately access the debate on a certa and discuss about it with others, in many strangers. Through the retweet function, decide to take a public stand and act as a of a certain information or position. Twitter has launched a program for acad researchers that allows – after the submi approval of a research project by the pro to access through the API v2 to the entir database of Twitter from its birth onward, passages could take some time, as they manual examination and approval by the

Figure 2. screenshot of the "packing into boxes" output: a matrix that reconfigured the information of the experiments.

which randomly generates fake biometric data. Hence, the authors compiled the matrix grouping tactics and tools inside a *box.* The reoccurrence of concepts laid the foundation for the design of a *glossary* of tools and tactics that would serve as a transversal means for all the experiments. Moreover, while *packing into boxes,* authors made ontological arguments about the structure of the experiments. Indeed, authors identified three nested levels of information:

- **Main-blocks** represent the higher level of information that structures type of contents transversally such as the *experiment presentation*, namely the overall description of each experiment provided by the groups. Authors identified *experiment presentation*, *context*, *research question*, *experiment setup*, *notes*, *step by step process*, *finding* and *references* as main blocks.

- **Sub-blocks**, namely the nested structure of each main block. Sub-blocks characteristics depend on the parent main block. Thus, some main blocks can lack sub-block. For example, while the *experiment presentation* can be composed by title, a brief description, a table of contents, and tools and techniques employed, the *references* block does not require an additional nesting level because it is a simple content list.

- Finally, the authors identified the **content blocks** that represent the most peculiar information, namely the text, images, videos, and multimedia contents that participants employed for describing their experiments.

Format layout design — The defined blocks were the starting point for translating the information into a graphic layout. Hence, by comparing the blocks across the experiments, authors identified common practices. For example: participants used

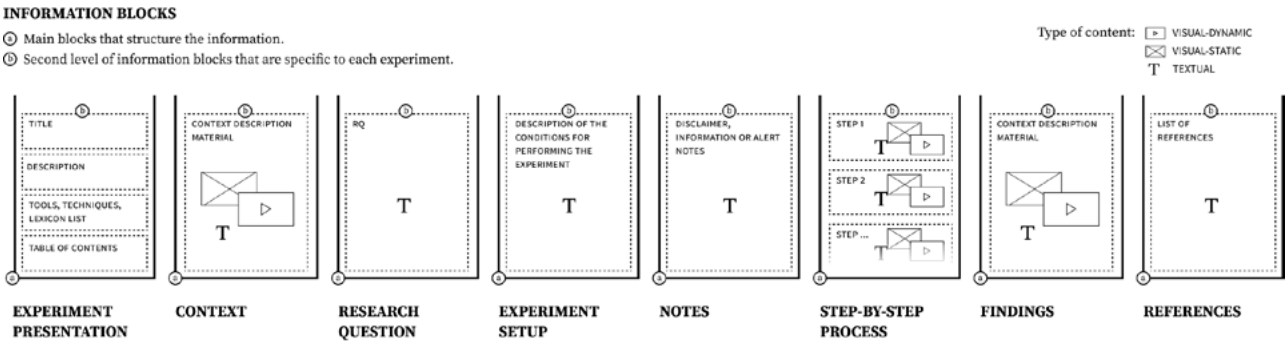


Figure 3. Diagram of information blocks that structure the experiments' content.

multiple media with various graphic styles and colors (static and animated images and visualizations, videos, links, among others) to narrate the process of the experiments. Moreover, authors noticed that experiments' blocks could be nested with different degrees of complexity and that it is necessary that the format allows the research to be read considering both the hierarchy and the sequence of the actions.

Following these observations, authors defined the design Requirements (**R**) of the format:

- **R1**: The graphic style of the format must not compete with the visual material and while emphasizing the structure of the information.

- **R2**: The format must give relevance to the use of visual and multimedia material present in the reports of each experiment.

- **R3**: The format must guarantee a fluid fruition of the contents from a desktop device.

The blocks identified above were essential to build the layout. Indeed, the authors translated them to fit into a clean visual grid that would allow to combine and nest different types of content (**R1**). Thus, while comparing blocks across the experiments, the authors decided to design a layout that could accommodate content in both one column and more than one column. This allowed authors to create diverse compositions including text, bullet points, and multimedia content. Generally, each *main-block* corresponds to a layout however, given the variety of *sub-blocks* and *content-blocks*, authors designed other layout options to meet any requirements.

Hence, as shown in Figure 4. format layout for composing the information blocks.

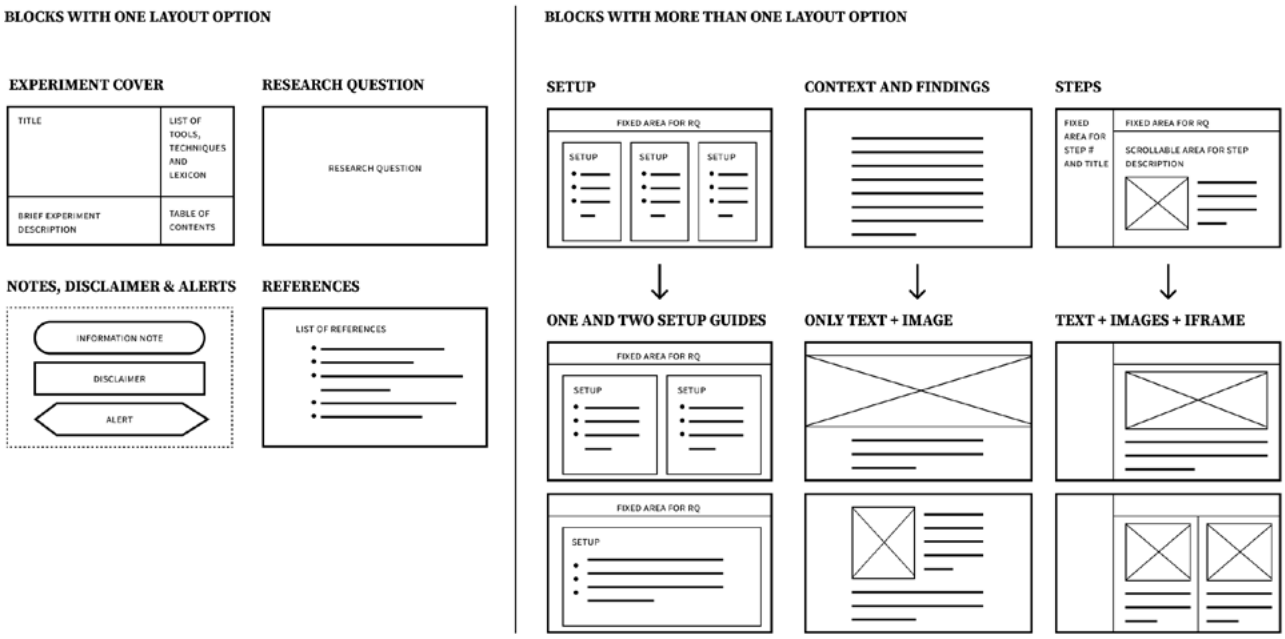


Figure 4. format layout for composing the information blocks. Some blocks have more than one layout depending on the specificity of the content. The diagram provides a sample of blocks that have been used in one experiment. Possible connections between experiments are depicted later (Figure 5. Overall structure of the dissemination format. Each project is a collection of experiments sharing a glossary which can be organized into categories, such as tools and techniques.).

Some blocks have more than one layout depending on the specificity of the content. The diagram provides a sample of blocks that have been used in one experiment. Possible connections between experiments are depicted later (Figure 5); the *experiment presentation*, the *research question*, the *notes,* and the *references* blocks have a fixed

layout. Differently, the *experiment setup*, the *context*, the *findings,* and the *step-by-step* blocks require multicolumn layouts or multimedia compositions to properly address design requirements (**R2**).

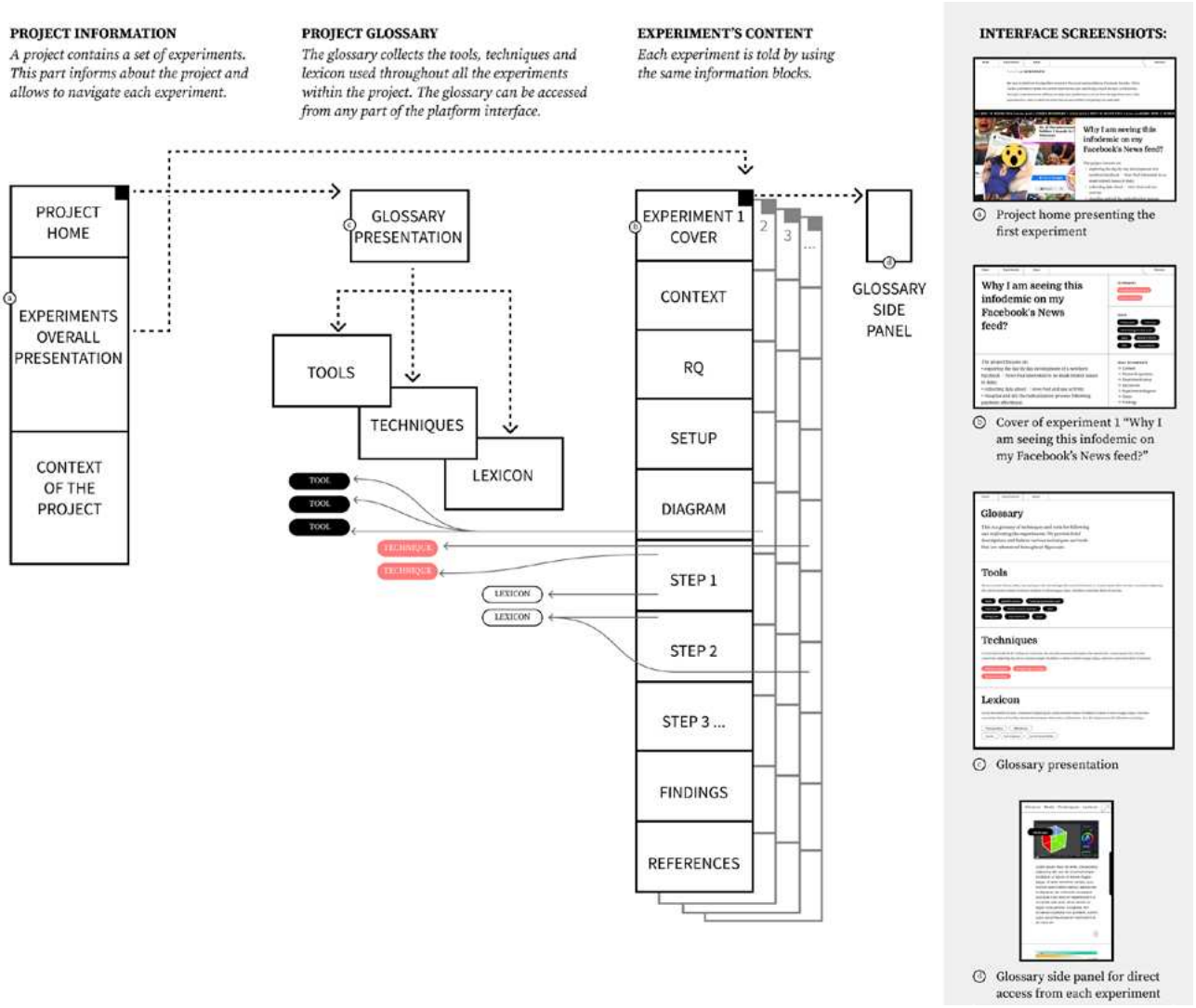


Figure 5. Overall structure of the dissemination format. Each project is a collection of experiments sharing a glossary which can be organized into categories, such as tools and techniques.

To guarantee fluid fruition of the contents, authors exploited interactivity giving priority to scroll, designing sticky and fixed elements, and including tools as magnifying lenses ad zoom for multimedia contents. For example, the layout envisioned for the step-by-step process (Figure 4. format layout for composing the information blocks. Some blocks have more than one layout depending on the specificity of the content. The diagram provides a sample of blocks that have been used in one experiment. Possible connections between experiments are depicted later (Figure 5)., last column) includes sticky elements (**R3**) which could support and improve the readability of the contents embedded into the scrollable area.

Finally, to make connections between experiments, the authors designed a layout

dedicated to the *Glossary* where tools and techniques could be gathered and described. The process of standardization of the possible compositions of the information is fundamental to achieve graphic consistency between all experiments, as well as to guarantee to those who will read the experiments a consistent way to find the needed information (Figure 5. Overall structure of the dissemination format. Each project is a collection of experiments sharing a glossary which can be organized into categories, such as tools and techniques.).

2.2. Staging the format: assessment through discussion and simulation of use

Once the layout of the format was defined, the authors addressed the focus group and stress-test activities to assess the format layout design.

Focus group — A wide range of disciplines (Silverman, 2016) has proposed *focus group methodologies* as "an informal discussion among selected individuals" (Becket al. 1986: 73) to gather valuable information about specific subjects. In such groups, participants are free to articulate their opinions in a spontaneous manner, which may facilitate the collection of more in-depth insights. On such intentions, the authors conducted a focus group (see Figure 6. Focus group participants from Political Sciences and Design fields.) of twelve participants (5 facilitators and 7 participants; almost all of them participated in the data sprint) belonging to both social sciences and communication design research fields, to test the structure, layout, usability, and flexibility of the dissemination template. The general aim was to collect feedback on those aspects which were not anticipated as critical during the design development. The focus group lasted approximately three hours and was structured in two main activities: a guided exploration of the format mock-up and a participatory discussion.

Figure 6. Focus group participants from Political Sciences and Design fields.

The first activity was carried out by a facilitator. During this phase, participants were invited to critically reflect on the different features showed. The guided exploration focused on different aspects concerning the layout design. For example, the facilitator asked participants to reflect on the flexibility of the column layout, on the design choices

of the Glossary, and on the arrangement of the visual elements in the cover. Participants were likewise provided with clear-cut questions to direct and nourish the discussion on the aforementioned aspects. For instance, when asked "What is the most suitable layout for your research?" most participants confirmed they would benefit from having the possibility to choose between a single-column or a two-column layout. At this stage the comments were rather positive and focused on the visual choices, the effectiveness of the layout and the usability of the platform.

Throughout the second part of active feedback exchange, the discussion evolved organically and was moderated by the facilitators. This activity was conceived to assess the completeness and flexibility of the template by asking the participants to imagine the application of other case studies. More specifically, they were asked to identify three features they found ineffective or complex along with the three they found useful. The discussion started with the latter, which proved to be a fruitful starting point for thinking about how to further improve specific details. For example, while the Glossary was found to be immediate and essential, the participants also pinpointed the following fine tunings: a) adding a customizable section and b) standardizing the layout for each section.

This phase was critical to confirm the effectiveness of the structure and the layout of the format to narrate the research process. Moreover, it emerged that the structure and the layout were reasonable flexible to accommodate different types of research. Finally, it allowed to identify the elements of the template to be improved.

Stress-test — Following the collection, analysis, and implementation of the feedback that emerged from the focus group, the succeeding activity involved a *stress-test.* The stress-test activity focused on evaluating the overall performance of the format. To assess it, authors populated a prototype version of the web-based format[8] with other contents. Although in the stress testing methodology, systems are intensively tested above normal operational capacity through automated techniques (Beizer, 2003; Nelson, 2009), authors' testing method consisted in creating a high-fidelity mockup of the format with two sample projects carried out during the DMI Summer and Winter schools and stored as reports on the website of the school Such a method allowed the authors to examine the efficiency of the format when dealing with experiments with variable structures and fields of application.

As a result, the focus group and the stress-test assessed the overall usability of the format, both when using experiments coming from the *infodemic* data sprint and from other similar contexts. Criticalities and suggestions emerged from the feedback session have been progressively solved and integrated within the format. For example, the authors refined the reference section and enriched the glossary creating cross-references between techniques and tools.

8. The prototype version of the web-based format is an early-stage Content Management System enabling authors to compose each experiment according to the designed blocks and layouts.

3. Using the format: a case study from the *Algocount* data sprint

As emerged from focus groups and stress tests, the format helps in disseminating data sprint research's processes and results, besides answering storage and durability issues of the information from the experiments over time, since it fills the gap described in Section 2 arisen both from literature and authors' experiences in the field. To better illustrate its usage, in the following section the authors will describe one of the seven experiments developed in the context of the data sprint, shedding light on the aspects of the format that make it a promising tool during usage.

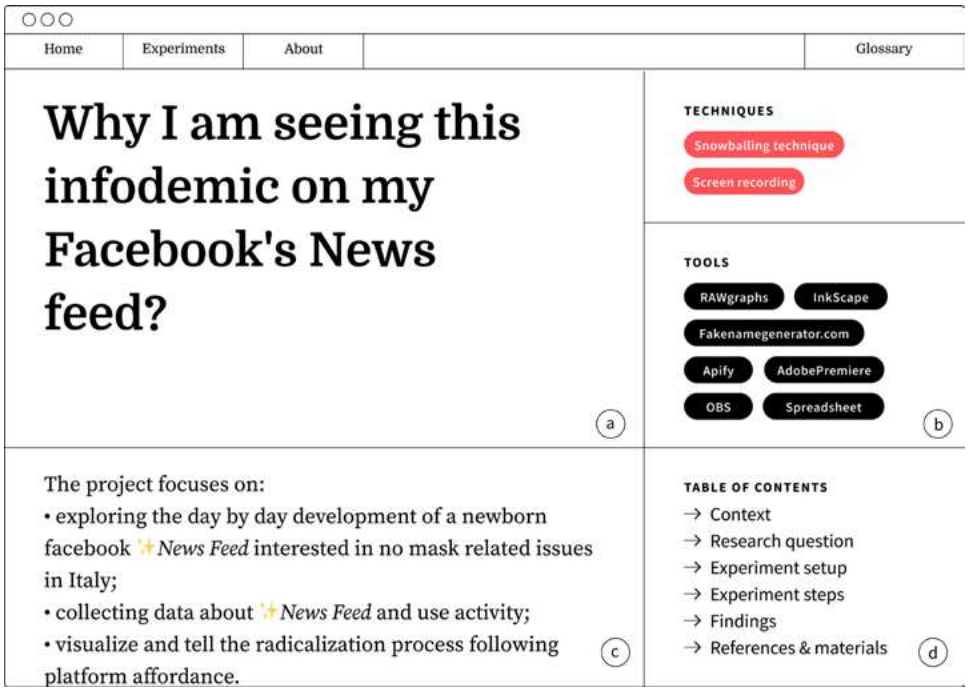


Figure 7. General presentation of the experiment "Why am I seeing this infodemic on my Facebook's News Feed" proposed as case study for this paper.

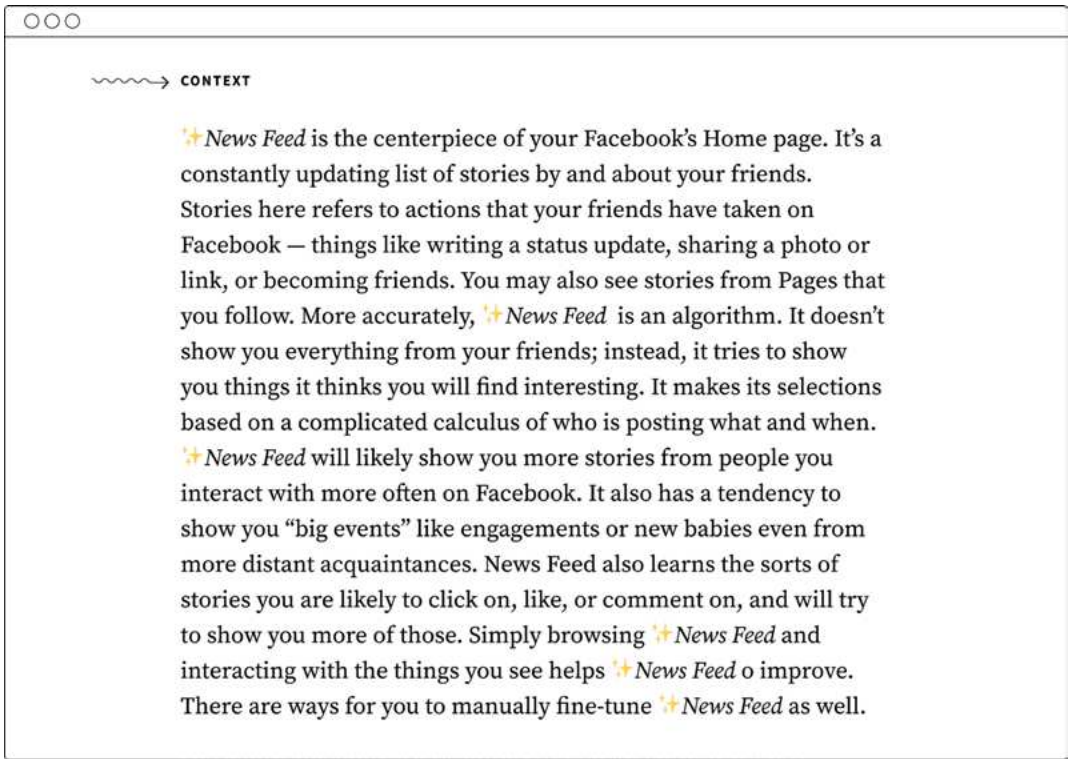
3.1. Case study: "Why am I Seeing this Infodemic on my Facebook News Feed"

The "Why am I seeing this *infodemic* in my Facebook's News feed?" experiment belongs to a collection of experiments developed during the data sprint. The experiment tries to answer the question "How does the Facebook algorithm facilitate the radicalization process of a newborn user through suggestions started from the query "*no-mask Italia*". As explained in the previous section, the format autogenerates this page, building it on top of the contents. "Why am I seeing this *infodemic* on my Facebook's News Feed?" focuses on (1) exploring the day-by-day development of a newborn Facebook account interested in no-mask related issues in Italy, (2) collecting data about newsfeed and user's activity manually, (3) visualizes and tells the radicalization process of the user.

Figure 7. General presentation of the experiment "Why am I seeing this infodemic on my Facebook's News Feed" proposed as case study for this paper. shows the general presentation of the experiment: the title (Figure 7. General presentation of the experiment "Why am I seeing this infodemic on my Facebook's News Feed" proposed as case study for this paper.a), the techniques and tools used to refer to the shared glossary (Figure 7. General presentation of the experiment "Why am I seeing this infodemic on my Facebook's News Feed" proposed as case study for this paper.b), the

brief description (Figure 7. General presentation of the experiment "Why am I seeing this infodemic on my Facebook's News Feed" proposed as case study for this paper.c), and the table of contents (Figure 7. General presentation of the experiment "Why am I seeing this infodemic on my Facebook's News Feed" proposed as case study for this paper.d), enabling a quicker navigation through the experiment.

Scrolling through the general presentation, as stated in the table of contents, the experiment contains the following blocks: the **context**, the **research question,** the **setup**, the **disclaimer**, the **visual index**, the research **steps,** and the **findings**. Among these, some are more relevant than others and deserve an in-depth description of their characteristics through the application of the case study under consideration. As emerged from the authors' experiences and focus groups' results, the main strengths of the format are the following:


CONTEXT

✨*News Feed* is the centerpiece of your Facebook's Home page. It's a constantly updating list of stories by and about your friends. Stories here refers to actions that your friends have taken on Facebook — things like writing a status update, sharing a photo or link, or becoming friends. You may also see stories from Pages that you follow. More accurately, ✨*News Feed* is an algorithm. It doesn't show you everything from your friends; instead, it tries to show you things it thinks you will find interesting. It makes its selections based on a complicated calculus of who is posting what and when. ✨*News Feed* will likely show you more stories from people you interact with more often on Facebook. It also has a tendency to show you "big events" like engagements or new babies even from more distant acquaintances. News Feed also learns the sorts of stories you are likely to click on, like, or comment on, and will try to show you more of those. Simply browsing ✨*News Feed* and interacting with the things you see helps ✨*News Feed* o improve. There are ways for you to manually fine-tune ✨*News Feed* as well.

Figure 8. Block describing the context, namely the research object and its features.

- It invites users to *contextualize the research object*, situating it in space and time.

- It helps in *structuring the experiment setup* suggesting the items to be considered.

- Its *visual index* summarizes the research steps in a diagram.

-It allows for combining different layouts, promoting the use of *parallel columns* for comparing data, processes, visualization, or findings.

-It allows to create *a glossary*, shared by all the experiments of the project.

Contextualizing the research object — The News Feed is the research object around which the experiment revolves. It is the centerpiece of Facebook's home page and is constantly updated with a list of stories by friends, advertisements, and pages

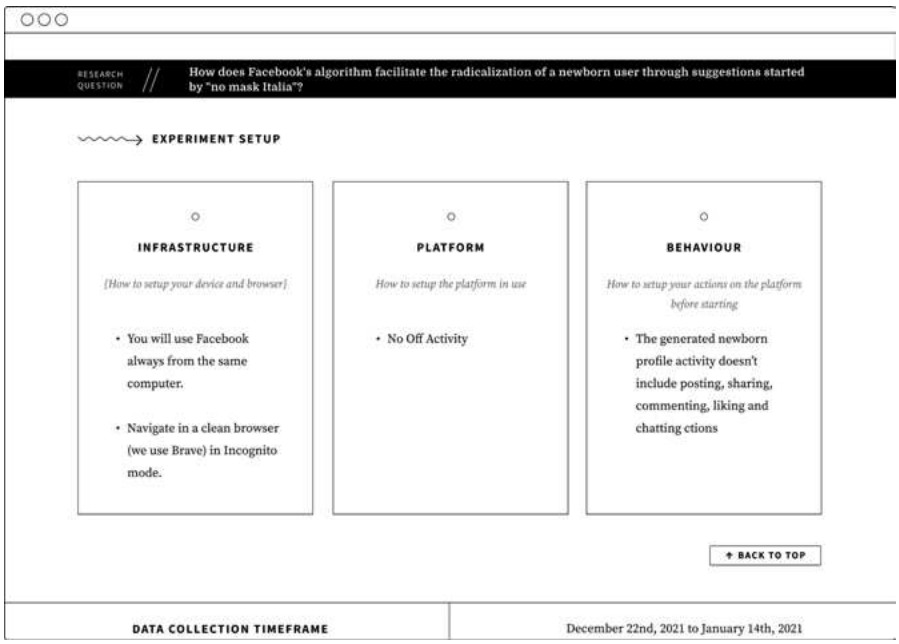


Figure 9. Experiment set-up building box in use. From left to right: "Infrastructure" is about describing how to set up the device and the browser. The "Platform" box is dedicated to settings related to the platform being analyzed. "Behaviour" defines which actions to perform within the platform.

updates. The variety of contents appearing on the News Feed is managed by a complex algorithm, which relies on many factors including user profiling, then users' choices and advertisements. At the time of the experiment (January 2021), given the General Data Protection Regulation (Regulation (EU) 2016/679 of the European Parliament and of the Council), users can also discover the reason why some contents such as suggested ADVs appear on their News Feed by clicking on the "Why Am I Seeing This" button. If in January 2021, the button was active just for mobile devices, one year later it is also active for desktops. Indeed, research objects are frequently evolving: they are revised and modified in terms of interface or privacy policy; hence it is important to keep track of their state of the art at the time of research.
Thus, the **context** block available in the format, which is usually plain text, guarantees an in-depth description of the research object, where users are invited to report its specificities at the time of the data sprint (Figure 8. Block describing the context, namely the research object and its features.).

Setting up the experiment — As emerged from previous discourse it is crucial to report the set-up of the experiment, namely what type of infrastructure and which devices have been used, on which digital platform the research has been done and what type of behavior was held by the researcher(s) carrying on the experiment. Indeed, defining the experiment setup is the first stage for making the procedure transparent and repeatable. In this case, regarding the infrastructure, participants decided to access Facebook always from the same device — a 15" MacBook Pro, and to navigate in a clean browser with the incognito mode on (Figure 9. Experiment set-up building box in use. From left to right: "Infrastructure" is about describing how to set up the device and the browser. The "Platform" box is dedicated to settings related to the platform being analyzed. "Behaviour" defines which actions to perform within the platform.).

Telling the process —Tracking and presenting the research process carried on during a data sprint is task-driven usually by designers (Mauri et al., 2019). According to Mauri et al., designing visual protocols is a common procedure for summarizing processes, making them accessible, transparent and repeatable, both during data sprints (Mauri et al., 2019) and as a didactic activity (Mauri et al., 2020). Drawing inspiration from

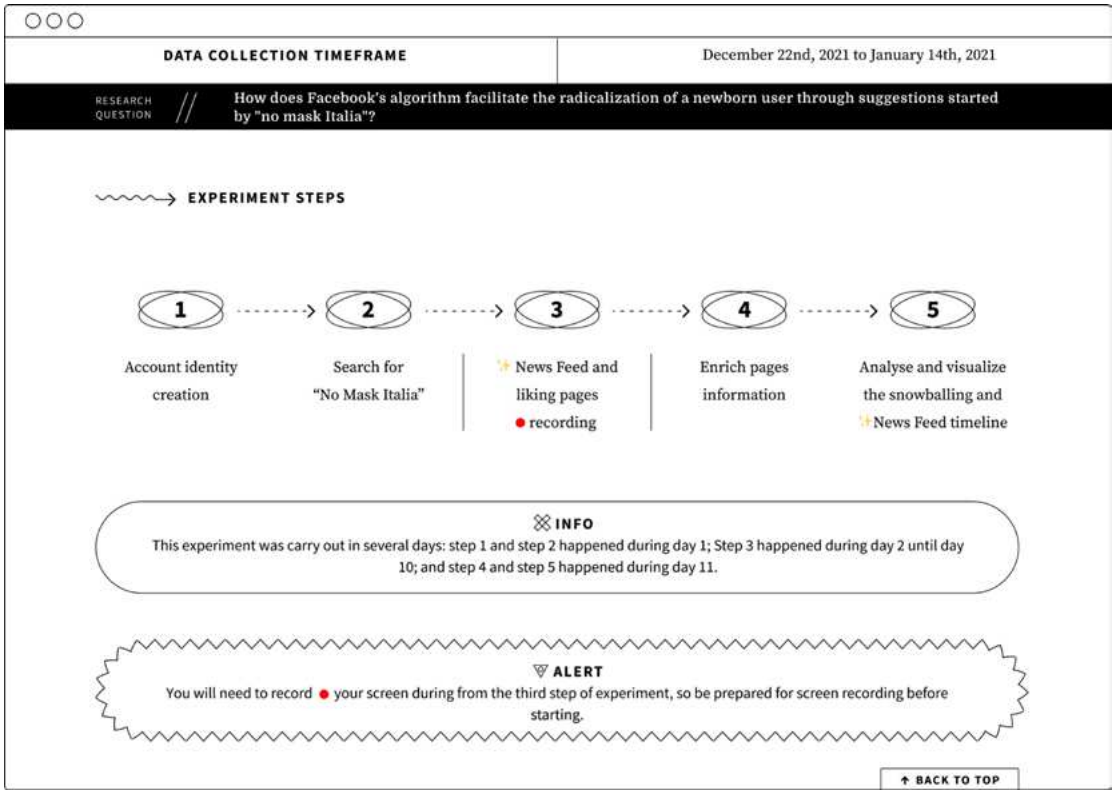


Figure 10. Experiment's visual index summarized in steps.

this tradition, the format allows users to self-generate a diagram working as a visual summary and index to quickly access the diverse steps of the research process. This feature not only makes the fruition simpler, but also helps to organize the experiment in a modular way, making it repeatable and transparent. Moreover, the format allows also to include supplementary information and disclaimers related to the process (Figure 10). As a first action, before starting to gather data, participants created a fake name and a new account on Facebook as depicted in Figure 11. Account creation for the fake Facebook user. The first fixed column on the left summarizes the step, while the double column layout allows to compare two actions, namely the fake name generation and the actual account creation.. During the two weeks before the data sprint, News Feed and user activity data were collected every two days employing techniques such as *screen recording* and *snowballing*. Through *screen recording* researchers stored a visual log of the actions done by the newborn user when collecting data appearing on the News Feed. *Snowballing*, namely the procedure whereby the researchers cascade likes or follows to what is suggested to them by the algorithm, was used to simulate a user activity based on the suggestions of the platform. Hence, given the long data collection process, a large part of the experiment was dedicated to constructing and enriching the

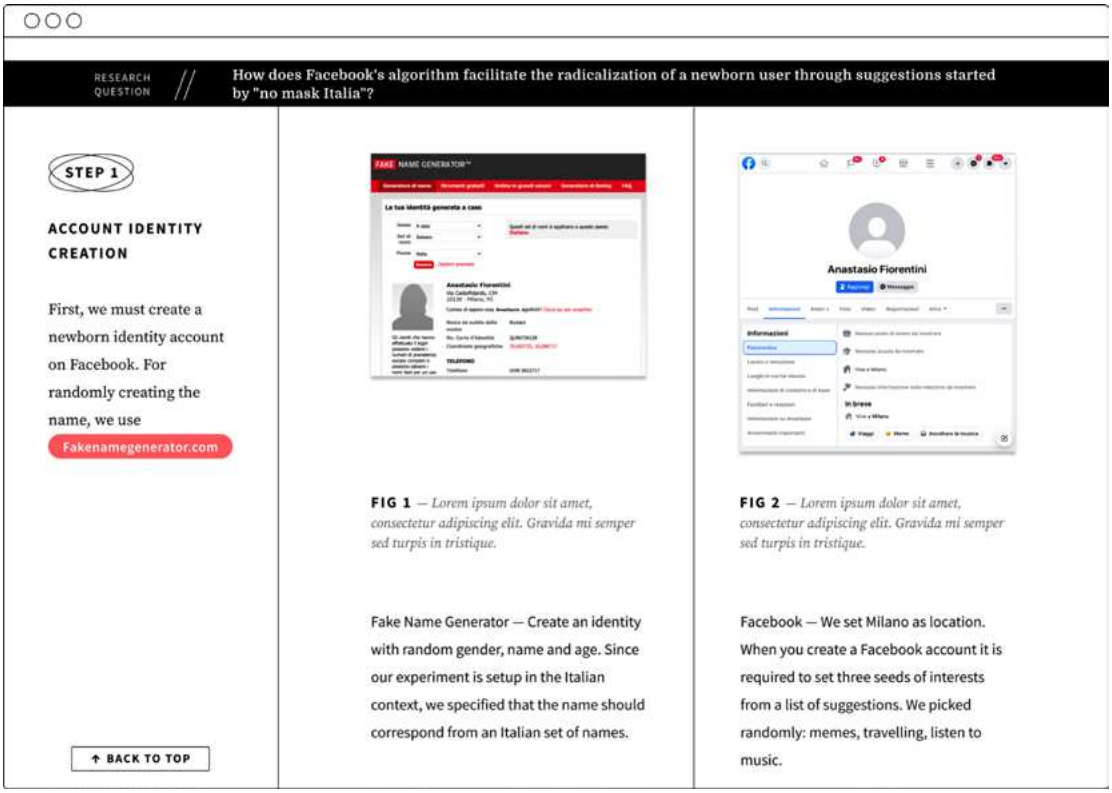


Figure 11. Account creation for the fake Facebook user. The first fixed column on the left summarizes the step, while the double column layout allows to compare two actions, namely the fake name generation and the actual account creation.

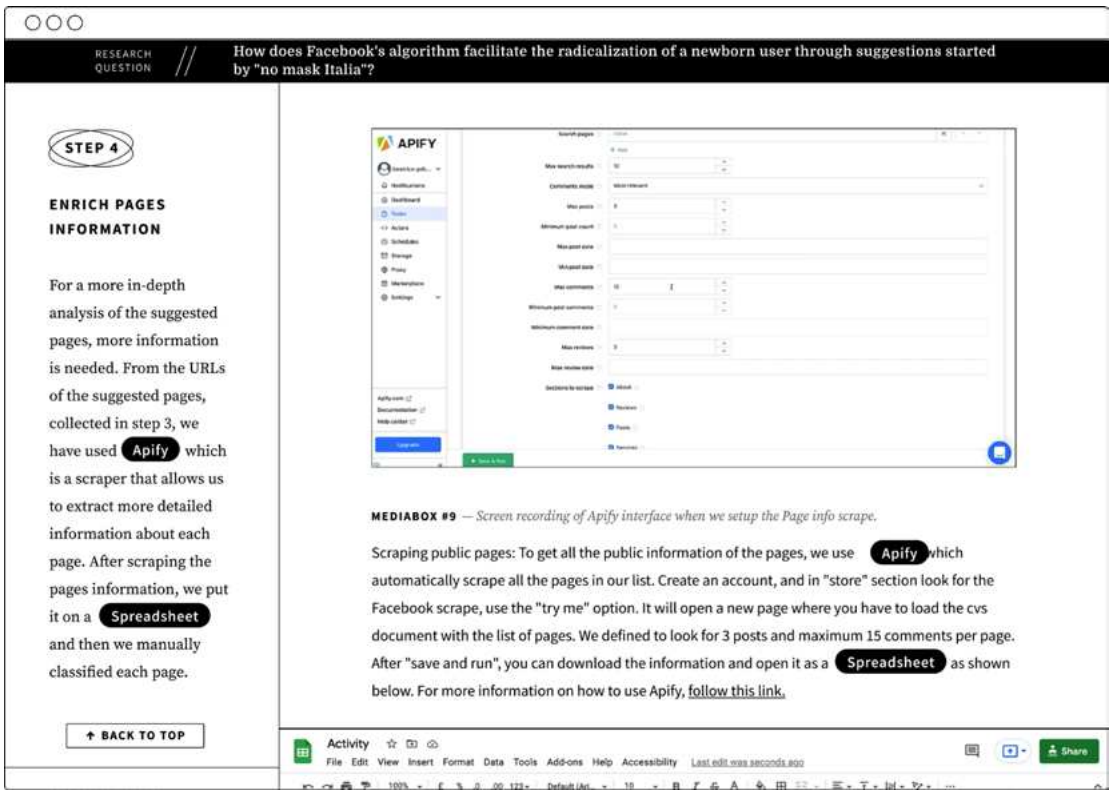


Figure 12. The column on the left summarizes the step, while the central column comprises two columns (top), or a single column with an animated GIFs and a caption.

main dataset. Indeed, to present how a new user interested in no-mask-related issues in Italy radicalizes on Facebook, researchers first collected the list of items appearing on the screen by scrolling through the News Feed. At the same time, they liked the suggested pages and, finally, they enriched the list of liked pages by crossing information coming from pages' descriptions. During the two weeks before the data sprint, News Feed and user activity data was collected every two days employing techniques such as *screen recording* and *snowballing*. Through *screen recording* researchers stored a visual log of the actions done by the newborn user when collecting data appearing on the News Feed. *Snowballing*, namely the procedure whereby the researchers cascade likes or follows to what is suggested to them by the algorithm, was used to simulate a user activity based on the suggestions of the platform.

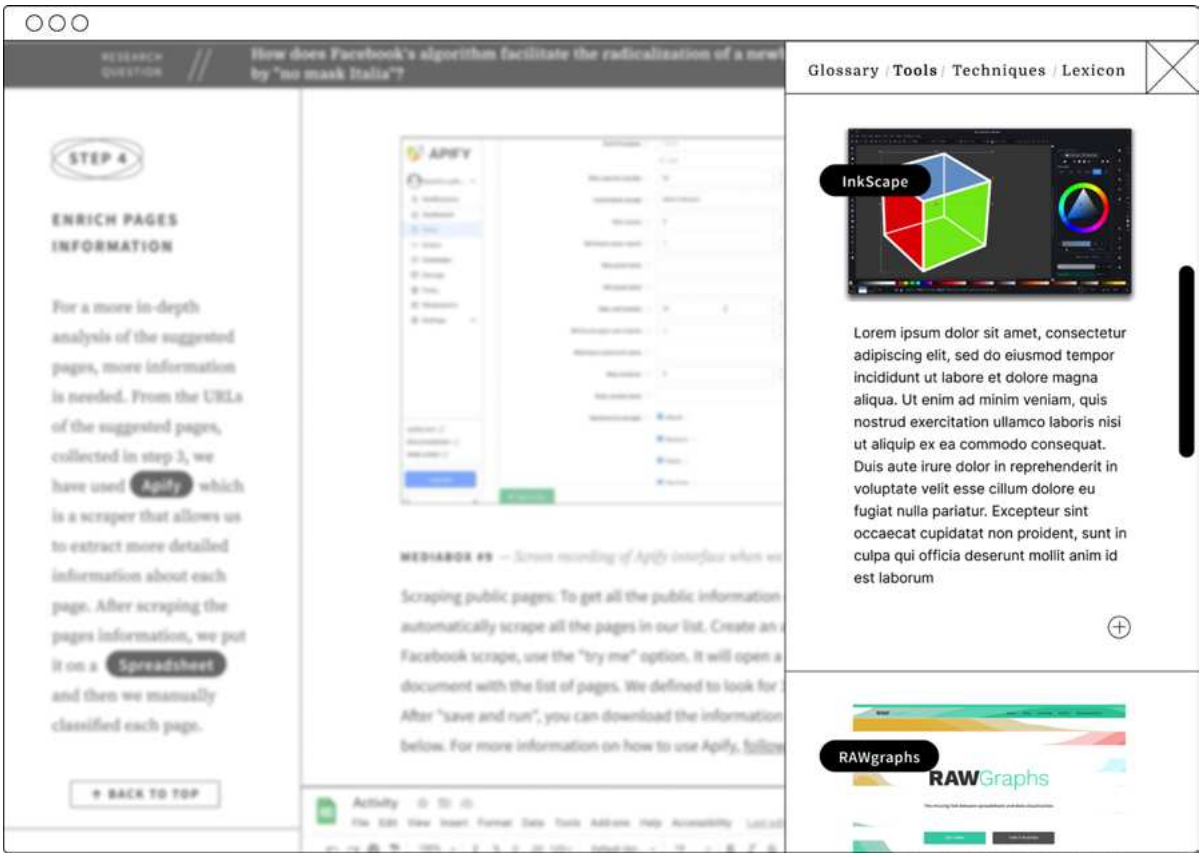


Figure 13. Screen of the Glossary which can appear both as a sliding panel or as an independent page, as depicted previously in Figure 5.

Comparing actions — When dealing with Digital Methods in data sprint-based research it is common to find experiments where some actions are compared in pairs: actions related to the same experiment step, the behavior of different social network accounts or the results emerging from two correlated hashtags —such as "program" and "anti-program" (Rogers, 2017, p. 82). To meet this need, the format allows users to populate a two-column layout where contents can be simultaneously read and compared. (Figure 11). Moreover, the format includes also a single-column layout, that users picked in this context for showing animated GIFs of a visualization, for embedding external video resources and for writing texts (Figure 12. The column on the left summarizes the step, while the central column comprises two columns (top), or a single column with an animated GIFs and a caption.).

Hence, given the long data collection process, a large part of the experiment was dedicated to constructing and enriching the main dataset. Indeed, to present how a new user interested in no-mask-related issues in Italy radicalizes on Facebook, researchers first collected the list of items appearing on the screen by scrolling the News Feed. At the same time, they liked the suggested pages and, finally, they enriched the list of liked pages by crossing information coming from pages' descriptions.

Generating a glossary — As emerged from authors' experience and from past data sprints documentation, it happens that experiments share tools, techniques, and lexicon. For instance, the previously described *Snowballing* technique has been used in other experiments as well as almost all the participants relied on RAWGraphs for designing visualizations.

Thus, the format integrates a feature for building a glossary and referring to its items throughout the whole experiment (Figure 5 and Figure 13). In this case, the shared glossary allowed users to self-create entries, take advantage of those integrated by participants who have worked on other experiments, and access items to be informed and inspired.

4. Conclusions: A flexible and thorough format for disseminating data sprint results

The paper describes the peculiarities of a new dissemination format by design practitioners for Digital Methods data sprints results. As described in Section 1, the necessity to design the format emerges from the authors' experience in the field. Indeed, the format condenses and reinterprets good practices already in use in current formats for disseminating data sprints. Authors acknowledge that the format presented in the paper requires users to try outside their usual way of collecting, compiling, and presenting the process and results. As stated before, contextualizing the *research object,* and structuring the *experiment setup* requires additional research and readings. Describing the *steps* of the research process forces users to include succinct text and accompanying images. Moreover, it is necessary for users to become familiar with the concept of building blocks for filling it. Working with different *layouts* and ordering blocks requires a clear perception of the results that are being presented. However, as emerged from the focus group, the stress test and the application case, the format gives users room to be thorough, and it provides complete information while still maintaining flexibility. The suggestion to contextualize the research object situates the experiments in space and time. Providing information about the characteristics of the research object at the time the experiment was done becomes significant when imagining researchers reading the experiment in a couple of years. Thanks to the possibility of adding datasets, tables, and animated images, the research process has more space in the format. Explaining the steps using screenshots simulating the actual use of tools not only gives researchers the freedom to describe the research as thoroughly as possible but helps future researchers repeat the process. Creating a glossary characterizes the project, enriching it with specific terms and techniques which can gain the attention of even readers who never used some of the mentioned techniques.

To be remembered, brought back to life and re-interpreted, knowledge needs to be dismantled, re-structured and stored in a *place* (Anceschi, 1992, p. 99). Similarly, the format presented in the paper is a digital *place* that allows data sprints' results to be stored, staged, and accessed after the research activities flexibly and thoroughly. Indeed, the format is designed by versed practitioners who condensed all their knowledge and experience to make future dissemination of data sprints more accessible. The format, available as an open-source Content Management System (CMS), will generate web pages to publish the research results online[9]. Even if the case study presented is framed within the context of a data sprint inquiring the public perception of algorithms on social networks, the authors foresee that the usage scenario of the dissemination format could be extended to others DM data sprints and advanced research carried on with DM. Thus, the next steps will be dedicated to testing with (1) participants of another data sprint and (2) students carrying on DM research activities, (3) DM researchers outside data sprints and (4) researchers not familiar with DM.

Acknowledgements

The paper is one of the results of the *Algocount* project, a collaboration between the Department of Political Sciences (Università degli Studi di Milano) and the Density Design Lab (Politecnico di Milano, Design Department). The *Algocount* project has been founded by Fondazione Cariplo.

Authors would like to thank Alessandro Gandini, the scientific coordinator of the project and the researchers from the Political Sciences department involved: Alessandro Gerosa, Luca Giuffrè, Diletta Huyskes, Silvia Keeling and Urbano Reviglio.

Besides, authors would like to thank participants of the described data sprint: Antonella Autuori, Irene Avossa, Laura Bruschi, Ludovica Corponi, Alessandra Facchin, Emma Fortunati, Giulia Giorgi, Michele Grilli, Giovanni Lombardi, Vincenzo Luise, Andrea Elena Febres Medina, Riccardo Pronzato, Camilla Volpe.

References

Anceschi, G. (1992). *L'oggetto della raffigurazione*. Etaslibri.

Bounegru, L., Gray, J., Venturini, T., Mauri, M. (2018). *A Field Guide to "Fake News" and Other Information Disorders*. Zenodo.

Ciuccarelli, P., Elli, T. (2019). Beyond visualization. In Hotson, H., Wallnig, T. (Eds.), *Reassembling the Republic of Letters in the Digital Age: Standards, Systems, Scholarship* (pp. 299–314). Göttingen University Press, Göttingen.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02016R0679-20160504&from=EN

Lupo, E., Gobbo, B., Lonardo, E. (2021). Towards a new design culture of scientific production – Innovating the formats of scientific publication of design. In *Design*

9. https://github.com/densitydesign/placplac

Culture(s). Cumulus Conference Proceedings Roma (pp. 1082–1097). Volume #2. .

Mauri, M., Briones, Á., Gobbo, B., Colombo, G. (2020). Research Protocol Diagrams as Didactic Tools to Act Critically in Dataset Design Processes. In *INTED2020. Proceedings of the 14th International Technology, Education and Development Conference* (pp. 9034–9043). https://re.public.polimi.it/retrieve/handle/11311/1132634/498643/2470.pdf

Mauri, M., Ciuccarelli, P. (2016). Designing diagrams for social Issues. Presented at the *Design Research Society Conference 2016*. https://www.drs2016.org/185

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., Azzi, M. (2017). RAWGraphs: A Visualisation Platform to Create Open Outputs. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*. ACM, Cagliari. https://dl.acm.org/doi/10.1145/3125571.3125585

Mauri, M., Gobbo, B., Colombo, G. (2019). O Papel do Designer no Contexto do Data Sprint. In Omena, J. J. (Ed.). *Métodos Digitais. Teoria-prática-crítica*. Livros ICNOVA.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy sciences, 4*(2), 155-169.

Rogers, R. (2013). *Digital methods*. The MIT Press, Cambridge, Massachusetts.

Rogers, R. (2017). Foundations of Digital Methods: Query Design. In M. T. Schäfer, & K. van Es (Eds.). The datafied society: Studying culture through data (pp. 75–94). Amsterdam: Amsterdam University Press.

Venturini, T. (2014). Climaps by Emaps in 2 Pages. *SSRN Journal*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2532946

Ziewitz, M. (2016). Governing Algorithms: Myth, Mess, and Methods. *Science, Technology, & Human Values, 41*(1), 3–16.