



What are data sprints for?

¿Para qué sirven los data sprints?

ID Janna Joceli Omena

Janna.Joceli.Omena@warwick.ac.uk
Centre for Interdisciplinary Methodologies,
University of Warwick
ICNOVA, iNOVA Media Lab; Public Data Lab

ID Lorena Cano-Orón

Lorena.Cano@uv.es
Universitat de València

ID Beatrice Gobbo

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

ID Ana Marta M. Flores

amflores@fcsh.unl.pt
ICNOVA, Universidade NOVA de Lisboa

ABSTRACT

The data sprint approach enables various objectives for the digital methods community, from fostering interdisciplinary collaboration to providing teaching-learning spaces regarding creative methods. However, data sprints' purpose, advantages and concrete results are still little known across disciplines. Thus, this paper presents four facets *pertaining to* and *deriving from* data sprints to explain their prospects for different (non-) academic communities. First, we define the data sprint approach, providing a detailed description of what data sprints are and what they involve in practice and, in turn, propose guidance to facilitate the replicability of this work method. Second, we elucidate how data sprints are (1) *a means of teaching and learning digital methods research*, arguing that the data sprint environment is not only (2) *a space for methods and tools creation* but also (3) *a reflective tool* to understand the triad of data-, software- and platform-oriented research (from the standpoint of practice). Therefore, data sprints offer researchers situational ways to access and critique scientific knowledge production. Finally, we address a standard post-sprint procedure, (4) *the reutilisation of data sprint reports for producing scientific knowledge*, through academic and non-academic publications, as an established research practice. The four facets unpack the data sprint approach for a broader audience whilst indicating the possible takeaways during and after such events. We conclude with reflections from and on data sprints.

KEYWORDS

data sprint, digital methods, science production, data analysis, data visualisation, interdisciplinary research

How to cite this article:

Omena, J. J.; Cano-Orón, L.; Gobbo, B. & Flores, A. M. (2022). What are data sprints for?. *Dígitos. Revista de Comunicación Digital*, 8: 9-30. DOI: 10.7203/drdcd.v1i8.253

RESUMEN

El método *data sprint* permite varios objetivos para la comunidad de métodos digitales, desde el fomento de la colaboración interdisciplinaria hasta la provisión de espacios de enseñanza-aprendizaje en relación con los diseños metodológicos creativos. Sin embargo, el propósito, las ventajas y los resultados concretos de los *data sprints* son todavía poco conocidos en todas las disciplinas. Este artículo presenta cuatro facetas que pertenecen y se derivan de esta metodología de trabajo para explicar sus perspectivas para diferentes comunidades (no) académicas. En primer lugar, definimos el método de los *data sprints*, proporcionando una descripción detallada de lo que son y lo que implican en la práctica y, a su vez, proponemos una guía para facilitar la replicabilidad de este método de trabajo. En segundo lugar, exponemos cómo los *data sprints* son: (1) un medio de enseñanza y aprendizaje de la investigación de métodos digitales, argumentando que el entorno de los *data sprints* no es sólo (2) un espacio para la creación de métodos y herramientas, sino también (3) una herramienta de reflexión para entender la tríada de la investigación orientada a los datos, el software y la plataforma, desde el punto de vista de la práctica. Por lo tanto, los *data sprints* ofrecen a los investigadores formas delimitadas de acceder y criticar la producción de conocimiento científico. Finalmente, abordamos un procedimiento estándar posterior al *sprint*: (4) la reutilización de los informes del *data sprint* para producir conocimiento científico, a través de publicaciones académicas y no académicas, como una práctica de investigación establecida. Estas cuatro dimensiones desvelan el enfoque del *data sprint* para un público más amplio, a la vez que indican las posibles conclusiones durante y después de tales eventos. Concluimos con reflexiones procedentes de los *data sprints* y sobre ellos.

PALABRAS CLAVE

data sprint, métodos digitales, producción científica, análisis de datos, visualización de datos, investigación interdisciplinaria



What are data sprints for?

1. Introduction

Data sprints are a well-known, conventional approach within the digital methods community. They can be part of master's and Ph.D. programs whilst also being available to a broader audience or dedicated to specific working groups and research projects. As a burst and intensive week of work with online data and methods, a range of tools and creative visualisations, the data sprint approach has made a name for itself and achieved status over the past decade. Its interdisciplinary profile plays a crucial role in this. It is no wonder that participants and collaborators from different disciplines co-operate in data sprint projects and workflows, contributing to achieving a common end, from answering specific research questions to experimenting with new methods or building research tools.

The data sprint approach has also inspired scholars, researchers and students to conduct research *with* and *about* digital platforms, online data and software. A data sprint project, for instance, can mobilise disciplines such as communication, design, sociology, history and computer science for the simultaneous study of climate change issues on social media and algorithmic recommendation systems. However, the

purpose, advantages and concrete results of data sprints are still little known across fields of study. So, what purposes does the data sprint approach serve? What does it bring to the scientific community? In addition, how does the data sprint approach work in practice? What is expected from all participants involved? This paper focuses on unpacking what data sprints are for, being organised as follows: First, we define the data sprint approach, providing a detailed description of what data sprints are and what they involve in practice; in turn, we propose guidance to facilitate the replicability of this working method. Subsequently, we present the four facets that *pertain to* and *derive from* data sprints, namely an understanding of data sprints as: (1) a means of teaching and learning digital methods research, (2) a space for methods and tools creation, (3) a reflective tool and (4) a means for producing scientific knowledge through interdisciplinary collaborations. By so doing, we argue that data sprints are more than an interdisciplinary collaborative environment; they are also a pedagogic and scientific tool. This reflection arises from the experience of the coordinators of the *Dígitos* monograph section 'The data sprint approach for research: experiments, protocols and knowledge' concerning the participation in, organisation of and collaboration within data sprints; and also from the experience of the authors and their work included in this special issue.

2. Data sprint: definition, characteristics and format

A data sprint is an interdisciplinary research practice that has been gaining ground in different fields and growing over the years. Unlike hackathon proposals, which bring programmers and other stakeholders together to build software programs or codes, data sprints focus on social, cultural and media projects. Although there are multiple variations related to diverse interpretations and fields, some common elements can be identified. Data sprints are time-limited and project-oriented, and they encourage co-participation amongst participants of different expertise and career levels. Research projects are developed within a period typically spanning 5 days and carried out in person, online or in hybrid modes. These projects often rely on data from the web and its technological environment to answer specific research questions. Data sprints may have different goals, but they are usually appropriate for tackling so-called ill-defined problems, namely situations where it is challenging to define an initial design brief and that require multi-disciplinary effort and expert knowledge to address the issue (Mauri *et al.*, 2019).

Data sprints have been implemented to foster and promote collaborative, interdisciplinary and experimental workflows. In this environment, participants with different backgrounds engage on specific projects whilst 'work(ing) together on a set of data and research questions' (Venturini *et al.*, 2018: 1). By providing a space for intensive collaborative workflows, tool creation, coding and methodological workshops, exploratory data analyses and visualisations, data sprints reflect the learning-by-doing principle. They thus require practical work and some technical knowledge in method implementation and data visualisation. With many people participating in real-time data analysis, data sprint activities are far from a segmented and isolated mode of

work. The richness of this approach lies in its interdisciplinary nature, since it enables ‘social scientists, developers and data designers together with relevant domain experts to explore research questions and create prototype digital methods projects’ (Munk, Madsen, & Jacomy, 2019: 110).

However, not all is rosy in data sprints, as Venturini and Munk (2022) warn when they define data sprint research as “a quick and dirty” approach. This is because of the intensity and speed with which teams shape research and test methods and generate results. Configured as an open-ended project, research in data sprints is shaped collectively by team members as data become available and analyses are carried out. This is why Rogers and Lewthwaite (2019: 19) state that data sprint projects operate “on the fly”. For Mauri and Ciuccarrelli (2016), data sprints often take the form of “unfinished products”. Indeed, at the end of the data sprint week, many of them will be discarded, but thanks to them, it will have been possible to open up new hypotheses and make progress while also answering research questions.

3. Understanding the Data Sprint Approach

We now move on to what happens before, during and after a data sprint to describe the characteristics and formats of a data sprint, as well as to explain what it is. We unpack how the data sprint approach works in practice, focusing on its different phases of preparation, realisation and results.

3.1. Before | Data sprint preparation: a checklist

Table 1 presents a data sprint preparation checklist according to the tasks assigned to organisers, collaborators and participants. This helps indicate how different activities are expected from all those involved in a data sprint, yet drawing particular attention to the role of participants.

Data sprint organisers secure funding from both inside and outside the institution they are a part of, choose the conceptual and thematic proposal, invite keynote speakers and collaborators (including those responsible for project pitches) and plan the data sprint’s programme. Organisers should consider, or try to balance the impact of, the proposed programme on both the project’s development and the participants’ learning expectations. The event’s logistics and communication management are also the responsibility of the organisers, from booking rooms or creating web-conferencing links to communicating with collaborators and attendees. Organisers coordinate activities not only before but also during and after the data sprint.

Data sprint collaborators (or facilitators) are crucial players in making the sprint happen owing to their specific field or discipline expertise, and are generally researchers, media scholars, issue experts or organisations willing to pitch a project or join a project. Also, early-career or senior professionals, such as developers, designers (also design students in training), data analysts and digital methods researchers, enable hands-on workshops and join a project.

Data sprint	Organisers	Collaborators	Participants
Checklist	<ul style="list-style-type: none"> ✓ Secure funding ✓ Define theme and practical labs ✓ Invite keynote speakers and collaborators ✓ Plan data sprint schedule ✓ Manage communication and logistics ✓ Provide a list of (video) tutorials, worksheets or recipes, as well as suggested readings for participants ✓ Coordinate activities not only before but also during and after the data sprint 	<p>[regarding hands-on workshops]</p> <ul style="list-style-type: none"> ✓ Practical labs preparation and presentation (face-to-face or pre-recorded section) ✓ Prepare method recipes and shared data folder for practical exercises <p>[regarding project management]</p> <ul style="list-style-type: none"> ✓ Provide a project description ✓ Create a project folder and datasheets for datasets ✓ Collect data and treat the dataset (if required by the project) 	<ul style="list-style-type: none"> ✓ Download and install tools and software suggested by organisers and required for the practical labs sections ✓ Watch suggested tutorials (or at least those of personal interest) ✓ Spend some time following up on the recommended tutorials and tools or codes ✓ Read suggested readings ✓ Engage with project work during and after the data sprint ✓ Collect data if required by the project

Table 1. Data sprint preparation checklist (for organisers, collaborators and participants)

Collaborators giving practical labs should prepare hands-on material to help participants use software or web-based tools, such as the Memespector Graphical User Interface (GUI) (Chao, 2021). Tutorials assist participants in making sense of digital records (i.e., hashtags, URLs and image metadata) and methodological approaches while using worksheets or method recipes for research (i.e., Public Data Lab and SMART Data Sprint recipes¹). Additionally, collaborators are also responsible for teaching specific data collection, analysis and visualisation techniques and how to make sense of innovative textual and visual methodologies. *Collaborators managing data sprint projects* have an even longer list of responsibilities. They should provide a project description containing, for example, the project title, a brief introduction to contextualise the project, research questions, what the methods, datasets and tools are, preliminary findings (if any) and references. This is where ethical considerations may arise, remaining relevant until the dissemination of the results of the data sprint. In addition, it is often recommended that data collection be done before the data sprint, leaving more space for exploratory and analytical tasks during the sprint. At this point, a shared folder for the project is created and organised with subfolders such as datasets, resources, analysis and visualisations (Figure 1). The organisation of the project folder guides how and where participants should enter their collaboration whilst providing common access to the results of the team’s work.

1 <https://recipes.publicdatalab.org/>, <https://smart.inovamedialab.org/digital-methods-recipes/>

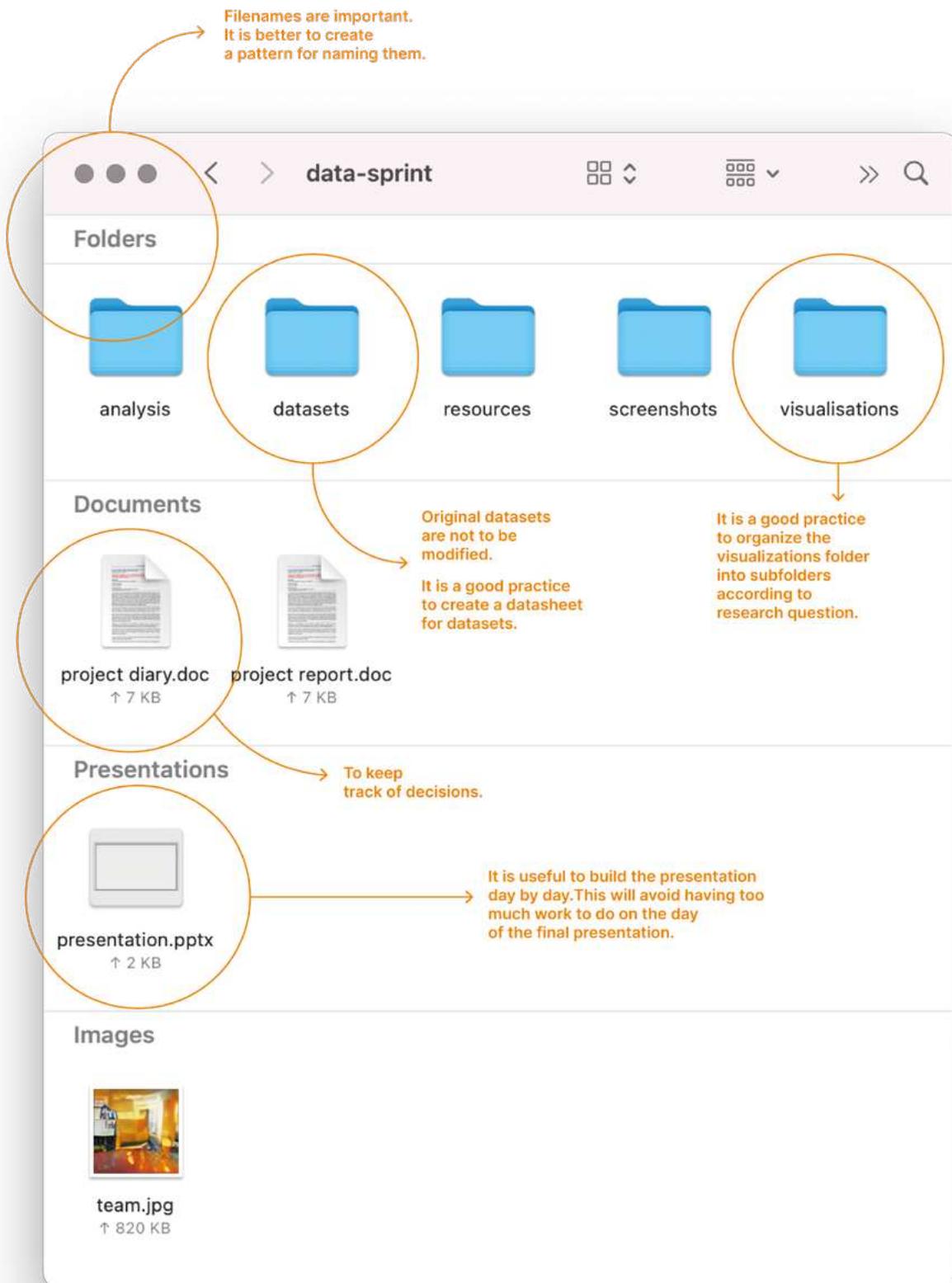


Figure 1. Sample folder for organising data sprint materials.

Project leaders are also expected to provide and look after good documentation for the dataset building (i.e., using dataset folders and datasheets for datasets) and project analysis and findings (i.e., using project diaries or shared files for analysis). Documenting practices before and during the sprint is crucial for facilitating data sprint workflows. For example, datasheets for datasets can document decisions concerning query design

Platform	Entry point(s)	Date— Data collection & extraction software	Parameters for data collection	Who collected the data?	Folder	Total number of images	Visualisation Folder	Computer Vision API & features in use	Memespector	Computer vision outputs
Facebook	"Covid Meme" Query	First week of January 2022 // Crowdtangle	(Not Meme Search) Pages and Groups Photos Only English Only January 1, 2020-January 3, 2022 Sorted by Total Interactions First 1000 Out of 8565		[link]	1005 images	[link]	Google Vision: web entities & web entities (text?)		[link]

Figure 2. Example of a *Dataset-building sheet* organised with participants one week before the 2022 Digital Methods Winter School. Project: What is a meme, technically speaking?, available at <https://wiki.digitalmethods.net/Dmi/WinterSchool2022WhatIsAMeme>

Platform/Case study	Folder-of-images	Total number of images	GCV API existing results from	Any computer vision output files from 2021?	CSV file with images annotated by GCV API	Subject of study	Descriptive info	Digital record	Entry point to data collection	Month/year of data collection & extraction software	Previous work
Tumblr - EleNao vs. EleSim	https://drive.google.com/drive/folders/1x-u5Yzrz8hpON5sB8tk0NHHY4FuW2fw?usp=sharing	567 images	October, 2018	No (also run web detection)	In this doc: annotated_ele nao_elesim_20181003-2040	Women leading protests against Bolsonaro before 2018 Brazilian presidential elections	Images of people protesting, selfies, posters/banners, food	hashtags	#EleSim, #EleNao	October, 2018 (Tumblr Tool)	https://thesocialplatforms.wordpress.com/2018/10/22/elenao-vs-elesim/
Facebook - Portuguese Universities	https://drive.google.com/drive/folders/1pB0Ch11UJNDQOU3SYnmN7HC3YkwkEEy?usp=sharing	22,594 images	May, 2018	No	In this doc: annotated_universidades_20180608-1603 or here: https://drive.google.com/file/d/1gk8bGuixTcnw76ZoQcK-cnGhVNE5jD_Q/view?usp=sharing	The imagery of Portuguese Universities on Facebook from 2009 to 2018. Timeline images of 15 official pages of Portuguese Higher education.	People attending academic events, sports, musical performance, buildings	image URLs from FB timeline publications	Page id	Between 2017 and 2018 (Netvizz)	https://icono14.net/ojs/index.php/icono14/article/view/1436
Instagram - Microcephaly	https://drive.google.com/drive/folders/1BbkcPCo5M0_LViWt5AhZSRcVgGfE7UJ7F?usp=sharing	10,797 images	February, 2018	Yes: June, 2021	https://drive.google.com/file/d/1iqz347c9a82fNiGjKEtjhUjaMob1gB/view?usp=sharing	The visualities associated with microcephaly in Brazil, how Brazilians families impacted by the virus shared their concerns and everyday life.	Images with infants with microcephaly accompanied by the mother or family, banners, events, spokespersons, food	hashtags	#microcefalia	Between June 2012 and October 2017 (Visual Tagnet Explorer)	Not published, but used as a reference in some presentations. Currently is part of our Disena paper.

Figure 3. Example of a *datasheet for datasets* organised weeks before the software-testing and cross-computer vision enquiry sprint at the Centre for Advanced Internet Studies in May 2021. Data sprint for testing the new version of Memespector Graphical User Interface (GUI) (Chao, 2021) and interrogating multiple computer vision APIs. Research software available at <https://github.com/jason-chao/memespector-gui>

strategies and data collection parameters, while serving as reminders of teamwork and decisions made before the subject of analysis. Figures 2 and 3 illustrate datasheets for datasets and the actions involving dataset building for studying meme collections (Figure 2), such as the choice of the platform, the query design strategy, entry points and parameters for data collection. Links to the original data folder (or files) and data processing go in this sheet. The different datasheets also show the documentation of datasets for software testing and cross-computer vision inquiry (Figure 3).

Data sprint participants extend across a wide range of academic and non-academic career levels, from master’s or Ph.D. students and early-career scholars to senior researchers and professors. Artists, journalists or data analysts are other examples of possible audience members. Data sprints demand a different level of self-sufficiency and anticipatory attitudes from their participants, unlike other academic events such

as conferences and symposiums. Therefore, the preparation of participants starts by committing to the instructions given by the data sprint organisers. Before the sprint, participants are often challenged to download and install research software and web-based tools, and spend some time following up on the recommended video tutorials and engaging with a list of suggested readings. Data sprint preparation for participants draws on the flipped learning approach (see Milman, 2012; Zuber, 2016). Students acquire knowledge before the class, using classroom time to practice and deepen understanding through interaction with peers. By following the preparatory instructions, participants can bypass the more overwhelming experience found in data sprint contexts, precisely because they have done some training in advance.

3.2. *During* | Data sprint realisation: hands-on tutorials, group work and final presentation

Keynote talks are often the opening activities of a data sprint, addressing the sprint theme and making room for reflections and discussions. However, the core activities are developed on the basis of the dynamic factors experienced by each participant. A data sprint implementation is composed of at least three phases: methods training and practising (hands-on tutorials), project development (group work) and presentation of the results (final presentation), as explained in detail below.

→ Phase 1 | Practical labs: training and practising methods

Firstly, there is the training phase. This is developed through attendance in the practical labs (or 'hands-on' tutorials) which explain the use of certain tools or the specific development of certain methods (Image 1). This is always useful, practical knowledge that is directly applicable to the participants' projects. Usually, the training is offered in parallel sessions so that a range of options can feed into different participant interests



Image 1. Practical lab on visualising data with RawGraphs during the 2019 SMART Data Sprint at NOVA University Lisbon, Portugal. Image source: https://smart.inovamedialab.org/wp-content/uploads/2019/03/IMG_0518.jpg

and project natures. It is here that data sprint collaborators share knowledge expertise, step-by-step protocols and other valuable teaching resources.

These meetings offer training that goes beyond formal education, utilising the methods of “learning by doing” and “learning from the community” methods. During data sprints, newbies and beginners acquire new skills and explore methods and tools that are not very familiar to them. This means stepping out of one’s comfort zone and opening one’s mind to new ways of approaching research in the digital environment.

➔ Phase 2 | Group work: project development

Secondly, there is the phase of project development on the basis of group work sessions, which is the longest part of the sprint, laying the foundation for the initiative. Data sprint projects are composed of the person or people responsible for the pitches, who have proposed and conceived the main objectives and methods of the project; the facilitators, who are the experts that collaborate and help participants to achieve the goals of the project (they can be communication designers, developers or methodological experts); and, of course, the participants, who are the working muscle of the project and who will shape the initial research over the course of the sprint. Participants are divided into projects according to their affinities for the topic or the methods they plan to use as input for each project. Once participants join a project, they all work together to achieve their goals.

These intense sessions usually last an entire working day. Participants work together on analyses and visualisations of research data. Here, participants and collaborators feed the project folder in a coordinated and organised way, keeping the original datasets intact whilst they work on copy versions. They discuss the methods used, try out different techniques and jointly interpret the results obtained, accounting for ethical considerations. During the group work sections, many types of output are produced: spreadsheets, tools, image collections, visualisations and (gif) animations, amongst others. This collaborative work contributes to the interdisciplinarity and richness of such study by allowing multiple points of view to contribute to the construction of the research.

The project members are encouraged to develop a draft report on a daily basis during the sprint, using for example a shared research diary or analysis files. This will make it much easier to return to the writing of the research later, when memory starts to fail regarding how and why certain decisions were made during the research process. In addition, as the research evolves over time, the report’s narrative changes; that is, new research questions and means of data analysis appear, adapting to the project’s evolution and the group work’s daily discussion. This is why writing up data sprint reports should be done with an open mind.



Image 2. Group working during the 2020 SMART Data Sprint at NOVA University Lisbon, Portugal. Image source: <https://smart.inovamedialab.org/editions/2020-digital-methods/photo-gallery/>

→ Phase 3 | Presentation of the results

Third, there is the presentation of the results to the entire forum of participants, facilitators, organisers and people responsible for pitches of that edition of the data sprint. This communication may be experienced with some anxiety. No matter how well organised the work is, there are always things to improve at the last minute. Nevertheless, this time constraint leads the projects' teams to produce clear and understandable conclusions for the audience, even if the results haven't met the initial goals. This exercise is fundamental after the hard work of the previous days so that the participants can give meaning to and create a narrative surrounding what they have developed.

Project groups usually present their research process, research questions and findings using slides, which are focused on highlighting the research path and describing the main findings and pitfalls, as well as future research opportunities. Here, researchers can insert hyperlinks to resources such as datasets and high-quality visualisations. At this stage, communication designers are key players as they present visualisations and support the organisation of the content. The presentation is a crucial moment when researchers have the opportunity to receive feedback and input for developing the project in the future.

At the end of the week, through formal, practical and collaborative learning, the acquisition of specific competences is achieved. In addition, engaging in data sprint projects not only offers ways of keeping up to date with the methods in use by the related research community but also stimulates future opportunities for collaboration and professional endeavours.

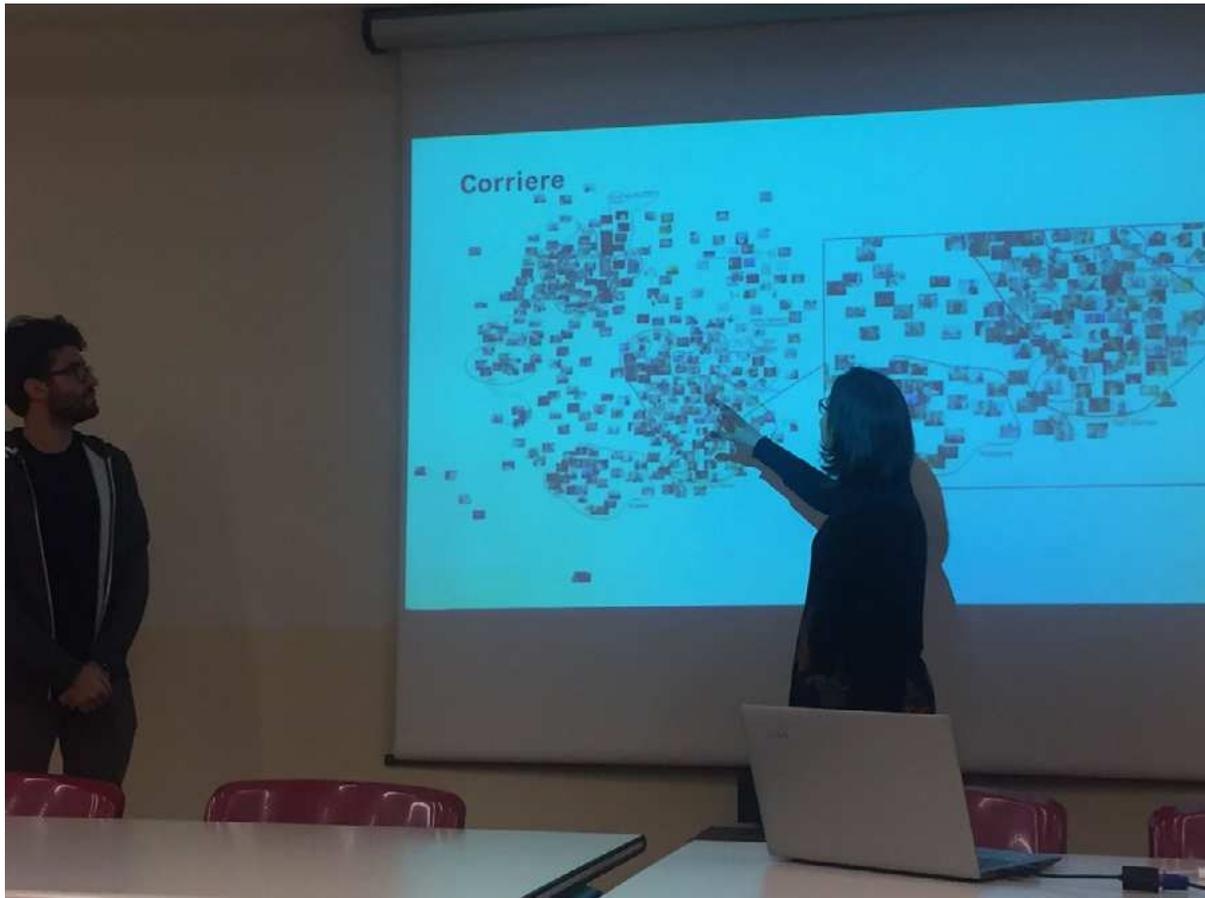


Image 3. Final presentation of project during the 2018 SMART Data Sprint at NOVA University Lisbon, Portugal. Image source: https://smart.inovamedialab.org/wp-content/uploads/2018/07/DVDCJY_XUAAtxJc.jpg

3.3. After | Data sprint results

Writing up data sprint reports

Once the results have been presented, the report that summarises the method and the results obtained to answer the research questions must be finalised. The data sprint is really only over when the project report is published on the venue provided by the organisers (i.e., the event's website). In this regard, it is good practice to draft the report throughout the sprint, and then refine its narrative, after the data sprint has ended.

Project reports derived from data sprints often have a particular structure that differentiates them from academic papers. Typically, their structure² is as follows:

1. First, **the summary of the main findings** in the research is presented. It is essential to write this part properly because it will spark interest in discovering the ins and outs of the research conducted (described throughout the report).

2. An example of a template for the Digital Methods Summer School 2022 report can be found here: <http://bit.ly/dmi22-ss-gdoc-template>; this short link is also available at <https://wiki.digitalmethods.net/Dmi/SummerSchool2022>

2. Second, **the contextualisation of the research** is displayed, detailing the subject matter and the research questions that motivated the work.
3. Third, **the methodology** followed is described: the data and methods used to answer the research questions. **Visual research protocols** (see Mauri et al., 2020) help to inform the platform(s) under investigation, dataset building, tools and analyses.
4. Fourth, **the results, discussions and conclusions** are addressed. These should be written in a concise but clear way, trying to use understandable language not only to those who know how the methods work but also to a wider audience. The conclusion section usually includes, if appropriate, possible applications of the methods and tools used in other types of research, as well as possible future research avenues to continue what was developed in the sprint.

Disseminating data sprint-based research

Research derived from data sprints, if well executed, is often rich and valuable to the academic community. This is because the research is conducted with a real multi-disciplinary perspective and an “open-mindedness” that is hardly found in more traditional work. This freshness allows work derived from data sprints to stimulate interest in the academic community, as it represents a “new” point of view on problems that have been worked on for a long time. Indeed, turning data sprint reports into academic articles has become the standard procedure for disseminating data sprint results. However, as Venturini and Munk (2022: 267) remind us, extensive refinement and documentation work is necessary to turn data sprints’ final presentations (slides or posters) and project reports into scientific research. As they state, without distilling the results, the “quick and dirty” data sprint approach would not meet the accuracy and robustness requirements of scientific research.

Disseminating the results of a data sprint may also require an interactive design process to make the content of the report browsable by a wider audience. Amongst them, web platforms and field guides are the most common communicative products in this context. Web platforms are web pages designed to collect the results of several data sprint experiments which are part of the same overarching research project (Venturini *et al.*, 2014). Unlike collections of slides and project reports, web platforms allow for the deep exploration of interactive digital content, such as networks, maps and huge collages of images. Whereas, field guides (Bounegru *et al.*, 2018) are aimed at stressing and narrating the methodological approach behind sprint-based research and packaging reusable practices to allow others to repeat them.

These are just some of the examples of how dissemination can occur as part of the current literature. However, the widespread use of the data sprint format and its various facets make it well suited as a source for results that are communicated to ever-varying audiences with diverse objectives, such as teaching resources and communication products addressed to the general public.

4. Four facets explain what data sprints are for

This section unpacks the four facets that *pertain to* and *derive from* data sprints, explaining their prospects for a broader audience (Figure 4). We first elucidate how data sprints are (1) a means of teaching and learning about digital methods research, arguing that the data sprint environment is not only (2) a space for methods and tools creation but also (3) a reflective tool to understand the triad of *data-*, *software-* and *platform-oriented research* (from the practice standpoint). Finally, we address a standard post-sprint procedure: (4) the reutilisation of data sprint reports for producing scientific knowledge, through academic and non-academic publications, as an established research practice.

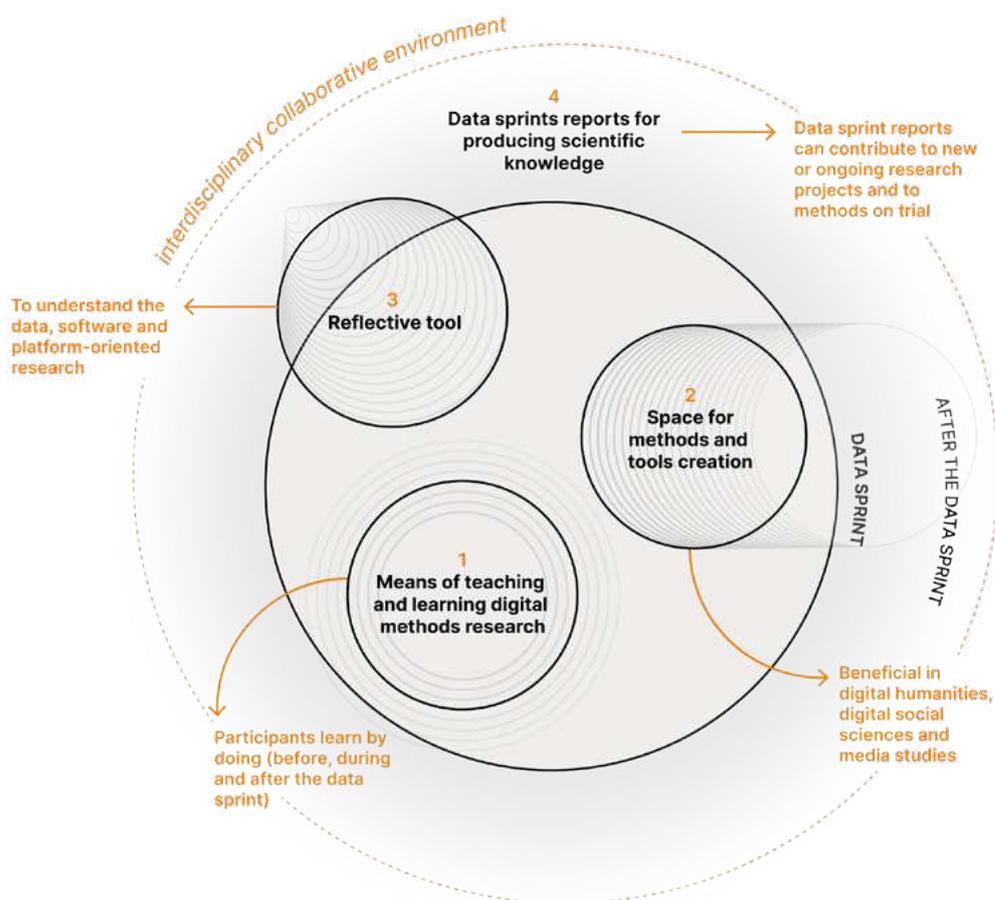


Figure 4. Four facets of data sprints

4.1. Data sprints as a means of teaching and learning digital methods research

After the a data sprint, competences in seeking out methods with different standpoints are achieved. This section elucidates data sprints as a means for teaching and learning digital methods (Omena, 2021; Rogers & Lewthwaite, 2019). Sprints offer a non-traditional sense of instructions, as Richard Rogers explains: the teaching method is tacit knowledge transfer, “where the students are gathering data, using tools, outputting analytical outputs, creating visualisations, and writing reports, all in a compressed timeframe”

(Rogers & Lewthwaite, 2019: 20). Data sprints can be considered research-led teaching devices as the support material (i.e., worksheets, methods' recipes, visual protocols and workshop slides) for the development of projects, and project themes themselves are often associated with broader lines of research and/or research projects followed by those who pitch project ideas/facilitators (Gray et al. 2022). Critical reflection on the role of platforms and software content whilst providing literacy in data analysis and visualisation is also a key teaching principle embedded in sprint activities.

Participants learn by doing by practising digital methods individually and with peers while engaging with a workgroup. By doing this, they understand in practice the rationale behind the digital methods approach (Omena, 2021). The learning process starts from both individual effort and the time pressure imposed on the collective actions and decisions during group work. If participants follow the data sprint preparation instructions (and not starting from scratch), they take significant advantage of the workshops (practical labs or tutorials). When attending these how-to sections, what is taught becomes less of a new element and more of an opportunity to clarify doubts and practice. Group work provides another opportunity for learning digital methods because projects offer "a good environment to understand not only how to make research questions but also learn in practical terms how to respond to these questions" (Omena, 2021: 42). Learning outcomes can thus be used in contexts other than the sprint.

In one way (teaching) or another (learning), being part of a data sprint also means understanding what makes sense when designing a research plan using digital methods and the reasoning behind this research practice. The relation between learning and teaching from data sprints is discussed by Mace Ojala, Laura Kocksch and Katharine Kinder-Kurlanda, who purposefully designed a sprint as a teaching situation. In their article 'Data sprint learning. Exercising proximity to data in teaching situations', they argue that data sprint learning is not exclusively for students; it can be for educators as well. They presented the data sprint preparation as part of a Ph.D. course in digital methods and data critique, and recognised that the teaching role is demanding and includes choosing a theme, collecting and cleaning data and informing students what they should do before the sprint.

Ojala, Kocksch and Kinder-Kurlanda brilliantly describe essential aspects of such learning processes without losing sight of their teaching role throughout the sprint. One situation informs when students should cope with feeling overwhelmed by "too many things to install, learn to operate and understand", from the intense and tight programming involved in data sprints to working with data. Unfortunately, as the authors explain, these tasks can also disappoint some students when they realise the extra effort demanded from them throughout the sprint. Another situation that refers to what the authors call "disconcertment over data" is when the sprint facilitators purposely cause students to feel unsettled, forcing them to, in turn, move from "moments of suspicion and scepticism and turn them into an analytic strategy". In this situation, the sprint facilitators should provide moments to engage with datasets as an opportunity to make sense of the collected data while also being aware of the data curation and collection

processes. These and other situations represent the first encounter of scholars at different carrier levels, not only that of the student, in a data sprint environment.

Òscar Coromina and Adrián Padilla, in their 'Hands On' article, incorporating data sprints and digital methods into digital marketing education (*'Hands On': incorporación de data sprints y digital methods en la docencia de marketing digital* [original language]), also reflect on their experience as teachers with master's degrees in digital marketing who have been organising data sprints with their students for 9 years. They see this practice as a way of presenting students with real challenges in which both students and working professionals collaborate. Coromina and Padilla emphasise that this type of workflow provides students with practical knowledge and digital and communicative competences that, beyond applied research, especially problem-solving practice, serves them in other environments outside the academy. In order for students to get the most out of the experience, the organisation already has workshops and specific teaching on data analysis to prepare them. Following common practice in sprint preparation, the authors include data collection in this phase prior to the sprint. In this way, they make sure that students participating in these *ad hoc* data sprints are at the same, or nearly the same, levels of digital literacy and make much better use of their time.

4.2. Data sprints as a space for methods and tools creation

This section demonstrates how data sprints are a space for methods and tools creation which has proven beneficial in digital humanities, digital social sciences and media studies. The making, testing and/or maintenance of research tools in sprint contexts has been substantial and has made a difference in several research fields, such as web archive research and (visual) network analysis (see Huuderman *et al.*, 2013 and Jacomy, 2022). Regarding methods made through data sprints, we start with Noortje Marres' Situational Analytics (2020), initially applied in a data sprint to analyse testing situations of Covid-19 on Twitter³. Her semi-automated method proposes an interpretative form of data analysis where researchers can actively curate the data and situation under study from large digital datasets. Another example is the three-layered (3L) perspective⁴ for addressing hashtag engagement by Omena, Rabello & Mintz (2020). The method was first proposed in 2017 and later refined and implemented in other projects and sprints. In addition to specific cases, data sprints play a role in a broader agenda for generating visual and vision-reproducible methods (see the work of D'Andrea & Mintz, 2019; Pearce & De Gaetano, 2021; Niederer & Colombo, 2019; and Rogers, 2021).

Giulia Tucci enters this space of developing new solutions and methodological outputs in the article "Visualizing an image network without rendering files: a method to combine user hashtags with computer vision labels". Her method proposal was developed during the 2021 Digital Methods Initiative Summer School. By avoiding the need for memory

3. Project report available at <https://wiki.digitalmethods.net/Dmi/SummerSchool2020COVIDTestingSituations>.

4. The first layer looks at potential differences in the use of hashtags by high-visible users and ordinary users. The second focusses on hashtag activity and the reutilisation of hashtags in social media databases. The final layer looks at the images and texts to which hashtags are related.

space on the computer when visualising networks with images, Tucci's work presents the potential of a creative technique for visualising networks of geolocated images without rendering the image files on the network. Specifically, the paper explores the Twitter case with geolocated images combining the #deepfake hashtags tweeted with the images and the Google Cloud Vision API best single expression to describe each image. She builds a network where nodes are *best-guess labels* and *country flags* while the edges are the images. In so doing, the author argues that it is possible to analyse large image datasets without necessarily seeing the images but with the help of computer vision.

The creative space of data sprints is not limited to methods or tools creation, as the work of Beatrice Gobbo and colleagues shows in "Staging and storing data sprint-based research results: communication design approach". The article's co-authors are a group of communication designer scholars, namely Maria de Los Ángeles Briones, Elena Aversa, Tommaso Elli, Andrea Benedetti, Michele Mauri and Gabriele Colombo. Faced with the challenges of better disseminating data sprint outcomes, the authors developed "a digital *place* that allows data sprints' results to be stored, staged, and accessed after the research activities". This *place* is a tailored platform to communicate the main results of projects developed in data sprints. As one feeds the platform, each step is an invitation to step back, reflect and define what and how to communicate the results. In other words, the authors operationalise ways of translating data sprint reports through the creation of such a platform.

As a source of creativity and spaces for searching for new solutions and building tools, data sprints allow the improvement of methodological processes.

4.3. Data sprints as a reflective tool

Being part of a data sprint working dynamic allows, at a later stage, reflection on how this has taken place, what decisions have been made and what the implications have been. Whether in an academic article or in a more informative way, some participants reflect on data sprints' modes of researching and relating to the community. This section discusses data sprints as a reflective tool to understand the triad of data-, software- and platform-oriented research. The post-sprint time is ideal for critical reflection on what was done, whether the results are helpful and if so how. The reflections come from the practice standpoint, when one should have at least a fair understanding of the rationale behind the adopted methods, proposed analysis and outcomes.

María Concepción Castillo-González, Dorismilda Flores-Márquez and Gabriela Elisa Sued address sprints as a reflective tool in "The data sprint method: An exercise in feminist reflectivity regarding the practices of knowledge production" (*'El método data sprint: un ejercicio de reflexividad feminista sobre las prácticas de producción de conocimiento'* [original language]). On the basis of their experience as well as a focus group with participants in a data sprint, they reflect on the phases of work, the difficulties encountered, the positive experiences of the participants and an evaluation from a gender perspective. Participating in these initiatives allows an "insider's view" of how the research takes

shape in a short period of time. From a critical perspective, they argue that these are also spaces with power differentials and social inequalities. In a sprint, participants with all kinds of backgrounds come together, and amongst them, there are also people who are practically strangers to these dynamics and working methods and have come precisely to learn. In a framework where speed is key to making decisions and getting results, this kind of digital skills gap creates feelings of frustration for some people on the team. The absence of skills is not only owing to lack of knowledge in the field of study, but also in terms of access to tools and training. As Castillo-González, Flores-Márquez and Sued state, tools of this kind are not intuitive and require a solid base for their full understanding, which cannot be achieved during a sprint. The authors also underline the positive effects of collaborative work, such as learning from peers, knowledge sharing and group support to achieve the goal. Regarding the gender perspective, they highlight that the participation of women in data sprints empowers them by promoting the appropriation of digital spaces in these activities.

Data sprints are therefore spaces where experience and observation allow participants to reflect on research processes and ways of working. In this way, thanks to this subsequent reflection, data sprints can also be drivers of change in and of themselves because they are also ideal scenarios for proposing new work dynamics on the basis of previous reflections.

4.4. Data sprint reports for producing scientific knowledge

This section reports another post-sprint activity: the project reports which often offer an updated version of the project's final presentation. Data sprint reports can contribute to new or ongoing research projects and to methods on trial. The reports also support researchers in grounding theory through empirical research and producing hands-on field guides (see Bounegru *et al.* 2018; Marres, 2020). Such reports do not necessarily have to remain just that; many academics appreciate their potential and continue to work on them afterwards, providing them with theoretical frameworks that support such work and that frame the research in a particular academic tradition. That is a scientific contribution. Nevertheless, the production of scientific knowledge generated in data sprints is not only about what is eventually published or even researched during the sprint. The concept is much broader, and also encompasses the participants' own appropriation of the methods they have learned to use and which they will then include in their subsequent research designs.

Jorge Martins Rosa, N. Gizem Bacaksızlar Turbic, Alda Magalhães Telles, Clara González Tosat, Cristian Jiménez Ruiz, Kalliopi Moraiti, Oğuz Özgür Karadeniz and Valentina Pallacci show how a data sprint can be a suitable framework for advancing scientific knowledge. As part of a much broader research project, they decided to use the collection of material to set themselves a challenge: to find the differences in how people engage with the content of the several Portuguese political party pages on Facebook during a period before and after two relevant election campaigns (one national and one European), and to detect what type of content received the most attention from Facebook users (understood as the number of reactions on that platform). This type of

research objective allowed the authors to apply methods and techniques developed in previous data sprints to explore content on Facebook.

Ilya Lavrov, Franziska Schranz, Chiara Miozzo and Marie Palaffre, in their contribution “Visual Communication strategy of populist leaders on Instagram in 2020”, present a study devoted to analysing and comparing the visual communication strategies of 12 populist leaders on Instagram. By tailoring the Tony Blair Institute methodology for identifying populist leaders, the research group analysed a pool of more than 3000 posts containing static images from various perspectives. After a comparative study considering the visual and textual content of the images using Vision API techniques, it turns out that populist leaders publish images that communicate national and personal power. Hence, they portray themselves as ordinary citizens when showing their strengths and reliability. Thanks to their work, it emerges that the use of Vision API and Digital Methods, combined with a qualitative categorisation and coding of emerging clusters, allows for the analysis of communication strategies on Instagram.

Whilst the data sprint workflow allows different methods to be explored collectively like a hive mind, in the post-sprint period, researchers can further analyse and refine the sprint outputs, approving or discarding what they are doing according to the needs of the study.

5. Reflections from and on data sprints

In this article, facets learned from the data sprint approach were identified. We have tried to elucidate important definitions for future applications in a proposal that compiles the anatomy of a data sprint, while taking into consideration its main purposes, uses and meanings that can be given to the practice. Considering the contemporaneity of teaching methods and learning tools in a highly digitalised and fragmented attention world, data sprints operate as a first-hand knowledge experience as few other tools can. The sense of experience and co-creation are perhaps the main differentiators of the practice, as the compact and intense format promotes the high performance and active involvement of the participants.

As a two-way street, the participation process of both those who develop or lead a project and those who are participants is a win-win situation. Collaboration in developing answers or solutions to issues focussed on a research question is key to a good outcome. More than solving technical or objective questions, a *Data Sprint serves as a means of teaching and learning digital methods research*. It is important to highlight that, when dealing with objects of an ever-changing nature, be it methodology, platform, software or empirical, the natural state of a researcher ready to participate is that of an eternal apprentice. This condition of learner that the sprints provide to all who are participating is especially reflected in the sense of being a space for the *creation of methods and tools*. In analogy to an algorithm which, in simplified terms, operates by means of a basic formula, one or more elements of this equation will always be variable. It is this variable that encourages a critical view and that allows us to fail and iteratively improve the research (Venturini & Munk, 2022). As in modern innovation processes, failure is itself

an engine for more challenging research with more interesting results.

Another important point to mention is an *analytical view of the sprinting processes* themselves as a meta-analysis exercise. Considering the short duration of sprints, it is not feasible to have a reflection space, because reflection requires time. In the post-sprint period, this exercise becomes possible. The critical evaluation of methodological and ethical decisions made can also be re-evaluated, generating new protocols and promoting a more solid base that can be replicated by peers and in future sprints. Besides specific technical issues, such as data access, information processing, writing and visualisation of results and so many other processes that are part of the work, it is important to consider lateral elements. Such elements include people management, sensitivity to deal with participants' varied cultural contexts, team development to collaborate and learn in a practical exercise of leadership and soft skills, processes that are rarely mentioned in scientific work in related areas.

Finally, the *variety of outputs* that can result from projects developed in data sprints are considered. Besides the dissemination patterns of scientific knowledge such as in articles and essays, case studies, conferences etc., the relationship of transdisciplinary visions suggests a valid and often unprecedented contribution to various areas of knowledge. In addition to the results themselves, the methodologies developed, the research frameworks and the guides for extraction, treatment and visualisation of data, as well as software and platforms that aid research in the area, are contributions that unite and enhance studies in digital methods on a global scale.

Acknowledgements

The authors want to acknowledge Thais Lobo for her comments and suggestions, and Nate Tkacz, Warren Pearce and Elaine Teixeira Rabello for their constructive feedback.

References

- Berry, D. M.; Borra, E.; Helmond, A.; Plantin, J. C.; Walker Rettberg, J. (2015). The data sprint approach: exploring the field of Digital Humanities through Amazon's application programming interface. *Digital Humanities Quarterly*, 9(3). Available at: <http://eprints.lse.ac.uk/65438/>
- Bounegru, L.; Gray, J.; Venturini, T.; Mauri, M. (2018). *A Field Guide to "Fake News" and Other Information Disorders: A Collection of Recipes for Those Who Love to Cook with Digital Methods*. Public Data Lab. Available at: <https://fakenews.publicdatalab.org/>
- Chao, T. H. J. (2021). Memespector GUI: Graphical User Interface Client for Computer Vision APIs (Version 0.2) [Software]. Available at: <https://github.com/jason-chao/memespector-gui>
- Colombo, G.; Niederer, S. (2021). Visual Methods for Online Images: Collection, Circulation, and Machine Co-Creation. *Diseña*, 19. <https://doi.org/10.7764/disena.19.Intro>

- D'Andrea, C.; Mintz, A. (2019). Studying the live cross-platform circulation of images with computer vision API: An experiment based on a sports media event. *International Journal of Communication*, 13: 1825–1845.
- Gray, J.; Bounegru, L.; Rogers, R.; Venturini, T.; Ricci, D.; Meunier, A.; Mauri, M.; Niederer, S.; Sánchez Querubín, N.; Tuters, M.; Kimbell, L.; Munk, K. (2022). Engaged research-led teaching: composing collective inquiry with digital methods and data. *Digital Culture & Education*, 14(3): 55–86 Available at <https://www.digitalcultureandeducation.com/volume-14-3>
- Hurdeman, H. C.; Ben-David, A.; Sammar, T. (2013). Sprint methods for web archive research. In *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 182-190). ACM. Available at: <https://dl.acm.org/doi/10.1145/2464464.2464513>
- Jacomy, M. (2022, 18 June). Call for participants: Gephi code sustainability retreat 2022. *Gephi Blog*. Available at <https://gephi.wordpress.com/2022/06/18/call-for-participants-gephi-code-sustainability-retreat-2022/>
- Mauri, M.; Gobbo, B.; Colombo, G. (2019). O papel do designer no contexto do data sprint. In J.J. Omena (Ed.), *Métodos digitais: Teoria-Prática-Crítica* (pp. 161–180). ICNOVA.
- Mauri, M.; Briones, M. A.; Gobbo, B.; Colombo, G. (2020). Research Protocol Diagrams as Didactic Tools to Act Critically in Dataset Design Processes. In *INTED2020 Proceedings* (pp. 9034–9043). Available at: <https://library.iated.org/view/MAURI2020RES>
- Mauri, M.; Ciuccarelli, P. (2016). Designing diagrams for social issues. *Design+ Research+ Society-Future-Focused Thinking*, 3(5): 941-958.
- Marres, N. (2020). For a situational analytics: An interpretative methodology for the study of situations in computational settings. *Big Data & Society*, 7(2), 2053951720949571.
- Milman, N. B. (2012). The flipped classroom strategy: What is it and how can it best be used?. *Distance learning*, 9(3), 85.
- Munk, A. K.; Madsen, A. K.; Jacomy, M. (2019). Thinking Through The Databody: Sprints as Experimental Situations. In M. E. Å. Mäkitalo, T. Nicewonger (Ed.), *Designs for Experimentation and Inquiry: Approaching Learning and Knowing in Digital Transformation* (pp. 110–128). Routledge. <https://doi.org/10.4324/9780429489839>
- Niederer, S.; Colombo, G. (2019). Visual Methodologies for Networked Images: Designing Visualizations for Collaborative Research, Cross-platform Analysis, and Public Participation. *Diseña*, 14: 40–67. <https://doi.org/10.7764/disena.14.40-67>
- Omena, J.J. (2021). *Digital Methods and Technicity-of-the-Mediums. From Regimes of Functioning to Digital Research* [Doctoral Dissertation, Universidade Nova de Lisboa]. <http://hdl.handle.net/10362/127961>
- Omena, J. J.; Rabello, E. T.; Mintz, A. G. (2020). Digital Methods for Hashtag Engagement Research. *Social Media + Society*. <https://doi.org/10.1177/2056305120940697>
- Ozdamli, F.; Asiksoy, G. (2016). Flipped Classroom Approach. *World Journal on Educational*

Technology, 8(2), 98–105. <https://doi.org/10.18844/wjet.v8i2.640>

Pearce, W.; De Gaetano, C. (2021). Google Images, Climate Change, and the Disappearance of Humans. *Diseña*, 19, Article.3. <https://doi.org/10.7764/disena.19.Article.3>

Rogers, R. (2021). Visual media analysis for Instagram and other online platforms. *Big Data & Society*, 8(1), 20539517211022370.

Rogers, R.; Lewthwaite, S. (2019). Teaching Digital Methods: Interview with Richard Rogers. Interviewer: S. Lewthwaite. *Diseña*, 14, 12–37. <https://doi.org/10.7764/disena.14.12-37>

Venturini, T.; Meunier, A.; Munk, A.; Borra, E.; Rieder, B.; Mauri, M.; Azzi, M.; Ciminieri, D.; Ubaldi, G.; Kaltenbrunner, A.; Laniado, D. (2014). Climaps by EMAPS in 2 pages (a summary for policy makers and busy people). SSRN. Available at: <http://ssrn.com/abstract=2532946>

Venturini, T.; Munk, A. K. (2022). *Controversy Mapping: A Field Guide*. John Wiley & Sons.

Venturini, T.; Munk, A. K.; Jacomy, M. (2018). “Data-Sprinting: A Public Approach to Digital Research”. In E. Lury, C., Fensham, R., Clough, P., Heller-Nicholas, A., Lammes, S., Last, A., Michael, M., Uprichard (Ed.), *Handbook of Interdisciplinary Research Methods Routledge*. Routledge International Handbooks. <https://doi.org/10.4324/9781315714523>

