



HAL
open science

Repurposing Digital Methods for Human-Centered Design

Donato Ricci, Agata Brilli, Roberta Tassi

► **To cite this version:**

Donato Ricci, Agata Brilli, Roberta Tassi. Repurposing Digital Methods for Human-Centered Design: Distilling Data-Driven Personas from Twitter Discussions: The case of Urban Nature in Paris. DRS 2018 - Catalyst, Jun 2018, Limerick, Ireland. Design Research Society, 2018. hal-01901091

HAL Id: hal-01901091

<https://sciencespo.hal.science/hal-01901091>

Submitted on 22 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Repurposing Digital Methods for Human-Centered Design

*Distilling Data-Driven Personas from Twitter Discussions.
The case of Urban Nature in Paris*

RICCI Donato^{a*}; BRILLI Agata^b and TASSI Roberta^c

^a Sciences Po

^b Oblo

^c Politecnico di Milano

* Corresponding author e-mail: donato.ricci@sciencespo.fr

doi: 10.21606/dma.2018.308

From logs and information left in online spaces to data points self-generated by connected devices, digital traces have become more diffused over the past years, prompting an expansion of Human-Centered Design methods. Along with some *big-data* approaches, Digital Methods of research – treating the actual content of digital users' manifestation on-line (i.e. tweets, Instagram pictures, comments) – offer the opportunity to better understand users through their online activities. This paper investigates how Digital Methods can be repurposed as a full-fledged approach for Human-Centered Design. Grafting on the NATURPRADI project – a research aimed at describing the debate raised by the re-vegetation of the city of Paris by analysing Twitter posts – in the paper we will explain how we have identified and described a set of *personas* characterized by different approaches towards the evolution of the urban nature issue. The final objective of the paper is to provide a first methodological tool created at the intersection of Digital Methods and Human-Centered Design discussing its opportunities and criticalities: *Data-driven Personas*.

personas; human-centered design; digital methods; mixed methods

1 Disciplines in transition: coping with socio-technical complexity, policy problems and digital data.

In the last two decades, the Design discipline has gone through a profound redefinition of its scope of action. We are acknowledging an entire disciplinary field experimenting with new “ways of thinking and doing” (Manzini 2016) in the face of growing environmental, technical and political issues in our society (Cross 2011; Ehn et al. 2014). These issues are characterized by the contested status of techno-scientific knowledge, the difficulties that traditional institutions show for



intervening in them and the necessity of redefining political and participative practices (Le Dantec 2017).

On one hand, more and more research institutions, consultancy agencies, think tanks and policy labs have started to apply User-Centered Design methodologies to address socio-technical complex and controversial settings (Kimbell 2015; Blyth et al. 2011). Positioning the user at the centre of the innovation processes helps organizations and institutions to come up with better answers to the complex challenges they are facing (Brown 2008). Furthermore, it allows to develop strategies and solutions grounded into a deep understanding of user needs, motivations and expectations. This understanding is generated through the application of design ethnographic techniques, and requires the researcher to spend time observing the users in their context in order to gain information and inspirations around their behaviours.

On the other hand, more germane fields of study to such complex socio-technical issues and problems (i.e. Science and Technology Studies, Political Sciences, Media Studies and Public Affairs) are experimenting new methodologies fitting into the so called *digital-turn*, expanding the notion of in-field research to the online domain. Research approaches based on the *digital* traces left over the Web and conducted in the framework of digital (Rogers 2009) and *quali-quantitative* methods (Venturini & Latour 2009) are opening the possibility to collect and analyse a wealth of data to observe and describe such complex environments.

The hypothesis of the present paper is that a promising way to cross these two tendencies, extending the contribution of Design disciplines in wider scientific contexts, is to continue and reinforce the circulation of approaches and methods between Design and Social Sciences. In other words, as Human-Centered Design has borrowed and repurposed Social Sciences methods (Hanington 2006; Ingold 2014) - from ethnography and anthropology, for example, to enquiry into situated practices and localised markets — this could continue re-imagining the use of digital data and methods in specific, controversial and complex Human-Centered Design contexts. The main question that this paper will tackle is: How can Digital Methods provide relevant insights and be integrated into a Human-Centered Design process?

2 Human-Centered Design, scope and application

Human-Centered design is a creative approach to innovation and problem solving that focuses on understanding people in order to ideate solutions that suit their needs (Brown 2009). The approach, also known as User-Centered Design, was initially introduced by Donald Norman (1986) – who evolved the concept of usability testing to embrace the need of learning more about user interests and needs. When the design practices encountered broader challenges due to the increasing introduction of technologies and connected systems as well as the strong push towards addressing complex societal transformations, Human-Centered Design approaches have become more and more relevant and asked to drive innovation processes by offering a way to handle that complexity and giving priority to the user perspective in generating and validating solutions (Venturi 2006).

User research is at the core of Human-Centered Design and defines all the activities that the designers put in place to engage users in the design process and gain inspirations and information from them (Steen 2012). User research requires to reach users where they are, spend time in their context, listen carefully. This type of research is by definition qualitative: it studies one individual behaviour at-a-time, and then elaborate on common patterns. *User frameworks*, such as *personas*¹ (Cooper 1998), are then used to crystallize and share those patterns, building a narration around the different types of behaviours that the research enlightened and turning those insights into actionable materials for ideation.

¹ Personas are fictional narratives that describe specific clusters of users and behaviours identified during a previous phase of data collection (defined as user research).

The extensive application of Human-Centered Design to address technical, environmental and political issues has raised the need to pair the qualitative understanding with quantitative data to better support decision-making processes. Surveys are a good example for quantitative research used in the Human-Centered Design process. Quantitative research can either set the ground for a more effective fieldwork - by doing an assessment of a large group of users before defining who to engage in the user research- or complete the picture offered by the qualitative study – by adding numbers that refine and enrich the insights emerged during the user research. On top of that, the increasing importance of the digital sphere and the high amount of data available there has challenged Human-Centered Design to look at the online dimension as an additional research context (Seemann, 2012). While marketing agencies have integrated web metrics and social listening tools in their practices to use the big data to better understand and segment their market², the Human-Centered Design discipline hasn't offered yet a set of methods to easily collect and use that information in the design process. This information might be extremely useful to mitigate and overcome the common limits (Chapman and Milham 2006) and scepticism (Cabrero et al. 2016) regarding the use of personas and more generally of “simplified accounts of people” (Turner and Turner 2011, Floyd et al. 2004).

With this paper we want to propose a hybrid approach that looks at the digital space as a field for quantitative and qualitative investigation in a sustainable and approachable way, providing a method that could be easily integrated in the Human-Centered Design toolbox.

3 Digital Methods

We argue that using digital data and analyse them both qualitatively and quantitatively can be extremely useful in complex social, technical and economic contexts where design is called to intervene. Proving this hypothesis requires to address different lingering challenges facing design theory and practice (The Design Collaborative 2014): How to cope with a heterogeneous and conflicting spectrum of values and interests? How to collaborate with other disciplines. How to stabilise specific research methods and protocols? How to test them in large scale empirical experiments?

While we will try to tackle the latter two questions further in the article, the first two ones are related to the controversial nature of the issues faced by design intervention. This is the case of all those issues where their very same definition is questioned by various actors and is redefined by the means of new technologies, governance settings and social representation (Venturini 2009). To study them, a specific research methods emerged, called Controversy Mapping (Latour 2007). Controversy Mapping proposes a data-acquisition protocol drawn on the theories and practices of Digital Methods of research (Rogers 2009, 2013). They exploit the wide range of traces that are left on the Web by the very actors of the issue under analysis. Digital Methods further a social research approach taking advantage of the empirical capacities embedded in online activities (Schneider & Foot 2004) with their unique dynamic nature – a mixture of ephemeral and permanent elements – (Hewson 2003).

Digital Methods differ from the *big-data* research programs. The emphasis of Digital Methods is not in the magnitude of digital data analysed but in the critical affordances deployed by the data-acquisition protocol. Digital Methods protocols are deriving significant findings from relatively small, *ad-hoc designed*, data-sets (Marres & Weltevrede 2013). By following a series of iterative steps and refinement procedures, the final formatted data are carved out of informational disarrays and unformed mass of online digital objects. This process has the advantage of avoiding the risk of projecting pre-existent categories on the issue in analysis. Being a challenge of data selection and curation, they provide, through the redaction of their protocols, an evident and traceable inspection of the qualitative decisions taken to compose datasets and corpora.

² One of the best example is offered by the Linkfluence (linkfluence.com).

4 Digital Methods and design disciplines

Until now the disciplinary exchanges between Digital Methods and Design disciplines mostly concerned two methodological procedures. On the one side design disciplines, and particularly Communication Design, have strongly contributed to produce the new graphical means required to visualise and represent digital data and information (Ricci 2010, Venturini et al. 2015, Mauri 2017), reinforcing the role of a *visual culture* into social research³. On the other, design disciplines have contributed during these experimentation and research projects with some organisational and management techniques. Typical of Digital Methods researches is the concept of sprints. These temporary gathering of social and computer scientists, designers and activists, are often traced back to the *hackathons* (Venturini et al. 2016) or to other compressed, rapid and collaborative forms of collaboration like *Book Sprints* (Berry 2015). Neglected in these two accounts is the contribution of the design discipline in installing a participatory-in-nature work environment, sustained by the creative engagement of the participants. A sprint, in design terms, would be called *design workshop*⁴ (Hanington 2006).

While the contributions on data visualization and on the reinforcement of non-conventional collective and participatory working practices have been fundamental to the development of Digital Methods, scarce are the traces of a reciprocal hybridisation. Nonetheless, Rogers (2014) argued that designers involved in Digital Methods learned more about other's disciplines procedure rather than improving on their own. We argue that, eventually, also Design discipline, and particularly Human-Centered Design, could benefit from the synergy developed with Digital Methods. In the following section we will describe an experiment in developing a proper method, the *Data-Driven Personas* that repurpose Digital Methods, for investigating, and intervening in complex social and technical environments.

5 Data Driven Personas

To propose a fertile synergy between Human-Centered Design and Digital Methods, and set the ground for a replicable empirical methodology, a full-scale test has been conducted. The initial hypothesis for this research is that it is possible to understand users' behaviours, needs and expectations we are designing for by collecting and studying their online traces. Distilling relevant information out of these online traces can lead to the identification of clusters of users to be then described as personas. The test has been set up drafting the *Data-Driven Personas* method on an ongoing research project investigating the issues raised by urban nature in Paris: The NATURPRADI⁵ project.

³ One of the best example is the EMAPS project (emapsproject.com), funded by the European Commission in the FP7 "science in society" programme. It was aimed at exploring "the opportunities and risks in the use of the web and the social media as a meaningful information tool and for developing a participatory communication between scientists and the different publics". The project focused on the emerging uses of the web as a tool of collective endeavour and public debate, in two particular techno-scientific issues (aging/life expectancy and climate change adaptation) and developed an "open-air" experiment developing an online interactive platforms (climaps.eu). Other experimentations acknowledged a peculiar assemblage of design research groups and design studios. A successful experiment (chemicalyouth.org/visualising-erowid) has been conducted in the Chemical Youth project by the University of Amsterdam that thanks to the collaboration with Calibro, a small design agency, developed original methods to collect and visualise information about the use of drugs from a specialised on line discussion space.

⁴ A typical design workshop such as a research-sprint, constitutes, as noted by Muller (2003), a "third-space" in which different knowledges are catalysed towards a common outcome "which emerge through negotiation and co-creation of identities, working languages, understandings, and relationships, and polyvocal (manyvoiced) dialogues across and through differences".

⁵ NATURPRADI – Urban Nature in Digital Practice(s) – is a 3 years research project funded by ADEME (French Environment and Energy Management Agency) as part of MODEVAL-URBA 2015 call for project. It started in September 2016 and is led by the LAVUE Architecture and Anthropology laboratory, which partnered with the City of Paris Office of Green Spaces, the Museum of Natural History Eco-anthropology and Ethno-biology laboratory, the Paris Urbanism Agency, and Sciences Po|médialab.

5.1 The NATURPRADI project and its Digital Methods campaign

NATURPRADI is aimed at observing and describing the effects of the many initiatives endorsed by the Paris municipality to revegetate the city. These initiatives are trying to produce *smart* solutions to a growing range of issues created by urban growth⁶. Nevertheless, there is no agreement on the imaginaries and technical practices that should be included into this new urban nature (Gandy 2006). Urban nature issue, while representing a vast field of socio-technical experimentations, on the one hand, nature is seen as an aesthetic strategy to create the perception of a city that integrates (itself in) nature to secure better health conditions. On the other, nature is seen as an engineered process and as a technical solution, for example to the “heat island” effect. But, beyond these objectives, green urban design has also been used as a “mask” for a capitalist urbanization resulting the social inequalities of “green gentrification”. Furthermore, engineered forms of urbanity are strongly entailed to corporate and technocratic visions. These mainstream narratives, obfuscates a muffled collection of contested natures (Macnaghten & Urry 1998), fabricated by technological and environmental objects and socio-cultural processes (Gandy 2002), urging for public participation and engagement.

To observe, monitor and, eventually, produce elements of reflections for future urban policies, the NATURPRADI project is mapping the symbolic and material elements of the urban nature debate (Ricci et al. 2017). The research project is aimed, at exposing the different social, political and technological issues associated to urban nature; its actors, and the controversies caused by alignment and misalignment of interests. The NATURPRADI project seeks to explore how the future mosaic of urban nature in Paris by observing how objects, places, practices and technologies are mobilised, re-appropriate and discussed in the public arena. To achieve its objective, NATURPRADI started a Digital Method campaign by collecting digital-native contents produced on Twitter. The online news and social networking platform has been chosen since it is broadly used by a variety of actors getting spontaneously organised around discussion topics by using hashtags. Twitter, presenting the concrete opportunity for “empirical sociocultural research” (Burgess & Bruns 2015), has become, over the years and despite its transformation, an object of study and a data source for research scopes (Rogers 2013b).

The core of the NATURPRADI project is to investigate and elicit users’ viewpoints and actors’ perspectives on urban nature. In the project it has been possible to observe and describe the different sub-topic composing the debate on such delicate issue. The project has identified these topics and how they are sustained by specific communities and populated by identifiable users, each of them proposing an instantiated vision of the future Parian urban nature. For this reason, NATURPRADI has provided the great opportunity to test a new process moving from users to personas, through a combination of quantitative and qualitative methods of analysis. Here we argue that the critical re-appropriation and the repurposing of Digital Methods could be an efficient resource to support the Human-Centered Design practice when dealing with complex issues.

5.2 The process of distilling Data-Driven Personas

The Data-Driven Personas method could be summarised into four macro-phases (Figure 1):

1. **Data collection**, aimed at grafting the Human-Centered approach on Digital Methods defining the nature and the scope of the data harvested as well as their limitations;
2. **Exploration of the discursive space**, aimed at finding an entry point of the investigated topic, displaying the constellation of debates emerged from the collected data.
3. **From cluster to communities**, aimed at finding the cohesive communities composing on a

⁶ In the last decade a vast array of new policy devices – the Budget Participatif initiative, the project Du vert près de chez moi, the regulatory grant Permis de végétaliser, the charter Objectif 100 hectares, the Parisculteurs call for project, a digital platform called Végétalisons Paris, have been developed to engage with citizens, national and multinational companies and local public actors, with a mix of bottom-up and economic incentives, to promote greening initiatives, share best practices, and develop communities of urban gardeners and biophilic supporters.

specific issue who can then be turned into personas;

Personas descriptions, aimed at describing each personas, making use of the most relevant qualitative and quantitative aspects of data, emerged during the research process.

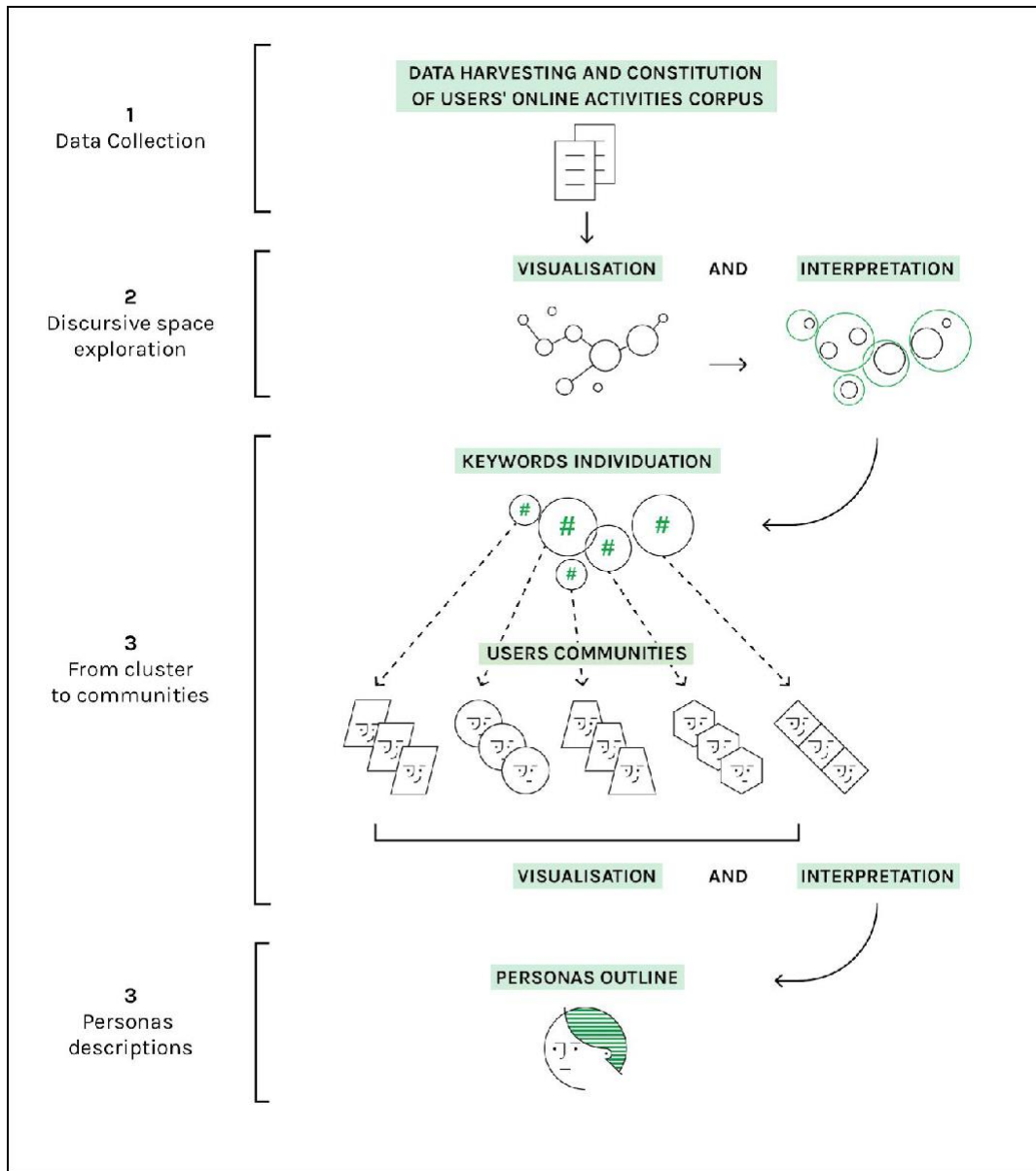


Figure 1. Methodology process

5.3 Phase n.1: Data collection

Objective: Define the nature of the information to be collected and set up the protocol to harvest the corpus of data.

To start harvesting and collecting the data for the distillation of the Data-Driven Personas, requires a deep understanding of the bias induces by the kind of users' traces (e.g. hyperlinks, tags and hashtags, threads, ranks or edits) collected for the analysis and by the services, platform and devices that are offering them (e.g. Facebook, Twitter, Wikipedia, blogging platform, search engines). Every Digital Method approach poses under scrutiny which digital objects are provided by the different devices. This has been the case for the NATURPRADI project. After having chosen Twitter for collecting the manifestations of interest, proposition and disagreement towards the Parisian urban

nature, among the several methods and technique for harvesting of tweets, the Streaming API⁷ has been adopted. It has required to acknowledge and consciously embrace its limitation⁸, specificity and embedded politics (Gillespie 2010, VanDijk 2013), technically, rhetorically and culturally expressed (Gillespie 2014). Only their clear understanding, obtained for example through the collaboration with Digital Methods experts, allows later on in the process to mitigate and validate the results of the process.

Among the different approaches for Twitter corpora building (see Mayr and Weller 2017) it has been chosen one based on key expression query. Through a collaborative and participatory procedure among the members of the NATURPRADI consortium a list of 158 expressions (Figure 2) has been used to capture the tweets mentioning them.

It has to be remarked that although this procedure has granted satisfying results, it is impossible to achieve a kind of “completeness” about the tweets referring to a specific issue. Users can always use different wording, or certain pertinent tweets may have been not capture because they didn’t mention any keyword of the list. As partial mitigation of this bias, once a tweet part of a conversation has been captured the entire discussion to which it belongs has been retrieved. To assure a territorial specificity to our corpus we queried only for French word. In addition to this, all the keywords are queried by adding the word “Paris”⁹.

agriculture - agricultures - agriculture biointensive - agriculture verticale - agroforesterie - agrosylviculture - alimentation proximité - #AMAP - aquaponie - arboriculture - arbre - arbre alignement - arbres alignements - arbre urbain - arbre_remarquable - arbres - arbres urbains - arbres_remarquables - arbuste - arbustes - architecture écologique - aromatique - aromatiques - association_végétale - bande_enherbée - biodiversité - biodiversité bâtiment - biodiversité jardin - biodynamique - biointensive - botanique - botaniques - cartographie végétation - chantier nature - compost - composts - compostage - compostages - conservation nature - corridor biologique - corridors biologiques - "coulée verte" - "coulées vertes" - développement - durable - éco-habitat - ecoagriculture - écocitoyenneté - écoconstruction - écologie urbaine - écologique - écologiques - écoquartier - écoquartiers - espace vert - espaces verts - ferme verticale - fermes verticales - fleurie - fleuries - flore des murs - floriculture - forêt - forêt urbaine - fragmentation écologique - fruitier - fruitière - fruitiers - génie écologique - génie_écologique - gestion différenciée - graminée ornementale - "guerrilla gardening" - horticulture - horticulture urbaine - îlot - îlot chaleur - îlots chaleur - incroyables comestibles - infrastructure verte - jardin - jardins - jardin communautaire - jardins communautaires - jardin ouvrier - Jardin-forêt - jardinage - jardinage collectif - "jardin familial" - "jardins familiaux" - "jardin public" - "jardins publics" - "jardin sauvage" - "jardins sauvages" - jardinage urbain - jardins ouvriers - matrice écopaysagère - micro-agriculture - micro-ferme - micro-fermes - mur végétal - mur végétalisé - murs végétalisés - murs végétaux - naturalité - nature - observatoire paysage - patrimoine naturel - paysage urbain - paysages urbains - permaculture - plan climat-énergie territorial - plante - plante grimpante - "plante utile" - plantes - plantes grimpantes - polyculture - potager - potagère - potagères - potagers - prairie - prairies - prévégétalisation - "produit local" - "produits locaux" - alimentation proximité - renaturation - "réseau écologique" - "réseaux écologiques" - résilience écologique - soutenable - soutenables - stratégie biodiversité - sylviculture - système d'information sur la nature et les paysages - terrasse végétalisée - terrasses végétalisées - toit-terrasse - toits-terrasses - toiture - toitures - trame verte - trame_verte - trames vertes - végétale - végétales - végétalisation - végétalise - végétalises - végétalisée - végétalisées - végétation - verger - vivrière - grenelle environnement - effet serre

Figure 2. The final list of expressions used to retrieve the tweets. All of them have been used in combination with the word “Paris”

⁷ These API offers the possibility to retrieve only live data, imposing bandwidth limitations coming into effect when the requested tweets exceed the 1% of the all traffic flowing in the platform.

⁸ A limitation affecting Twitter based researches is linked to its representativeness (see Blank 2016). Although Twitter is widely used all across the world, its adoption rate changes accordingly to different social milieux and the way it is used may differ significantly from country to country. In the NATURPRADI project there is no assumption about the possible exact extension of the observed digital population to the general one.

⁹ To assure that the final corpus would not be biased by tweets not related to Paris or to the urban nature, a further curatorial procedure has been applied. Through a custom and open-source software (the source code is available here: <https://github.com/medialab/catwalk>), every tweet has been read by the research team and evaluated in terms of its pertinence. This approach, distinguishes the NATURPRADI project from many other big data ones. Furthermore, the close reading of the tweets enabled us to have a constant overview of the state of the discussion, gaining a deep understanding of the dynamics of the issue. This aspect resulted to be extremely useful in the analysis and interpretation of the data.

5.4 Phase n. 2: Exploring the discursive space

Objective: Visualise the corpus in order to get a synoptic view on the issue under analysis and identify the main components.

Many controversial and complex issues, as the case of urban nature, emerge from social, biological, political and technological interplays, requiring a series of mixed analysis and visualisation approaches able to document and expose visually the different discourses that emerge from them. In order to grasp the complexity of the issue, a set of visualisations aims at exploring the different topics that constitute it and how they are related to each other constituting in this way a heterogeneous discursive space.

Identifying the clusters of people and verbal expressions for each topic results in a first formulation of *personas hypothesis*¹⁰ (Cooper 2014).

In the case of the test with NATURPRADI, the visualisations consisted in a series of graphs¹¹ showing how different users are attached to each other and the specific words used in the tweets. Reading and interpreting the graph it could be remarked the centrality of an institutional cluster featuring linked to the municipality (@Paris, @Anne_Hidalgo, @PKOMITES and @vegetalisons). On the left of this group, there is a cluster concerning urban farming and bio-agriculture activities. In this cluster is located the municipality's initiative *Parisculteurs* that finances interventions relate to green and agricultural urban areas. On the right of the institutional accounts, lays a cluster about participative initiatives deployed by the municipality of Paris to promote the citizens' engagement, like the *Permis de Végétaliser*. Furthermore, another relevant cluster is located under the institutional accounts, it features ecology and recycling related topics. Looking at the right edge of the discursive space, there is the cluster where all the famous green areas of Paris are mentioned. Finally, at the opposite edge, can be detected the cluster about innovative agricultural techniques and startups.

On the basis of this quali-quantitative interpretation of the graph, sustained by a visual analysis of the network¹², five key clusters can be identified (Figure 3):

- A. **Technological Development**, featuring innovation initiatives and project in the agricultural field;
- B. **Urban-Agriculture**, featuring bio-agricultural projects developed inside the city of Paris;
- C. **Co-design of Public nature**, featuring all the debates around the participatory activities endorsed by the municipality;
- D. **Ecological attitude**, featuring the concerns about the ecological transition like the domestic recycle of wastes.
- E. **Relaxed Contemplation**, featuring the discussion about outdoor activities around Parisian gardens.

The exploration of the discursive space and the description of these key-clusters allowed to identify the topics most vibrant and relevant for the users are the most vibrant and. Drawing upon these overlapped delimitations, the process has moved towards a finer description and characterisation of the users present in each cluster.

¹⁰ The persona hypothesis is a first attempt to define the different kinds of users for a product or service. The hypothesis serves as an entry point to start with the interview planning.

¹¹ The graphs have been produced using the tweets of the trimester January-March 2017, the same range has been used throughout the entire process. To produce the networks, the elements contained in a single tweet —mentions to other users and hashtags and other relevant n-gram has been transformed in a clique. The combination of these cliques produced the final network.

¹² The visual network analysis is based on a spatialization achieved through the ForceAtlas2 algorithm (Jacomy et al. 2014).

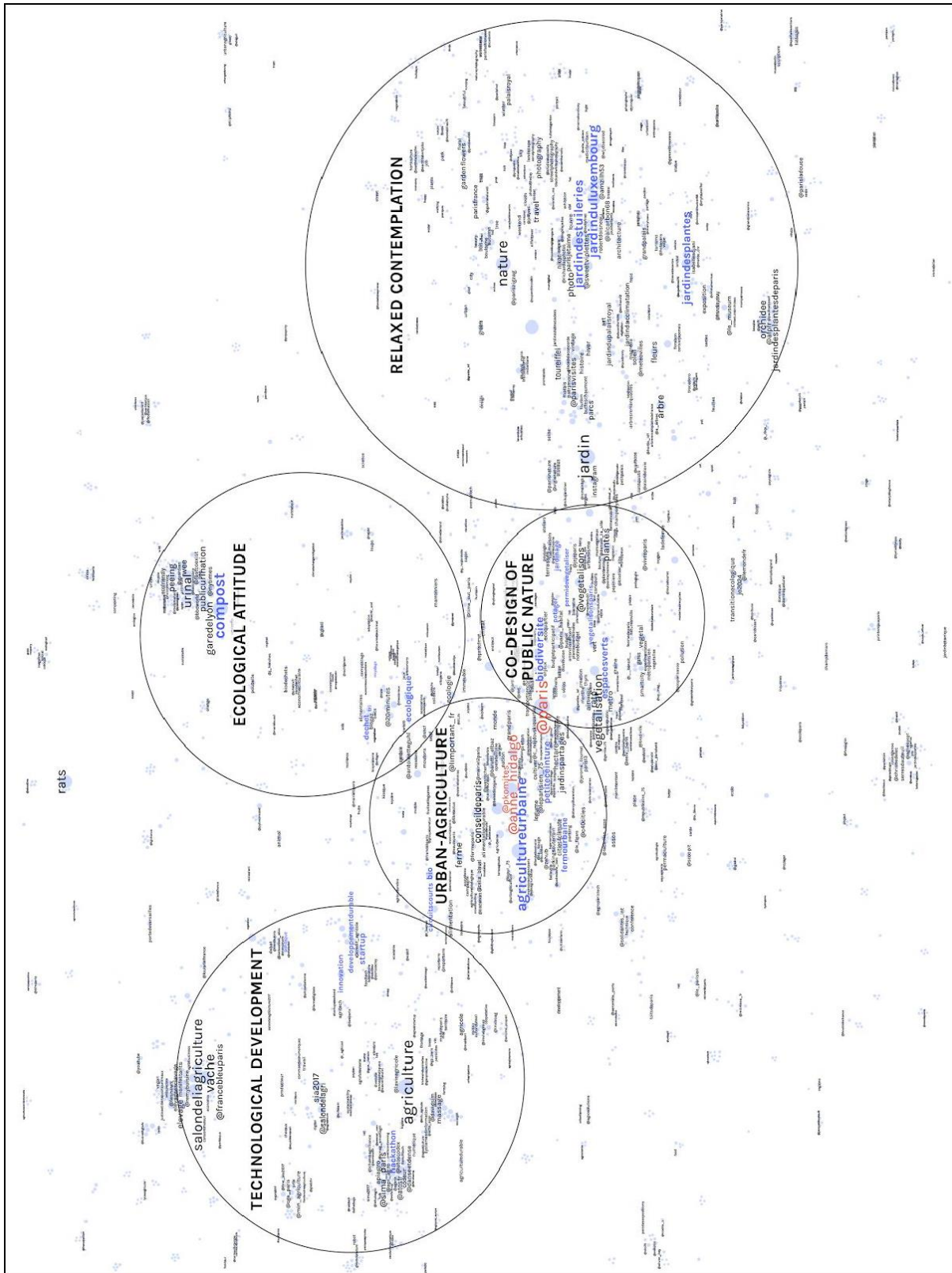


Figure 3. The trimestral tweets user-object networks

5.5 Phase n. 3: From cluster to communities

Objective: Analyse each sub-topic, identify the communities linked to them, and understand if they represent a unique behaviour.

While the *exploration of the discursive space* allows to quickly highlight the most relevant topics and clusters composing a complex issue, through a deeper analysis it is possible to closely identify the users that populate them and discover if they represent a cohesive community. To achieve this objective, it is necessary to detect some distinctive features characterizing each cluster. They might be the most specific used terms or external references used in the discussions. Then the collected corpus of digital traces has to be subdivided into sub-corpora using the detected features in order to isolate the users forming each cluster.

In the NATURPRADI test, for each cluster a list of keywords¹³ was produced to assign the users to one or more specific communities. For instance, a user is part of the *Ecological Attitude* community whenever she used at least one of the keywords *compost, ecologique, dechet, tri*. If that same user wrote any keyword related to other clusters, she would appear also in those communities¹⁴ (Figure 4).

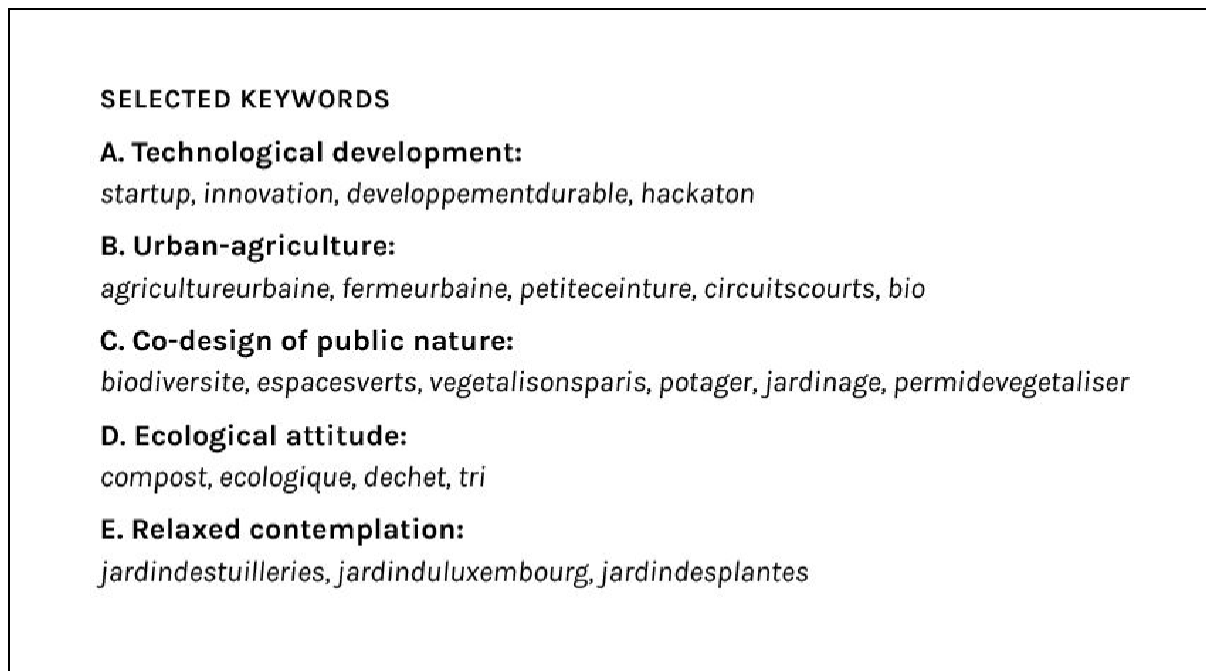


Figure 4. The selected keywords for each cluster used to retrieve the communities of users which used these words in their tweets

The corpus is then used to understand if, besides debating about the same topic, they also debate in a similar way. Just like when the researcher carries out field investigations to collect more insights about how people live, the visualisation and interpretation of different aspects of users' online activities become an iterative process that enables to progressively validate the cohesion of communities. This iterative process consists in visualising the multiple dimension of the corpus (e.g. images, texts, links), then in interpreting the results to understand if there are similar communities

¹³ The keywords have been selected by their frequency, their specificity (for example nature has been discarded because too generic) and their non-ephemeral nature (for example fashionweek). Furthermore, when there two or more terms share same root, only the most recurring one has been chosen (e.g. ecology and ecological).

¹⁴ A user can belong only to one community as well as more than one. In the latter case, one could imagine analysing more thoroughly only those users belonging both communities to see if they constitute a separate and distinct behaviour.

which can be merged together and considered as a unique behaviour or, on the contrary, if inside a community more than one distinctive behaviour can be discerned.

In our test, we have focused on the two main elements of a tweet, its textual content and the possible images attached to it. In this way the iterative process leading to the definition of the validate behaviours has been concerned with both the textual and the visual sphere and their overlap.

5.5.1 *Digging into the textual sphere*

Analysing the most frequent and relevant vocabulary elements that a community use to discuss a given issue, is extremely relevant to identify both the commonalities and distinctive traits of each community. In our test, the visualisation of the textual sphere shows the 150 most recurrent terms for each community, sorted from the most to the least frequent (Figure 5). The size of each bubble is proportional to the frequency of the word. The colour of the bubbles describes how much each word is shared with other communities: the lightest the colour, the most shared the word; the darkest the colour, the least share the word – which means that it is uniquely used by a specific community.

While the most used words by the *Technological development*, *Urban agriculture and Co-design of public nature* have proved to be in accordance with the initial depiction of these communities, the interpretation of this visualization lead to an interesting observation concerning the *Relaxed contemplation* and the *Ecological attitude* communities. About the *Relaxed contemplation*, the names of several famous French photographers occurred among the most frequent words. This could suggest the presence of a smaller community with an interest in photography and in the historic representation of the city. The *Ecological attitude* community seems mostly linked to the uproar news of the urinal-vases installed in Paris by the municipality. Furthermore, the textual sphere appears comparable with the vocabulary used by the *Urban-agriculture* community. This could suggest that the two communities share parts of same debates.

5.5.2 *Digging into the visual sphere*

Analysing the images that are produced and shared by the users enables a quick introduction to the imagery of each community. While in Human-Centered researches, visual information, to better understand the users, are collected through mostly analogue methodologies like *photo-journals*¹⁵, Data-Driven Personas make use of the images produced by users during their online activities. In order to depict the visual sphere, the images are automatically plotted according to their content similarity¹⁶.

By interpreting the visualisation, it can be remarked that, in the *Relaxed contemplation* cluster, the sphere depicts the most famous Parisian architectures and green areas (Figure 6). This seems to be coherent with the fact that this cluster is composed mainly by tourists and those who appreciate the aesthetic role of the Parisian nature. In the lower part of the network there is a significant group of historic images, which corroborates the presence of a sub-community of *Nostalgic* users.

Repeating the process for the other clusters, the visual sphere analysis helps to understand communities' cohesion (Figure 7-10).

¹⁵ The photojournal consists in asking a group of user to take some pictures in accordance with an assignment. BROBERG, Ole; ANDERSEN, Vibeke; SEIM, Rikke. Participatory ergonomics in design processes: The role of boundary objects. *Applied ergonomics*, 2011, 42.3: 464-472./

¹⁶ For this task, it has been used the Image Tagging API provided by IMAGGA (<https://docs.imagga.com/#auto-tagging>). The model used by this service is trained on generic images, therefore it was considered suitable for the NATURPRADI dataset, composed of very different type of images (photos, drawings, posters, screenshots). In this way a set of tag is generated and attached to each image. A weighted bipartite network is then built using images and tags. The weight of the edges corresponds to the confidence interval that the algorithm gives for the association image-tag.

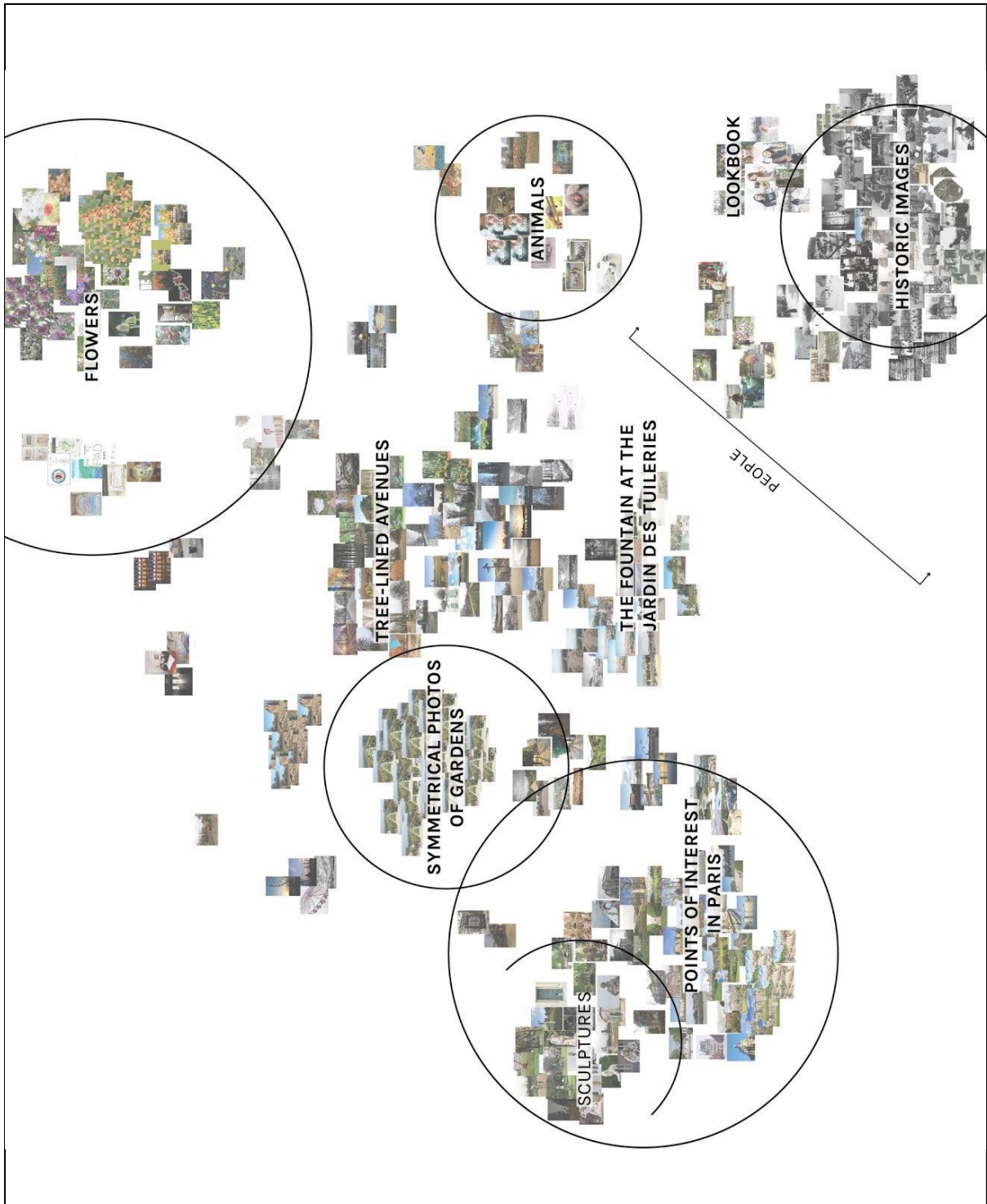


Figure 6. The visual sphere of the “Relaxed contemplation” community

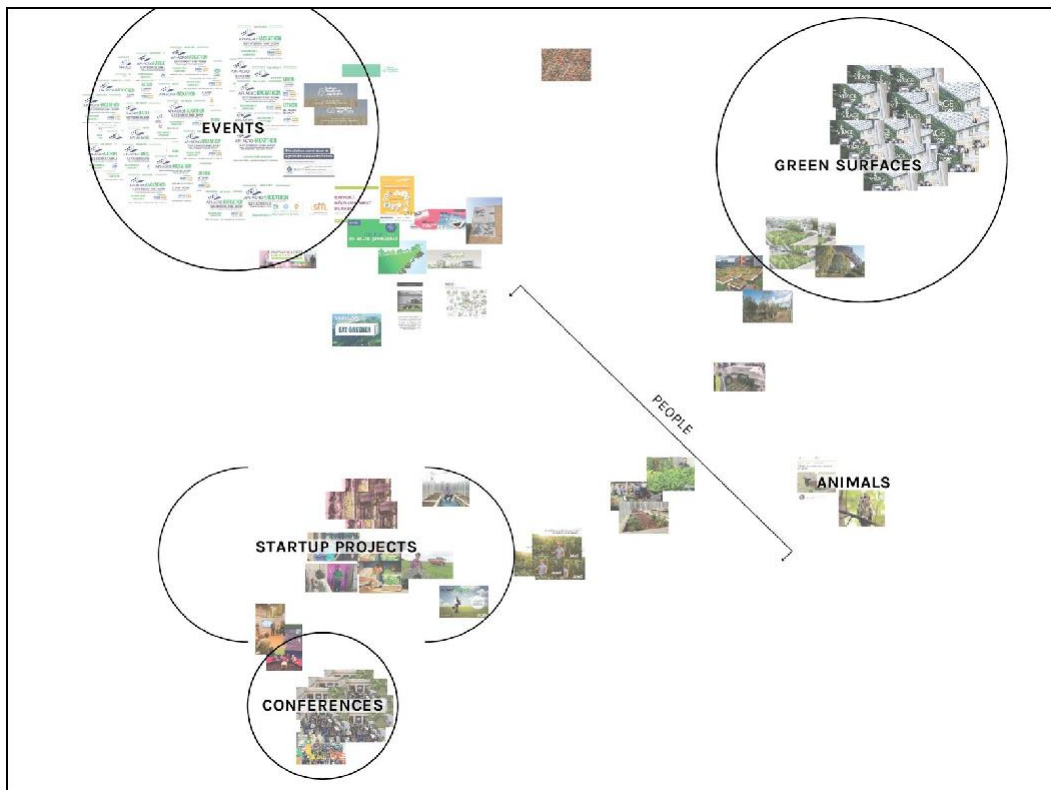


Figure 7. The visual sphere of the "Technological development" community

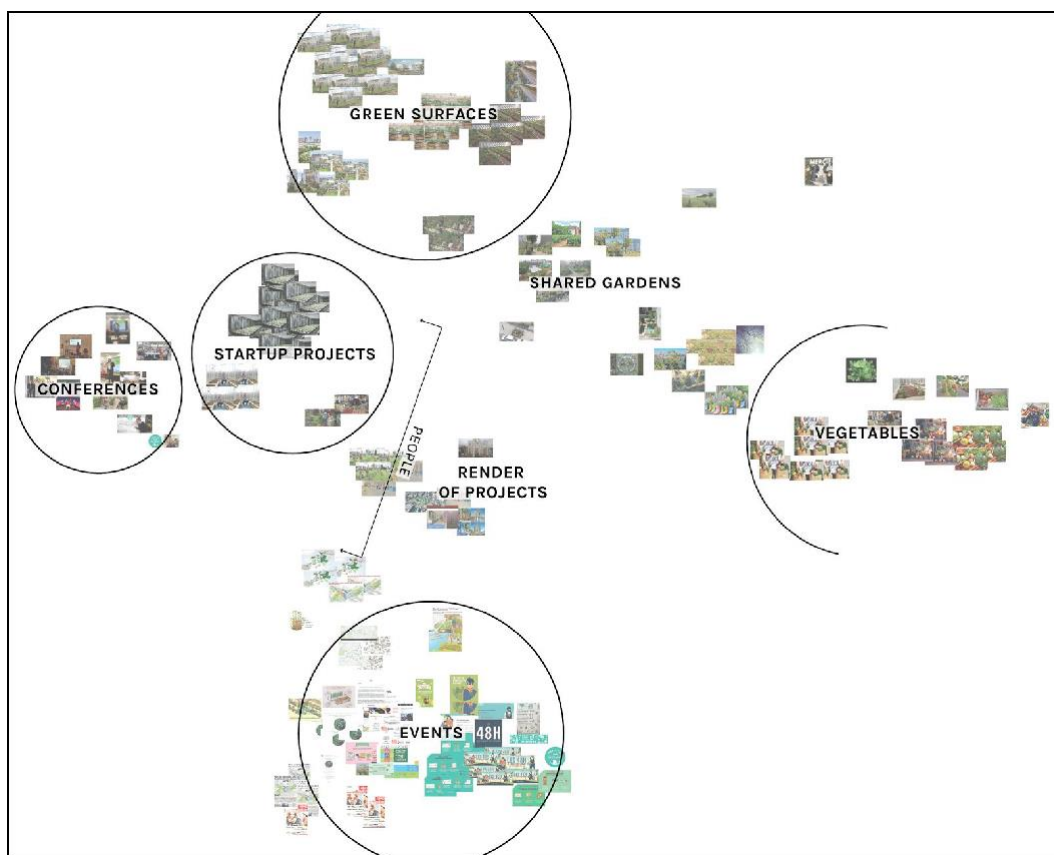


Figure 8. The visual sphere of the "Urban-agriculture" community.

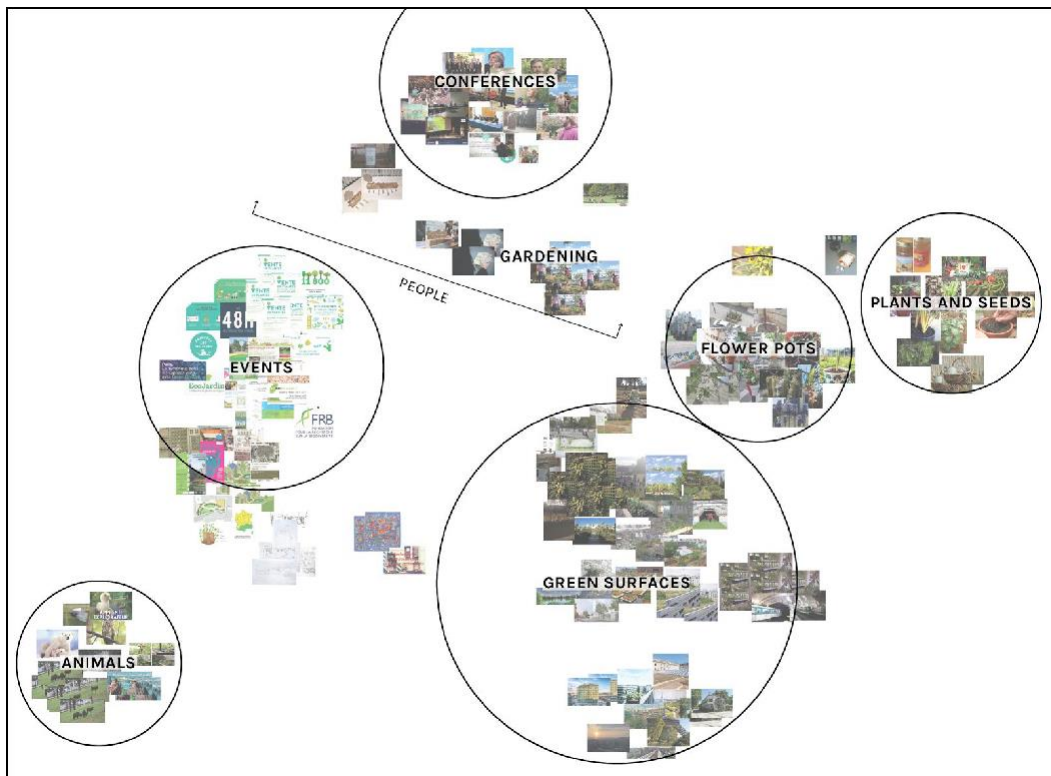


Figure 9. The visual sphere of the "Co-design of public nature" community

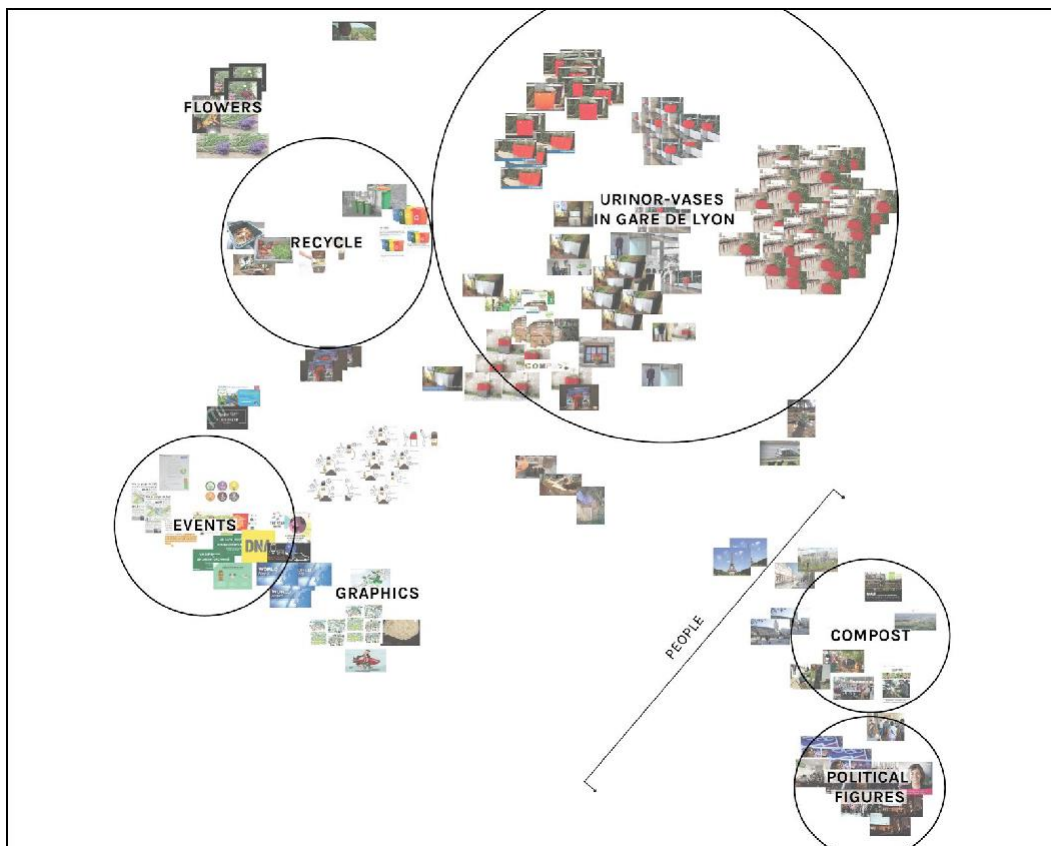


Figure 10. The visual sphere of the "Ecological attitude" community

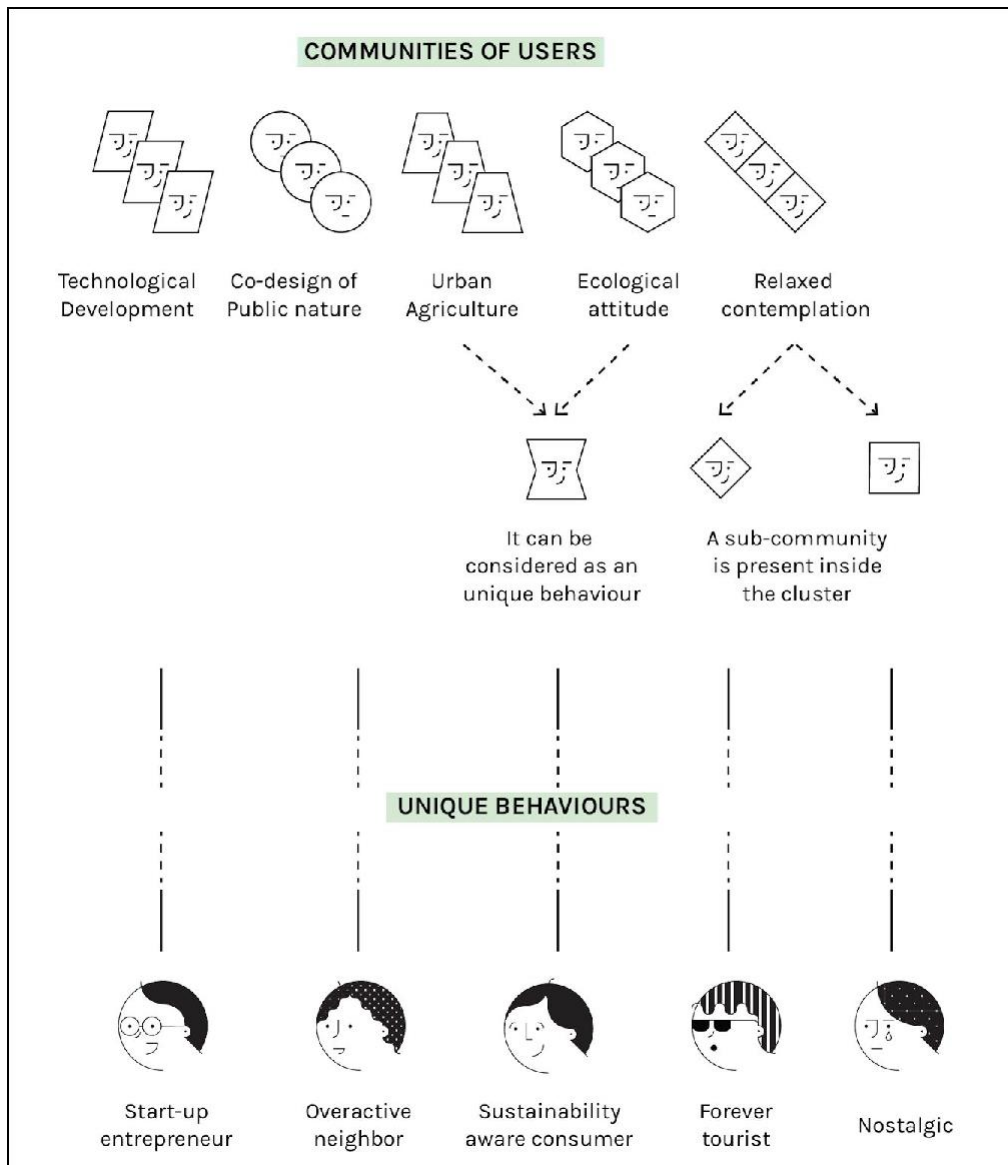


Figure 11. The visual sphere of the “Ecological attitude” community

5.5.3 Phase n. 4: Personas descriptions

Objective: Outline personas through the most relevant insights gathered during the process.

The last phase consists in moving from validated communities to personas and to build an engaging narration around those personas. Vivid descriptions of personas enable to bring fictional user profiles to life (Cooper 1998, Grudin & Pruitt 2003) and —as regards to Data-Driven Personas— this can be done by using the data generated by users during their online activities.

An advantage of the Data-Driven method is that researchers can build the profiles in a semi-automatic way by mapping users’ information on the various aspects of the persona profile. At this stage, data visualisation shifts its objectives from an analytical research tool to a way of synthesising the main aspects of each persona.

Inspired by the key elements that traditionally characterize the narration of a persona - such as a portrait photo, a relevant quote, a day-in-the life and key attributes - the descriptions generated by Data-driven personas can offer a similar narration, built through the type of data available in the digital observed sphere. The elements that compose the description of a Data-Driven Persona are

susceptible of variation, since they depend on the nature of the data collected, the digital platform used for the research and the insights gathered during the process.

In the test on NATURPRADI, each persona has been outlined with the data produced by the group of users from whom that personas was created. The narration starts with the picture and the name randomly picked from those of real users belonging to that persona. The keywords which initially brought to the definition of the community and then of the relative persona are listed as its most connoting hashtags. A tweet has been selected from the data corpus, in order to represent the usual way that persona would talk about urban nature in Paris. The Twitter descriptions of users are used to narrate how each persona would describe themselves: a bubble chart visualise the most occurred terms. The most engaging images of each persona are also part of the narration, showing their visual imagery. The relationship between personas, as well as their similarity, is represented by a diagram showing how many users are unique to the personas and how many are shared with other personas, since a user could be present in more than one cluster. Each persona can be also located along the "expertise axis", telling us if - in their relationship with nature - that persona is more a contemplator or an expert. Finally, the tweets activity over time of each persona can help understand the engagement with the topic (Figure 12-16).

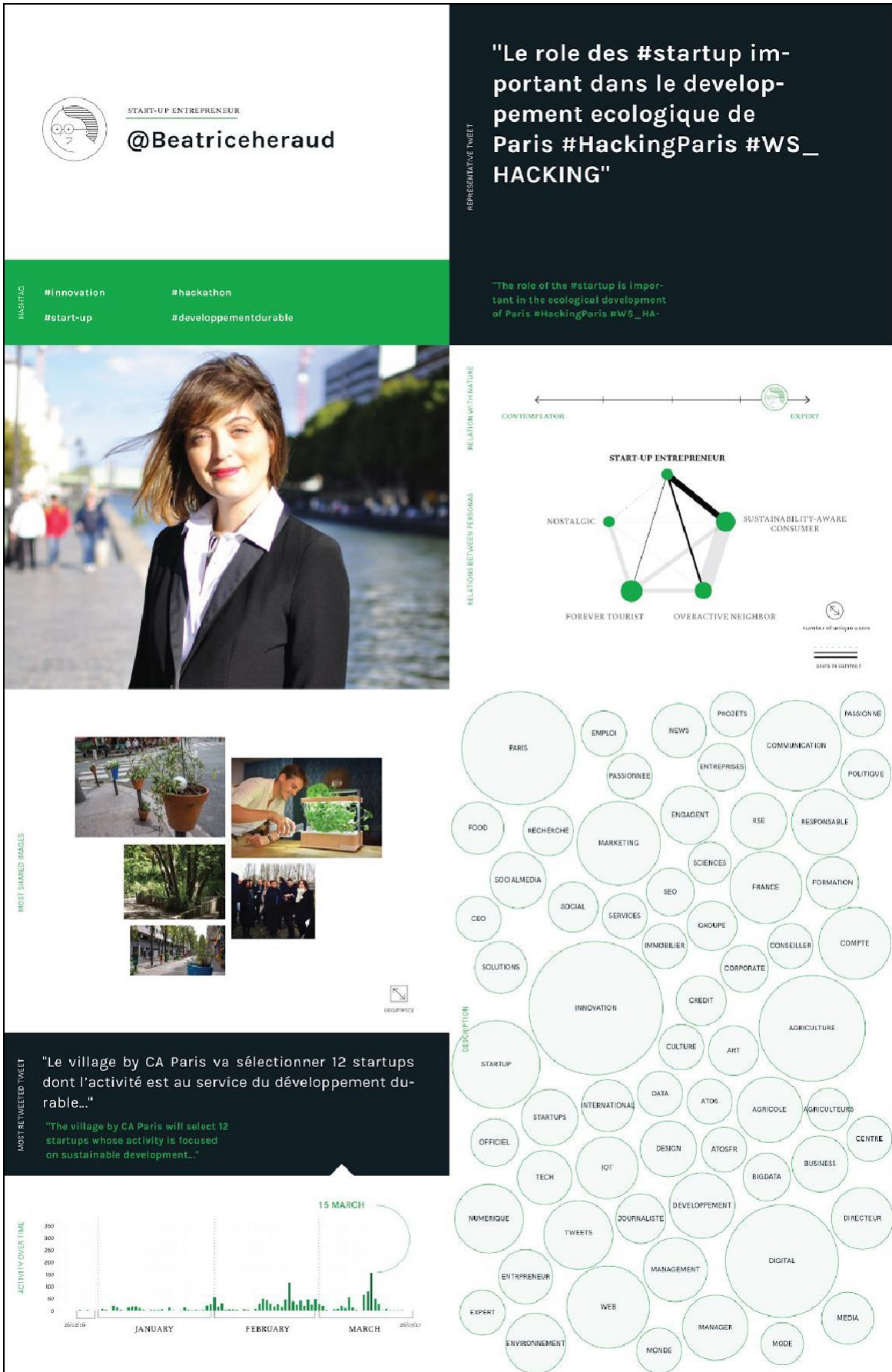


Figure 12. Description of the "Start-up entrepreneur" persona

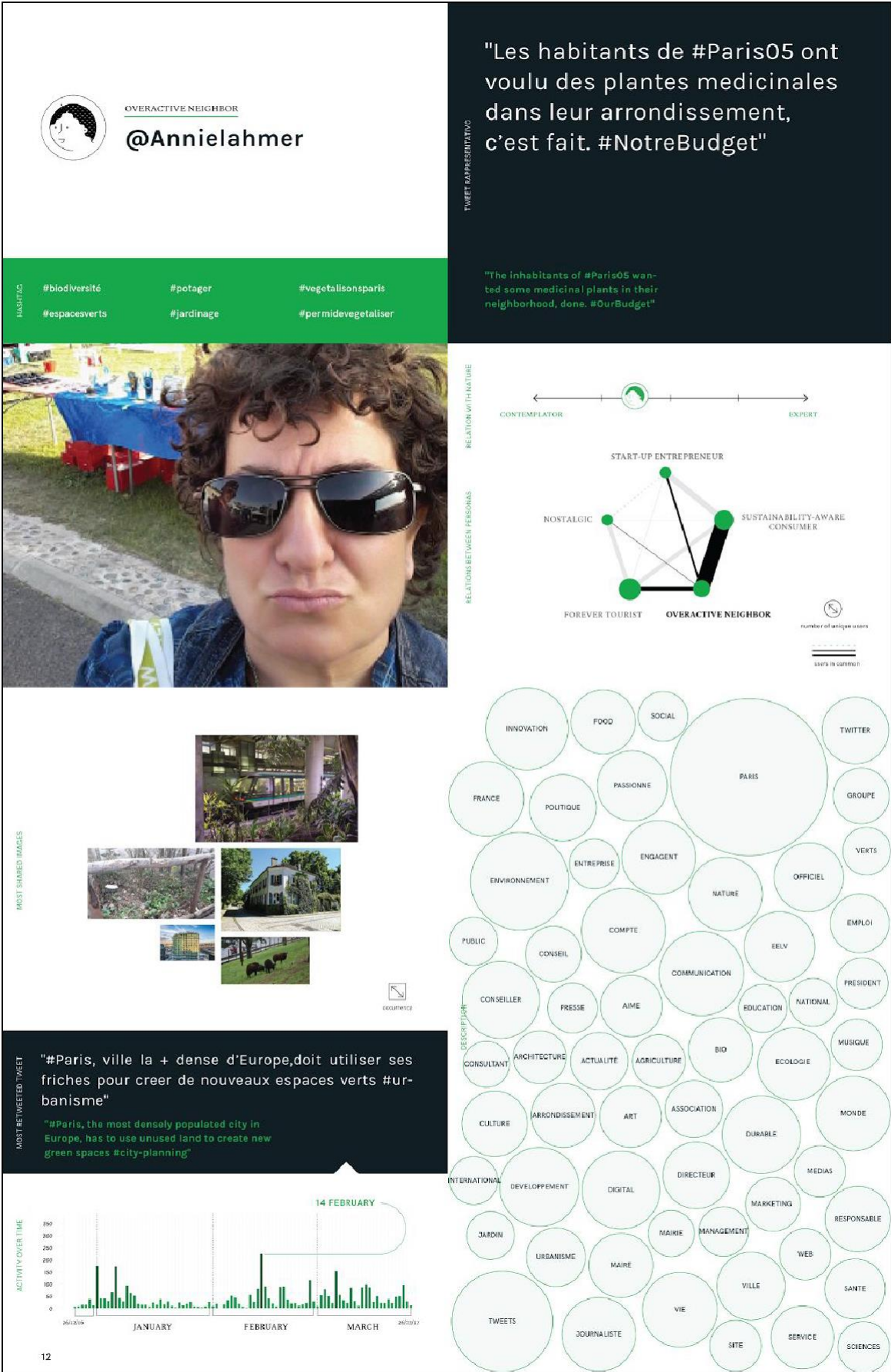


Figure 13. Description of the "Overactive neighbor" persona

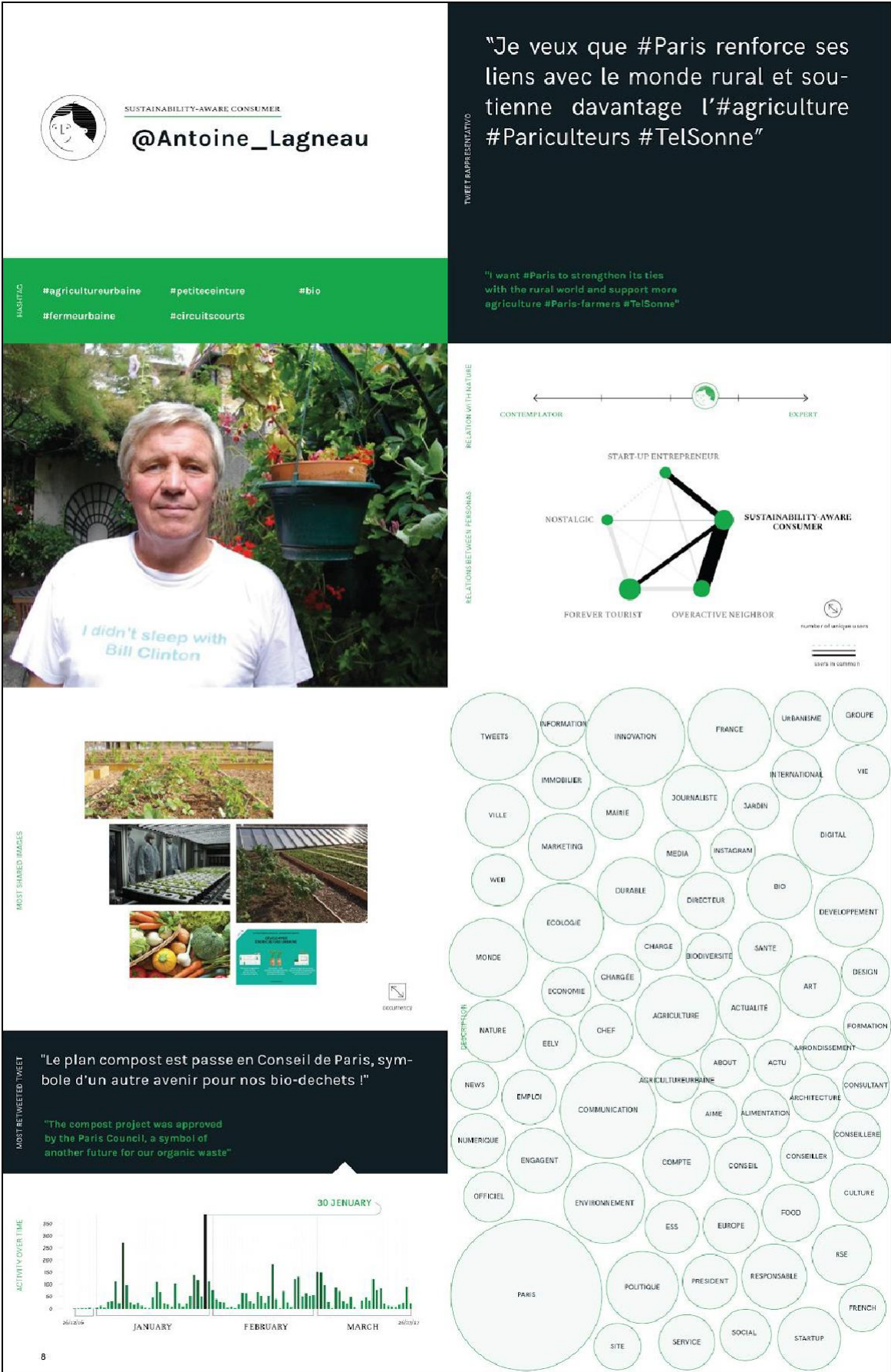


Figure 14. Description of the "Sustainability aware consumer" persona

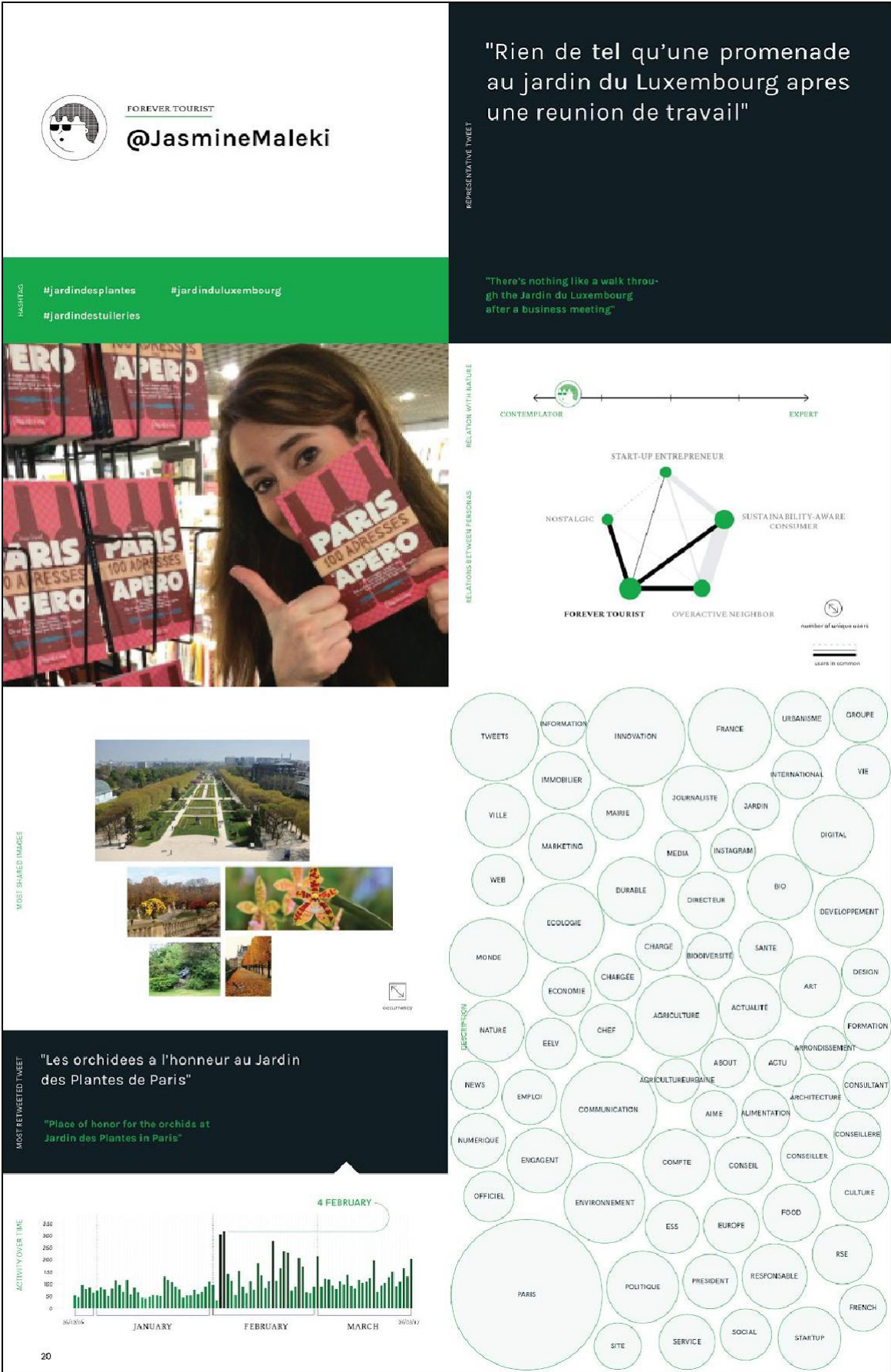


Figure 15. Description of the "Forever tourist" persona

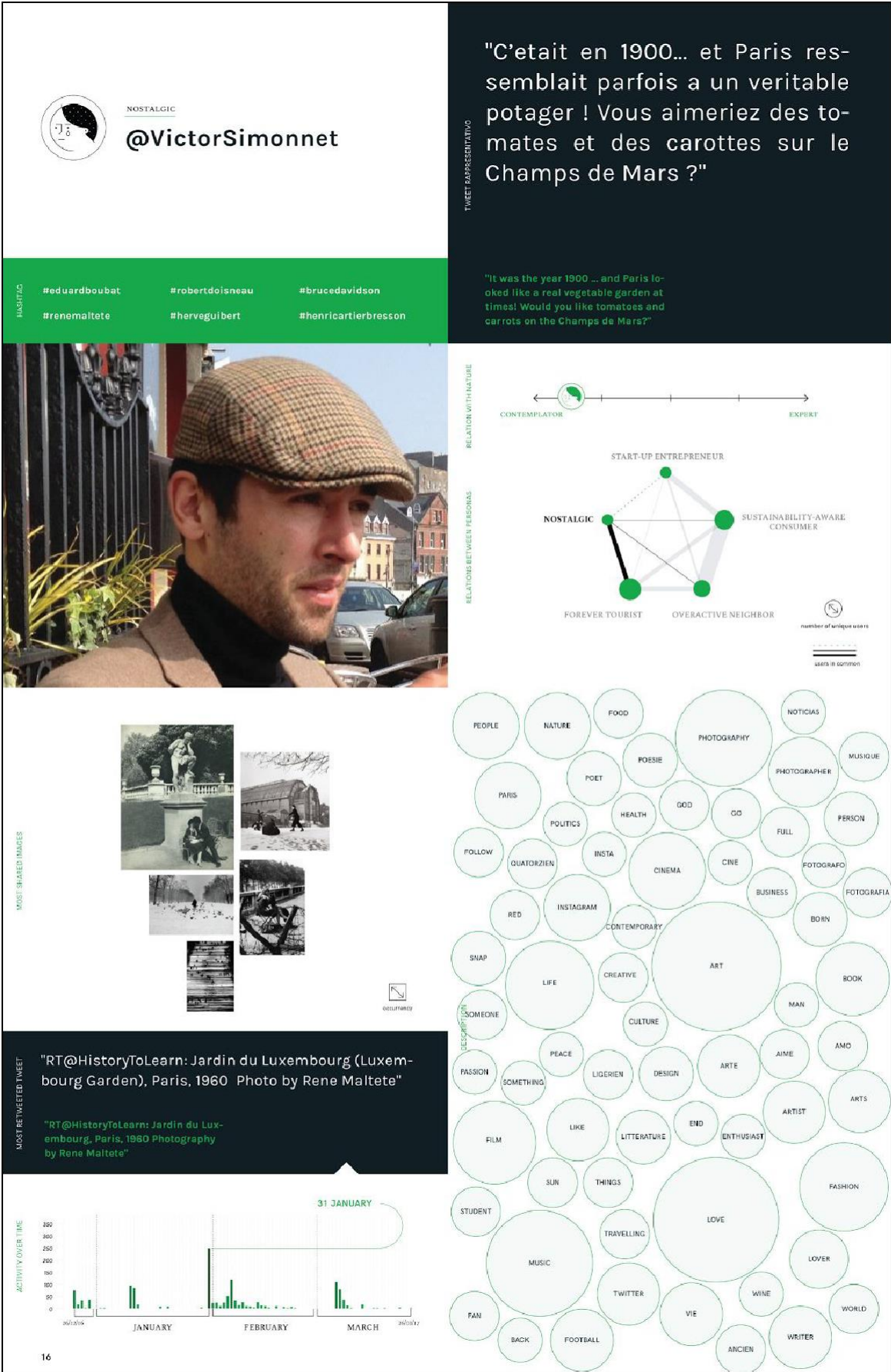


Figure 16. Description of the "Nostalgic" persona

6 Opportunities and Challenges

The new Data-Driven personas method aims to expose how Digital Methods can be integrated into Human-Centered Design, deploying new “*techniques which communicate, interact, empathize and stimulate the people involved*” (Giacomin, 2014).

Digital Methods could allow to scale up the magnitude of data and information collected. The proposed approach offers significant advantages in terms of time and costs, if compared to traditional qualitative research techniques: it allows to quickly collect and analyse a wide dataset and develop key insights even before activating the field-research and start investing on it.

Nevertheless, there are some activities in the process that shouldn't be underestimated, such as:

- **Data collection:** setting up the necessary infrastructure for collecting data might take some time. Depending on the scale of the data to be collected, simple solutions like storing it into spreadsheets or plain text files (i.e. CSV) might be not appropriate and the setup of a proper and efficient database might be required. Furthermore, the API provided by the digital platform, as well as the interface through which the data might be scraped, tend to change rapidly. This may affect the quality of the harvesting, or at least, require a continuous monitoring, tracking and adjustment of the collection procedures. Working with digital data means to respect the ever-changing privacy policies and terms of use of the platforms involved in the research. Along with the respect for these standards, an ethical reflection on how to handle personal identifiable information is always needed.
- **Data cleaning:** in some cases, sorting noise out of the stream of data collected can be done in a quick way (e.g. filtering out objects that are less frequently encountered, or conversely, the ones that are mentioned too much). In other cases, as for the NATURPRADI project, a careful reading of the collected data is necessary. Regardless the specific strategy adopted for reviewing and cleaning the dataset, a constant control of the data harvested is always necessary. While this operation helps the researcher explore the material they are going to work with, it also requires the setup of an appropriate infrastructure (i.e. from generating reports containing random samples of the data to *reading* the single data points one by one).
- **Data visualisation:** distilling information out of a dataset is more and more simple thanks to the growing numbers of techniques, libraries and software. Nevertheless, the ultimate scope of the visualization, exploratory visualisations and procedures are needed in order to continuously offer different views on the data through multiple and non-exclusive visual models, especially in the first part of the process. The production of interactive visualisation should be preferred to the static ones, to better support the exploration of the views and ease the identification of key insights and learnings.
- **Data interpretation:** what Digital Methods offers to Human-Centered Design is a better understanding of the user space and consequentially the reduction of the methodology concern raised by Chapman about the risk of projecting pre-existing categories, or missing unknown parts of the debate. (Chapman & Milham, 2006). The insights and clusters that emerge during this type of analysis need to consider possible limitations and bias. Collecting data over social media, for example, implies to cope with the digital divide issue and with the different *platform culture* (e.g. the more or less marked propension to use hashtags) that might be present in different geographical location. The outcome of this type of analysis needs to be seen as part of a wider range of research and design thinking activities aimed at understanding the people and context of use. The insights can be used for example to inject new hypothesis in a user research phase, or to enrich the existing knowledge with a different perspective.

Digital Methods need to be considered an opportunity in integration, and not in replacement, of current HCD tools and techniques. For example, along the process of creating Data-Driven Personas, the researchers may use the emerging clusters as a way to define potential participants for a set of in-depth interviews. The interviews will provide both a validation of the analyses carried out previously

and, above all, add a deeper qualitative layer to the understanding of the different personas. Following this example, Digital Methods can ease the preparation of a field-research, by raising important themes upfront and offering an alternative strategy to recruit research participants.

Another possibility for synergies is offered by the attempts to make the Human-Centered Design process more lean and agile, in order to cope with the challenges and speed that are typical of digital transformation processes. In order to streamline certain steps, the Lean UX approach has introduced, for example, the idea of proto-personas (Gothelf 2013). Proto-personas are user profiles developed as assumptions, by elaborating on the existing knowledge of a specific team or organization, prior to conducting an ethnographic study. Similarly, to Proto-personas, Data-Driven personas offer an assumption based on data that is already available in the digital space, and can be used as starting point to understand user behaviours and attitudes, before activating a field-study.

The Data-Driven Personas protocol is a first attempt to provide a sustainable and replicable approach to effectively apply Digital Methods to support the Human-Centered Design research process. This protocol is applicable to all those cases where the research environment involves a community of users who drive a series of debates inside an online space. Other protocols could be explored in the future, investigating the possibility to derive a broader set of user-centered frameworks from the analysis of the online discourse (e.g. experience journeys, system maps, future narratives).

We have argued that personas can be created on the basis of online information shared by users, using Digital Methods as a set of practices to gather this information in a renewed way for the design research. This first attempt requires additional experiments to pursue new ways of repurposing Digital Methods at different stages of the HCD process.

7 References

- Berry, D. M., Borra, E., Helmond, A., Plantin, J.-C., Walker Rettberg, J., & Walker, J. (2015). The data sprint approach: exploring the field of Digital Humanities through Amazon's application programming interface. *Digital Humanities Quarterly*, 9(4).
- Blank, G. (2016). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review*. <http://doi.org/10.1177/0894439316671698>
- Blyth, S., & Kimbell, L. (2011). *Design Thinking and the Big Society: From solving personal troubles to designing social problems An essay exploring what Design can offer those working on social problems and how it needs to change*. London.
- Broberg, O., Andersen, V., & Seim, R. (2011). Participatory ergonomics in design processes: The role of boundary objects. *Applied Ergonomics*, 42(3), 464–472. <http://doi.org/10.1016/J.APERGO.2010.09.006>
- Brown, T. (2008). Design Thinking. 2008. *Harvard Business Review*, 6.
- Brown, T. (2009). *Change by Design: How Design Thinking Transforms Organizations and Inspires*. *Innovation Harper Business*. New York, USA.
- Bruns, A., & Burgess, J. (2015). Twitter hashtags from ad hoc to calculated publics. In N. Rambukkana (Ed.), *Hashtag publics : the power and politics of discursive networks* (p. 293).
- Cabrero, D. G., Winschiers-Theophilus, H., & Abdelnour-Nocera, J. (2016). A Critique of Personas as representations of the other in Cross-Cultural Technology Design. In *Proceedings of the First African Conference on Human Computer Interaction* (pp. 149-154). ACM.
- Chapman, C. N., & Milham, R. P. (2006). The personas' new clothes: methodological and practical arguments against a popular method. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 5, pp. 634-636). Sage CA: Los Angeles, CA: SAGE Publications.
- Cooper, A., Reimann, R., Cronin, D., & Noessel, C. (2014). *About face: the essentials of interaction design*. Indianapolis: Wiley.
- Cooper, A. (2004). *The inmates are running the asylum:[Why high-tech products drive us crazy and how to restore the sanity]*. Indianapolis, IN: Sams.
- Cross, N. (2011). *Design thinking: Understanding how designers think and work*. Oxford: Berg.
- Ehn, P., Nilsson, E. M., & Topgaard, R. (2014). *Making futures : marginal notes on innovation, design, and democracy*. Cambridge, Mass.: The MIT Press.
- Floyd, I. R., Cameron Jones, M., & Twidale, M. B. (2008). Resolving incommensurable debates: a preliminary identification of persona kinds, attributes, and characteristics. *Artifact*, 2(1), 12-26.

- Gandy, M. (2002). *Concrete and Clay*. Cambridge, Mass.: The MIT Press.
- Gandy, M. (2006). Urban nature and the ecological imaginary. In N. Heynen, M. Kaika, & E. Swyngedouw (Eds.), *In the Nature of Cities: Urban political ecology and the politics of Urban Metabolism* (pp. 63–75). New York: Routledge.
- Giacomin, J. (2014). What Is Human Centred Design? *The Design Journal*, 17(4), 606–623.
<http://doi.org/10.2752/175630614X14056185480186>
- Gillespie, T. (2010). The politics of “platforms.” *New Media & Society*, 12(3), 347–364.
<http://doi.org/10.1177/1461444809342738>
- Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: essays on communication, materiality, and society*. Cambridge, Massachusetts: MIT Press.
- Gothelf, J., & Seiden, J. (2016). *Lean UX: Designing Great Products with Agile Teams*. O'Reilly Media, Inc.
- Hanington, B. (2003). Methods in the Making: A Perspective on the State of Human Research in Design. *Design Issues*, 19(4), 9–18. <http://doi.org/10.1162/074793603322545019>
- Hewson, C. (2003). Conducting research on the Internet. *The Psychologist*, 16(6).
- Ingold, T. (2015). Design Anthropology is not and cannot be Ethnography. In *Research Network for Design Anthropology*.
- Kimbell, L. (2015). *Title Applying Design Approaches to Policy Making: Discovering Policy Lab Type Report*. Brighton.
- Latour, B. (2007). La cartographie des controverses. *Technology Review*.
- Le Dantec, C. A. (2016). *Designing publics*. Cambridge, Mass.: The MIT Press. Macnaghten, P., & Urry, J. (1998). *Contested natures*. London: Sage.
- Manzini, E. (2016). Design Culture and Dialogic Design. *Design Issues*, 32(1), 52–59.
http://doi.org/10.1162/DESI_a_00364
- Marres, N., & Weltevrede, E. (2013). SCRAPING THE SOCIAL? *Journal of Cultural Economy*, 6(3), 313–335.
<http://doi.org/10.1080/17530350.2013.772070>
- Mauri, M. (2014). An interview with Richard Rogers: repurposing the web for social and cultural research. Retrieved November 6, 2017, from <http://www.densitydesign.org/2014/05/an-interview-with-richard-rogers-repurposing-the-web-for-social-and-cultural-research/>
- Mayr, P., & Weller, K. (2017). Think before you collect: Setting up a data collection approach for social media studies. In L. Sloan & A. Quan-Haase (Eds.), *The SAGE handbook of social media research methods* (p. 679). London.
- Muller, M. J. (2003). Participatory Design: The Third Space in HCI. In A. Sears & J. A. Jacko (Eds.), *Human-computer interaction. Development process* (pp. 165–185). Boca Raton: CRC Press.
- Norman, D. A., & Draper, S. W. (1986). *User centered system design. New Perspectives on Human-Computer Interaction. New Perspectives on Human-*. Hillsdale: L. Erlbaum Associates.
- Pruitt, J., & Grudin, J. (2003, June). Personas: practice and theory. In *Proceedings of the 2003 conference on Designing for user experiences* (pp. 1-15). ACM.
- Ricci, D. (2010). Seeing what they are saying: Diagrams for socio-technical controversies. In D. Durling, R. Bousbaci, L.-L. Chen, P. Gauthier, T. Poldma, S. Rowoth-Stokes, & E. Stolterman (Eds.), *DRS2010 - Design & Complexity proceedings*. Montreal.
- Ricci, D., Colombo, G., Meunier, A., & Brilli, A. (2017). Designing Digital Methods to monitor and inform Urban Policy. The case of Paris and its Urban Nature initiative. In *3rd International Conference on Public Policy (ICPP3)*.
- Rogers, R. (2009). *The End of the Virtual: Digital Methods*. Amsterdam: Amsterdam University Press.
- Rogers, R. (2013a). Debanalizing Twitter. In *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13* (pp. 356–365). New York, New York, USA: ACM Press. <http://doi.org/10.1145/2464464.2464511>
- Rogers, R. (2013b). *Digital Methods*. Cambridge Mass.: The MIT Press.
- Schneider, S. M., & Foot, K. A. (2004). Web Sphere Analysis: An Approach to Studying Online Action
 Forthcoming in *Virtual Methods: Issues in. In Social Science Research on the Internet*. Oxford: Berg.
- Seemann, J. (2012). Hybrid Insights : Where the Quantitative Meets the Qualitative. *Rotman Magazine*, 58–61.
- Steen, M. (2012). Human-centered design as a fragile encounter. *Design Issues*, 28(1), 72-80.
- The Design Collaborative. (2014). DesignX: A Future Path for Design - [jnd.org](http://www.jnd.org). Retrieved October 1, 2016, from http://www.jnd.org/dn.mss/designx_a_future_pa.html
- Turner, P., & Turner, S. (2011). Is stereotyping inevitable when designing with personas?. *Design studies*, 32(1), 30-44.
- VanDijck, J. (2013). *The culture of connectivity: a critical history of social media*. Oxford: Oxford University Press.

- Venturi, G., Troost, J., & Jokela, T. (2006). People, Organizations, and Processes: An Inquiry into the Adoption of User-Centered Design in Industry. *International Journal of Human-Computer Interaction*, 2(21), 219–238.
- Venturini, T. (2009). Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science*, 0963662509102694v1.
- Venturini, T., & Latour, B. (2009). The Social Fabric: Digital Traces and Quali-quantitative Methods. In *Proceedings of Future En Seine*. Paris.
- Venturini, T., Munk, A., & Meunier, A. (2016). Data-Sprinting: a Public Approach to Digital Research. In C. Lury, P. Clough, M. Michael, R. Fensham, S. Lammes, A. Last, & E. Uprichard (Eds.), *Interdisciplinary Research Methods*.
- Venturini, T., Ricci, D., Mauri, M., Kimbell, L., & Meunier, A. (2015). Designing Controversies and their Publics. *Design Issues*, 31(3), 74–87. http://doi.org/10.1162/DESI_a_00340

About the Authors:

Donato Ricci is researcher using Design Methods in Human and Social Sciences at Sciences Po, médialab. With Bruno Latour, he co-curated the *Reset Modernity!* exhibition. He teaches *Knowledge and Representation* at Universidade de Aveiro and he is part of the *SPEAP* programs at Sciences Po.

Agata Brilli is an information designer. She collaborates with research groups including SciencesPo, médialab and the Digital Methods Initiative at UVA. She combines Human Centered design tools, digital methods and data visualization.

Roberta Tassi is a researcher and service designer, expert in combining ethnography design methods, participatory design and information visualization in complex design challenges.