Defining Beneficiaries of Emerging Data Infrastructures Towards Effective Data Appropriation

Insights from the Swedish Space Data Lab

Aya Rizk[1](✉), Cathrine Seidelin[2], György Kovács[1], Marcus Liwicki[1], and Rickard Brännvall[3]

[1] Luleå University of Technology, 97187 Luleå, Sweden
aya.rizk@ltu.se
[2] University of Copenhagen, 2100 Copenhagen, Denmark
[3] RISE Research Institutes of Sweden, 97347 Luleå, Sweden

Abstract. The increasing collection and usage of data and data analytics has prompted development of Data Labs. These labs are (ideally) a way for multiple beneficiaries to make use of the same data in ways that are value-generating for all. However, establishing data labs requires the mobilization of various infrastructural elements, such as beneficiaries, offerings and needed analytics talent, all of which are ambiguous and uncertain. The aim of this paper is to examine how such beneficiaries can be identified and understood for the nascent Swedish space data lab. The paper reports on the development of persona descriptions that aim to support and represent the needs of key beneficiaries of earth observation data. Our main results include three thorough persona descriptions that represent the lab's respective beneficiaries and their distinct characteristics. We discuss the implications of the personas on addressing the infrastructural challenges, as well as the lab's design. We conclude that personas provide emerging data labs with relatively stable beneficiary archetypes that supports the further development of the other infrastructure components. More research is needed to better understand how these persona descriptions may evolve, as well as how they may influence the continuous development process of the space data lab.

Keywords: Data infrastructure · Data lab · Beneficiary · Data appropriation · Persona

1 Introduction

One vision that has motivated the early calls for open data efforts and mandates was that making data accessible for all will enable data-driven innovation. Scholarly research has investigated this vision and the means by which the openness drives innovation in the public sector e.g. [1]. Since then, various governments and organizations in the public sector are doing efforts of digitalization, data collection and standardization for open data [2]. However, these are very complex and challenging tasks. Therefore, the efforts done

A. Lopata et al. (Eds.): ICIST 2021, CCIS 1486, pp. 32–47, 2021.
https://doi.org/10.1007/978-3-030-88304-1_3

to build the required data infrastructures, and corresponding data capabilities, towards the realization of data-driven innovation and evidence-based policy making have benefits that are yet to be realized [3]. Organizations, both public and private, are also struggling to identify and establish the data culture necessary in realizing those objectives, from commitment to data-based decision making, to the improvement of their data analytics processes and capabilities [4].

Data labs (also known as data hub, data factory, or policy lab) are currently being established as a result to the increasing amount of these open data initiatives. Data labs are initiatives that aim to help governments and other organizations make their collected data easily accessible and usable for evaluation, research and innovation [3]. In the United States, more than six data labs have been established over the past decade [3]. This paper focuses on a data lab initiative, which is situated in Scandinavia. In Sweden, the Swedish Innovation Agency is funding data labs and data factories to act as national resources in specific domain areas [5]. The objectives of funding theses labs range from lowering the barriers to data use to the use of data analytics, machine learning and Artificial Intelligence (AI) in to solve problems in said domains (ibid).

In order for data labs to be useful and value generating, it is crucial for these institutions to ensure that the data products and services they offer are relevant and appropriate for its intended beneficiaries. Beneficiaries are decision makers in organizations who decide on, and expected to benefit from, the adoption of data products/services offered by the data lab. Their buy-in is considered crucial for any data lab's establishment and survival [3]. Given the nascence of these initiatives, very little is currently known about how to identify and understand those beneficiaries so that this form of appropriation can be done [6]. In addition, in the early stages of establishment of data labs, there may be some ambiguity around key elements such as the core offerings (i.e. data product/service) and the talent required to deliver them [3]. With those ambiguous elements, the appropriation of data to the beneficiaries becomes even more challenging. Accordingly, we address the following research question: *How can the necessary appropriation of data for beneficiaries of data labs take place when the beneficiaries are undefined?*

We propose that developing personas can help us reach a preliminary understanding of beneficiaries and their needs. Hence, we also address the sub-question: *How can personas help identify and define beneficiaries of emerging data labs?* To address these questions, we designed a qualitative study in which we use the persona method [7, 8] in order to elucidate the beneficiaries of the Swedish Space Data Lab (SDL), as well as their characteristics as data infrastructure beneficiaries. Thus, the paper starts with reviewing related work on data lab initiatives, infrastructural challenges relevant to data labs and the status quo of earth observation data in Sect. 2. Section 3 describes the research method, followed by the findings in terms of persona variables and descriptions in Sect. 4. Finally, we discuss the implications of the identified personas on the SDL, as well as on the infrastructural challenges pertaining to data labs, before concluding in Sect. 5.

2 Background

2.1 Data Lab Initiatives

Data Labs are emerging in the public sector due to the need for data infrastructures that can support the technological development and its many possibilities for data-driven innovation [2]. The last decade's increase in wholesale digitisation has brought more citizens and organisations to make use of digital ways to connect with the public sector and vice versa (e.g., the healthcare system, their local municipality etc.). These new digital platforms have led to a vast collection of data on citizens' and organisations' preferences and behaviours, which are increasingly becoming more trackable and possible to access. Simultaneously, the developments in analytical techniques have created possibilities for understanding this data and applying it for different purposes and to varying actors [9]. However, despite the many promising possibilities for data analytics, governments are struggling with developing the infrastructures necessary for multiple stakeholders with varying interests to make use of relevant public data. These challenges pertain to, in part, the lack of understanding on how data drives development and innovation, as well as its associated mechanisms when such actors are from heterogenous knowledge backgrounds [10, 11]. Data Labs are emerging as a means to address this issue [3].

The Swedish Innovation Agency, who finances and supports national data labs, envisions data labs as national resources for their respective domains; a resource through which the domain's data is available, developed and used (e.g. for AI), and where different actors are encouraged and supported to use it for data-driven innovation [5]. A common goal of a data lab is "to improve public services using evidence-based decision-making which, in many cases, leveraged [administrative] data." [3, p. 7]. Various data labs are already established and have been serving their respective beneficiaries for years. California Policy Lab (CPL) provides analytics services that help different beneficiaries - at city, county and state level - evaluate and improve their public programs. UK-based Justice Data Lab helps organizations that work with offenders, gain access and make use of re-offending data, so that they can optimize their rehabilitation programs, predict and avoid re-offending. Even though these data lab initiatives have been ongoing for a few years, the scholarly literature is extremely scarce on them. A clear exception is Lyon et al.'s [12] evaluation of the Justice Data Lab.

Furthermore, more data labs are in the process of being established. At the time of writing this paper, the Swedish Innovation Agency has funded 16 data labs [5]. The Swedish Language Data Lab aims to develop the Swedish language reference dataset and models for Natural Language Processing (NLP). On the other hand, the Ocean Data Factory aims to "enable Sweden to be a global leader in sustainability and innovation in the digital blue economy" through the applications of AI in areas such as shipping and logistics, emissions both to air and water, and climate change [13].

2.2 Lessons Learned from Existing Labs

Existing data labs differ in their models around lab ownership, who the beneficiaries are, what services are offered, and the source of their analytic talent. Each dimension has its tensions. For instance, the decision of lab ownership has implications on its governance, data sharing agreements, the degree of buy-in required from various beneficiaries, and what projects, services or problems to pursue [3]. Furthermore, the relevant talent, that is able to generate insights and/or actionable recommendations from the analytical results, moderates these aspects and is often challenging to find. The CPL director notes the conundrum of data labs as follows: "Often the people most interested in research, do not have the relevant data, and the people charged with stewarding the data do not have the resources to pursue research." (Evan White, cited in Dinesh, 2017, p. 8).

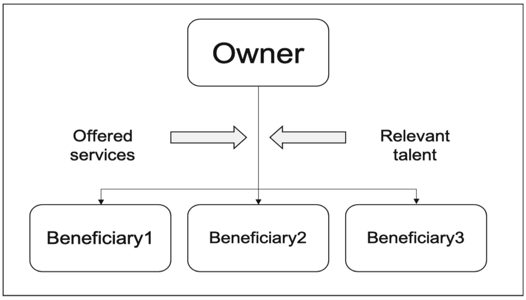


Fig. 1. Elements of a data lab

Accordingly, the four key elements (seen in Fig. 1) seem to be interdependent when defining and establishing a data lab. While the literature provides insights on identifying and the incentives to attract the relevant talent and build an appropriate data science culture [4], much less guidance is offered on how to define prospective beneficiaries and their needs. The current literature also assumes that the other three elements are fixed and known. However, if the beneficiaries are diverse, their needs and competences are unclear, it becomes problematic to define the other elements.

2.3 Infrastructural Challenges of Data Labs

The ability to share and reuse data can enable innovation [1]. However, it is a challenge to understand how data should be made available when the coordination and social practices pertaining to the data lab and its intended beneficiaries are at a very nascent stage. As a way to address this key infrastructural challenge, we include related work from the eScience and cyberinfrastructure streams to help us think about the emergent SDL. The eScience community is well-known for design and development of software pipelines that support data production, processing, and analysis within and across different scientific communities [14].

Previous research has investigated the development and use of such software pipelines and associated cyberinfrastructures, in particular with a focus on the organisational work it takes to create and maintain such infrastructures [15, 16]. Star and

Ruhleder [16] propose that infrastructure is inherently a substance for relations (i.e. between things and people) rather than a substrate. For example, Bietz et al. [15] look at how different stakeholders maintain the development of cyberinfrastructure. In this case, they explore how developers work to preserve a path that ensures continuous development of the infrastructure. The study shows that the developers do so by adjusting their own work to certain domain science projects and funding streams [15]. With the rise of the big data phenomenon, Demchenko et al. [17] proposed that we start considering the "data infrastructures" as distinct from the cyberinfrastructure due to their different physical and logical representations. Moreover, Lee et al. [18] propose the notion of "human infrastructure" which refers to different forms of organizing among stakeholders that overlap, might change over time, and might be used simultaneously to create and maintain a cyberinfrastructure. They propose several practical implications that should be considered in the context of infrastructure development. For instance, they argue it is essential to reduce the ambiguity and to embrace fluid organisational structures. Finally, the blurred perception of group membership among distinct stakeholders demands efforts that help to ensure broad participation from multiple stakeholders [18]. In this study, we aim to reduce such ambiguity and capture the stakeholders' perceptions in order to facilitate the appropriation of data in the SDL.

2.4 Earth Observation (EO)

The lab in focus of this study, the SDL, is set up with the objective to make EO data more accessible and usable for innovation, policymaking, and monitoring in Sweden. This data is obtained through the Copernicus program. The program launched a family of satellite missions, namely Sentinel-1 to Sentinel-6 [19], each containing two satellites in order to fulfil the revisit and coverage requirements. Data from the Sentinel missions are available in five cloud-based platforms, as well as in various download services. To be able to efficiently exploit this data, the SDL uses tools from the Open Data Cube (ODC) project [20]. ODC is an open source data management and analysis software project originally based on the Australian Geoscience Data Cube [21], designed to organise vast amounts of EO data into an efficient database structure. For the purposes of the SDL project, the Sentinel-2 data of Sweden from years 2017 and 2018 was batch-downloaded, and new data is continuously downloaded during the operation of the project, and as needed by the individual pilots.

Working with EO provides another dimension to this study, bringing the practical component of the research problem. It is not uncommon for domain experts who are not working with EO data to face challenges in making significant use of it for public benefit [6]. During a Copernicus networking event in 2019, various domain experts (e.g., on climate adaptation, forestry, etc.) expressed their interest to the Swedish space agency - one of the SDL consortium partners - in using the SDL data. However, it was unclear how and why they want to use it.

Taken together, these different perspectives highlight the challenges of emergent data infrastructures in terms of its ambiguity and fluidity of its actors and relations. The SDL can be viewed as an emergent infrastructure, and thus it becomes relevant to regard the above-mentioned challenges and recommendations, in particular, that there is no "one size fits all" solution. It is therefore important that the SDL takes into account the varying (data) needs of multiple different beneficiaries throughout the design process. We have created persona descriptions as a design tool to support the design team in creating a data lab that addresses the needs of various beneficiaries.

3 Research Method

The appropriation of data for the beneficiaries entails understanding the meanings they assign to data, their current mental models, their usage needs, and their expectations. As a way to develop a joint understanding of SDL's beneficiaries, we chose a qualitative approach to develop Personas, as this method is particularly useful for generating recognisable and distinct fictional character of typical user groups [22]. In this section, we first give a brief account on the Persona method, followed by a description of the SDL, which constitutes the research setting. Finally, we elaborate on our development of persona descriptions for this context.

3.1 The Persona Method

Developing a persona is a user-centred design method commonly used to understand user characteristics, needs, and goals in order to condense valuable insights that can inform a design process. The method emerged in the late 1990s as an attempt to communicate an understanding of the users to support IT system development [8, 23]. Personas constitutes abstract user representations, or character archetypes, which are most often developed by designers based on user data [8, 22].

> "*Personas are an efficient design tool because of our cognitive ability to use fragmented and incomplete knowledge to create a complete vision about the people who surround us. With personas, this ability comes into play in the design process, and the advantage is a greater sense of involvement and understanding of reality"* [8, p. 24].

Nielsen et al. [24] propose that the advantages of personas can be divided in three categories: mental models, data storage, and prioritization. The first category refers to how personas can support the creation of joint mental models, which can help to challenge assumptions, both on individual and organisational levels in design processes [25]. The second category emphasises benefits of how personas (and the information they hold) often are easy to access and communicate. Finally, the third category encapsulates personas' ability to prioritize audiences. Thus, as a design tool, personas aim to represent current and/or potential beneficiaries and users. This is highly relevant in our case.

3.2 The SDL Context

This study is conducted as part of the SDL project. The project was launched in June 2019 for two years. It aims to make EO data accessible and usable for public benefit and innovation. Representatives from four organizations make up the project consortium: the Swedish space agency, a university, a research institute and an NGO. When the project was initiated, it was a major challenge to pinpoint for whom the lab is developed. The project plan included five pilot projects with five different beneficiaries, however, it was not clear who these beneficiaries are and what their needs are. Therefore, the design decisions were informed as the project and pilots were ongoing, and the personas method facilitated such discussions.

3.3 Developing SDL Personas

Overall, we followed Goodwin's [7] guidelines to develop our personas. Eight individual in-depth semi-structured interviews and one workshop comprised our primary data collection. The interviewees were selected from current and potential pilot project leaders (beneficiaries) and each interview lasted between 45 and 70 min. An interview protocol was designed and guided the interviewers. Seven interviews were conducted in English and one in Swedish, albeit with English notes. All interviews were recorded. The workshop was conducted with five other prospective beneficiaries, in Swedish, and coordinated by one of the project researchers. The workshop had two sections: one where participants described their pain points when working with EO data, and the other where they were asked to describe a typical user for the SDL. The workshop was documented in notes organized by those sections.

The interview data analysis was conducted in four phases. First, two interviews were listened to and analysed to extract the first set of persona variables. The analysis was largely inductive where the interviews were subject to thematic coding [26]. This yielded 9 continuous, 5 discrete and 3 demographic variables. Two of the researchers then read all the notes and met to discuss the variables. This resulted in the addition of 4 continuous, 2 discrete and the omission of one demographic variable. Second, all interviewees were mapped along the variables by two researchers, followed by the analysis of visible patterns. This process yielded 3 clear personas. During the mapping process, use case-specific missing data led to the exclusion of the following discrete variables (and including them with demographics in an ongoing questionnaire study complementing this one): user homogeneity and number of simultaneous users. Third, a description of each persona was written down in terms of behavioural descriptions and persona goals. The workshop notes provided triangulation and richness to the persona descriptions. Each description was written by one researcher and cross-checked by another, going back to the notes for consistency. Fourth, the personas were written in a résumé format (see example in the Appendix).

4 Findings

The variables identified from the analysis could be organized in different ways. The first categorization was structured according to whether they were discrete or continuous.

Then, they were also categorized based on whether they described the persona as an individual, their attitude in relation to a specific use case or scenario, or in relation to their organization. In this section, we report them along those two dimensions and describe the personas accordingly.

4.1 Persona Variables

These persona variables were identified and refined with the objective of pinpointing the beneficiaries' priorities and expectations engaging with space data and the SDL. Accordingly, rather than focusing on the demographics of individuals, we focused on their data needs, especially when it comes to spatial data and their (potential) use cases. Throughout the study, their goals from space data and current pain points were of main concern. Table 1 below summarizes the variables in focus that helped us develop the persona descriptions further.

Table 1. Persona variables.

	Continuous	Discrete
Individual	General technical proficiency GIS proficiency Attitude towards collaboration	Nature of role Interest in space data
Use case	Specificity of use case Temporal/feature coverage Temporal/spatial coverage Spatial/feature coverage Financial resources Dependency on SDL	Motivation for collaboration Anticipated benefits from SDL (with regards to domain, and data/information/knowledge) Most valuable aspect of SDL
Organization	Organizational/technical scope Level of outward data sharing Level of inward data sharing	Data sources Types of analyses

4.2 Persona Characteristics

Three clear personas were found among our beneficiaries. Table 2 shows the complete list of different personas characteristics, and a summary is provided thereafter to highlight the similarities and differences across the personas.

Table 2. Characteristics of beneficiary personas in the SDL.

	AB: the coordinator	CG: the consultant	ED: the GIO
Technical proficiency (remote sensing)	Non-technical	Technical	Technical
GIS proficiency	Basic to moderate	Advanced	Advanced
Role	Coordination	Operational	Executive
Interest	Brings opportunities - explorative	Neutral - business necessity	Must use it - legal compliance
Collaboration	Neutral	Follower	Leader
Use case specificity	Ambiguous	Clear	Clear
Coverage priorities	1. Temporal 2. Spatial 3. Features	Exhaustive	1. Spatial 2. Features 3. Temporal
Financial resources	Needs external financing	Full financing possible	Needs partial financing
Dependency on SDL	Direct service needed	Minimum support needed	Moderate support needed
Motivation	Exploration & innovation	Strategic business development & access to infrastructure	Legal compliance & operational efficiency
Anticipated benefits (domain)	Monitoring & responsiveness	Cost efficiency & service quality	Cost efficiency & service quality
Anticipated benefits (DIK)	Outsourcing of skills	Data quality & wider coverage	Accessibility (e.g. through SLAs)
Most valuable aspect of SDL	Data products & competence	Tools & infrastructure	Data products
Organizational scope	Geography- and domain-specific	Technology-specific	Domain-specific
Data sharing	Outward conditional Limited inward	Inward only	Outward Limited inward
Data sources	"GUI Maps" - readily computed models	LiDAR Radar National land survey Ground measurements Questionnaires Aerial photography Raster & vector	LiDAR Aerial photography Satellite imagery DEMs Land surveying Vector models

(*continued*)

Table 2. (*continued*)

	AB: the coordinator	CG: the consultant	ED: the GIO
Analyses	Temporal analysis Forecasting Change detection	Pre-processing atmospheric corrections Time series Fixing geometry	Pre-processing Fixing geometry Raster functions Overlays

The first persona (AB) assumes a coordination role in a county administrative board, mostly focused on specific phenomena (e.g., climate adaptation, flooding, etc.). Thus, they are primarily interested in their county's geographical area when it comes to data coverage. They also work with monitoring and preparedness and are interested in change detection and long-term temporal analysis for their respective phenomena. They often have elements of critical infrastructure within their scope; hence, they believe the data should not be entirely availed publicly. They have the least technical skills of all three personas, manifested in their ability to only work with readily computed "maps" accessible via a friendly user interface.

The second persona (CG) is a consultant and business associate who identifies as a "space actor", meaning they are familiar with the space vertical from satellite technologies to their end-user needs. They are focused on technology and have the most advanced technical (both on the programming and GIS) skills of the three personas. Their interest in the data coverage is paramount - the more data the better. The interest in the SDL and motivation for collaboration is strategic and regard the technical infrastructure as the SDL's main advantage.

The third persona (ED) assumes the role of a Geographic Information Officer (GIO), who is responsible for the acquisition, management and dissemination of geographic data for their agency. They also possess advanced technical GIS skills. The agency they work for oversees the use and management of natural resources all over the country, owned both by private and public entities. Thus, the GIO's interest in data coverage is wider than that of AB, but not as exhaustive as CG since they are interested in specific features. This also means that the areas they cover may be subject to some degree of confidentiality (for private owners) or protection (for certain protected areas).

5 Discussion and Conclusion

In this section, we discuss the insights which were generated through the development of the three personas and the usage of the descriptions as a tool to facilitate the establishment of the SDL, how the infrastructural challenge with its emergence is addressed, and implications on the SDL design of offered products and services.

First, the personas have elucidated the difference between beneficiaries, who are experts in specific technologies (e.g., remote sensing and space technologies) and application domains (e.g., forestry, climate change and sustainability, etc.) in their mental models, expectations and data needs, and motivations for collaboration. While the former group define themselves to be "space actors", the latter are relatively technology agnostic. EO data and the SDL provide a common language by which those different actors could communicate and a bridge between technology and domain. This presents yet another relation in the space of relations complementing Star & Ruhleder's [16] view of infrastructure. However, it also reveals that EO data users (and resulting personas) represent a continuum, rather than user/non-user distinction adopted implicitly in the literature [6]. It is difficult to draw the line between users and non-users, since data representations, granularities and media evolve from one user group to another, and from one infrastructure to another (i.e., from Copernicus, to ODC to the SDL).

Second, in the absence of fixed organizational structures and practices, ambiguity of products and services, and uncertainty of relevant analytical talent, the personas enabled the SDL to focus on relatively stable archetypes of beneficiaries as a starting point to address the other elements. In that way, the focus on the human infrastructure when the former elements are ambiguous or changing supports the development of the technical infrastructure [18], as well as the data infrastructure [17].

Third, the direct practical implications of the personas lie in the design decision taken by the SDL to shape the data lab's offerings. In the proposal to Vinnova, the SDL described the aim of offering access to the ODC and tools for analysis and visualization. These components were primarily designed for programmers. However, the design evolved to include three environments through which the SDL can be accessed and used, taking into account the skills and possible access to talent available to each beneficiary: a) analysis lab that enables Python programmers take full advantage of machine learning techniques and methods (most relevant to persona CG), b) a GUI that delivers readily computed data models (most relevant to AB), and c) integration services to allow connecting EO data with other software solutions (e.g. GIS) - most relevant to ED. The three environments are illustrated in Fig. 2 below.

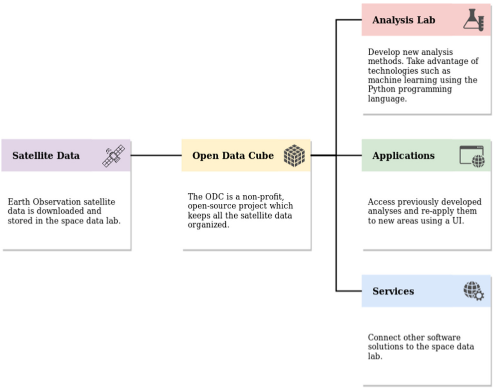


Fig. 2. SDL infrastructural design.

Another key implication is privacy and security measures related to EO data infrastructures. EO data is initially thought to contain no sensitive data, and thus should all be openly accessible. However, with increasing granularity and resolution, this notion is challenged. Personas AB and ED both highlighted the necessity to consider the implications of the open access on specific lands and properties that are critical in their respective domains, such as specific forests serving long-term measurements, properties that are part of critical infrastructures or locations that need to be protected for their biodiversity. Having data about these locations as open data poses risks to them. In addition, prioritization with regards to data selection, integration and selection of future pilots is considered based on the developed personas.

This study provides early insights on the infrastructural challenges with emergent data labs; in particular those of ambiguity of its various stakeholders, and how persona development can help tackling these challenges. Applying the persona method in the context of data infrastructures also revealed that the distinction between users and beneficiaries (typically referred to as customers in the persona literature) is not clear, since data usage represents a continuum of processing, analysis and use. The study also provides practical insights on how to identify beneficiaries of emerging data labs using the persona method, as a step towards lowering the entry barriers for them to be engaged in using its constituent offerings for public benefit [3].

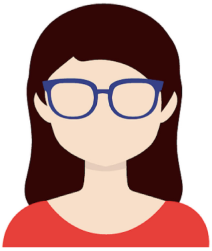
However, the study has a few limitations. The personas developed are as representative of the interviewees as the selected variables, which were driven by their current practices rather than their future use of space data. The current study also does not account for the role of the different consortium members on the personas' characteristics, such as their motivation or attitude towards space data. Future work will address these limitations by testing the persona variables through a questionnaire to a wider sample of beneficiaries. This will enable us to also develop the tools to keep the personas as "live" representations of the beneficiaries. We will also observe how the inclusion and usage of personas influences the development of the data lab into the future. Further research is also needed to explore how the human, data and cyberinfrastructures notions may relate to one another and interact in the context of data labs.

Appendix

ED

EBBA DANIELSSON

GEOGRAPHIC INFORMATION OFFICER
BENEFICIARY | GOVERNMENTAL AGENCY



BRIEF PERSONA DESCRIPTION

Ebba is a geographic information officer in a governmental agency. She is responsible for the acquisition, management and dissemination of spatial data required for her organization to perform its function. Her role is of an executive nature and directly enables the agency's domain-specific operations. She often needs to make strategic decisions in relation to the three responsibilities mentioned earlier.

Ebba's decisions regarding GIS technologies are directly driven by her agency's needs. Her interest in space data grew when current surveying technologies impose limitations to their operations.

SKILLS

Technical background in remote sensing

Advanced GIS user, including data management & administration

ArcGIS, QGIS

USE CASE NEEDS

Ebba has detailed knowledge about applied use cases for space data. They have clear questions, data sets and ways to evaluate it in mind. In case compromise is required, wider spatial coverage is prioritized over longer temporal coverage. Detailed features (i.e. what is being measured) are also prioritized over longer temporal coverage, as long as specific change is detectable.

The motivation for her use cases is largely driven by enforcing regulations, complying with the law and enabling data-driven policies. Ebba anticipates benefits from SDL in terms of feasible wider spatial coverage and better SLAs (i.e. data availability).

To achieve these benefits, close collaboration and service from SDL is required. She is very motivated for such collaboration but will require (partial) external funding to help her realize those use cases. Ebba perceives SDL's data products as its most valuable aspect.

DATA SOURCES

The agency currently sources both vector and raster such as:

- LiDAR
- Aerial photography
- Satellite imagery
- DEMs
- Land surveying
- Vector data & models from other agencies (e.g. EPA)

ORGANIZATION AND SDL

Ebba's organization operates all over Sweden. They are currently generating their own data from surveying, land measurements and area photography. The data is shared with county administrative boards and municipalities.
They have also started using satellite data for some time now, but are facing tremendous challenges regarding data quality, accuracy and availability.

ANALYSES

Ebba does not currently conduct analyses herself but prepares and manages spatial data for other users. She prepares it to be analyzed via:

- GIS systems (e.g. overlays)
- Raster functions

GOALS

- Prepare data products for executive partners faster and with higher quality
- Enforce regulations in a more cost-efficient manner
- Activate data-driven policy making across their organization

EBBA SAYS...

"We have the pressure to manage a lot of [natural and financial] resources that depend on geographic data. It is very costly to do extensive manual surveying. And the European Commission is pushing for us to conduct area monitoring using Sentinel data. But it is also difficult to work with this data. We need an automated way of obtaining and using it… and also develop a decision support system that can help [our operations]. We were hoping that the Space Data Lab will solve these problems for us. That's why we became very interested in taking part in this" (Interviewee E)

References

1. Jetzek, T., Avital, M., Bjorn-Andersen, N.: Data-driven innovation through open government data. J. Theoret. Appl. Electron. Commer. Res. **9**(2), 100–120 (2014)

2. OECD: The Path to Becoming a Data-Driven Public Sector. OECD (2019). https://doi.org/10.1787/059814a7-en
3. Dinesh, A.: Building the smarter state: the role of data labs. Medium, 13 December 2017. https://medium.com/data-labs/building-the-smarter-state-the-role-of-data-labs-5b5428920f0f. Accessed 19 Nov 2019
4. Kremser, W., Brunauer, R.: Do we have a data culture?. In: Data Science – Analytics and Applications, Wiesbaden, pp. 83–87 (2019). https://doi.org/10.1007/978-3-658-27495-5_11
5. Vinnova: Datalabb och datafabrik som nationell resurs (2020). https://www.vinnova.se/e/datadriven-innovation/datalabb-datafabrik/. Accessed 07 Feb 2020
6. Yang, C., Huang, Q., Li, Z., Liu, K., Hu, F.: Big Data and cloud computing: innovation opportunities and challenges. Int. J. Digit. Earth **10**(1), 13–53 (2017)
7. Goodwin, K.: Designing for the Digital Age: How to Create Human-Centered Products and Services. Wiley, Hoboken (2011)
8. Nielsen, L.: Personas-User Focused Design. Springer, London (2019). https://doi.org/10.1007/978-1-4471-7427-1
9. Bright, J., Ganesh, B., Seidelin, C., Vogl, T.M.: Data science for local government. SSRN J. (2019). https://doi.org/10.2139/ssrn.3370217
10. Rizk, A.: Data-driven innovation: an exploration of outcomes and processes within federated networks. Doctoral thesis, Luleå University of Technology, Luleå (2020)
11. Rizk, A., Ståhlbröst, A., Elragal, A.: Data-driven innovation processes within federated networks. Eur. J. Innov. Manag. (2020). https://doi.org/10.1108/EJIM-05-2020-0190
12. Lyon, F., Gyateng, T., Pritchard, D., Vaze, P., Vickers, I., Webb, N.: Opening access to administrative data for evaluating public services: the case of the Justice Data Lab. Evaluation **21**(2), 232–247 (2015). https://doi.org/10.1177/1356389015577507
13. Vinnova: Datalabb och datafabrik som nationell resurs. Vinnova, Sveriges innovationsmyndighet, 2019–01038 (2019). https://www.vinnova.se/globalassets/utlysningar/2015-07023/omgangar/60709b5a-db86-421c-bb86-05616f1f2d1e.pdf932684.pdf. Accessed 24 Nov 2019
14. Paine, D., Lee, C.P.: Producing data, producing software: developing a radio astronomy research infrastructure. In: 2014 IEEE 10th International Conference on e-Science, vol. 1, pp. 231–238 (2014). https://doi.org/10.1109/eScience.2014.41
15. Bietz, M.J., Paine, D., Lee, C.P.: The work of developing cyberinfrastructure middleware projects. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, pp. 1527–1538 (2013)
16. Star, S.L., Ruhleder, K.: Steps toward an ecology of infrastructure: design and access for large information spaces. Inf. Syst. Res. **7**(1), 111–134 (1996)
17. Demchenko, Y., Grosso, P., De Laat, C., Membrey, P.: Addressing big data issues in scientific data infrastructure. In: 2013 International Conference on Collaboration Technologies and Systems (CTS), pp. 48–55 (2013)
18. Lee, C.P., Dourish, P., Mark, G.: The human infrastructure of cyberinfrastructure. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, Banff, Alberta, Canada, pp. 483–492 (2006). https://doi.org/10.1145/1180875.1180950
19. ESA: Copernicus Overview. United Space in Europe (2020). http://www.esa.int/Applications/Observing_the_Earth/Copernicus/Overview4. Accessed 26 Jan 2020
20. Killough, B.: Overview of the open data cube initiative. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 8629–8632 (2018). https://doi.org/10.1109/IGARSS.2018.8517694
21. Lewis, A., et al.: The Australian geoscience data cube—Foundations and lessons learned. Remote Sens. Environ. **202**, 276–292 (2017). https://doi.org/10.1016/j.rse.2017.03.015
22. Adlin, T., et al.: Putting personas to work. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 13–16 (2006)

23. Cooper, A.: The Inmates are Running the Asylum. SAMS, Macmillan, Indianapolis, IA (1999)
24. Nielsen, L., Jung, S.-G., An, J., Salminen, J., Kwak, H., Jansen, B.J.: Who are your users? Comparing media professionals' preconception of users to data-driven personas. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction, Brisbane, Queensland, Australia, pp. 602–606 (2017). https://doi.org/10.1145/3152771.3156178
25. Floyd, I.R., Cameron Jones, M., Twidale, M.B.: resolving incommensurable debates: a preliminary identification of persona kinds, attributes, and characteristics. Artifact **2**(1), 12–26 (2008). https://doi.org/10.1080/17493460802276836
26. Gibbs, G.R.: Analyzing Qualitative Data, vol. 8. SAGE Publications, London (2007)