



Exploring the Challenges of Integrating Data Science Roles in Agile Autonomous Teams

Ivar Hukkelberg^(✉) and Marthe Berntzen

Department of Informatics, University of Oslo,
Gaustadalléen 23B, 0730 Oslo, Norway
{ivarhuk, marthenb}@ifi.uio.no

Abstract. The notion of autonomous teams is core to agile software development. However, autonomy in agile teams is challenged by increasingly complex software projects, large-scale agile and perhaps increasingly multidisciplinary teams. At the same time, data science roles are making their way into agile teams as companies seek to reap the potential advantages of using data to develop better and more competitive services and products. The consequences of implementing such roles in traditional agile teams are largely unknown. In this paper, we take an exploratory approach to the topic of data science roles in agile teams by a set of interviews with five data scientists as well as three members of an agile software development team. Based on the interviews we identify a set of challenges associated with incorporating the role in agile autonomous teams. Based on these challenges we discuss preliminary recommendations for companies seeking to integrate data science roles in agile teams.

Keywords: Data science · Agile · Software development · Teams · Autonomy

1 Introduction and Background

During the past decades, agile practices have spread beyond the traditional software development team to include other roles, parts of the organization, and even the organization as a whole [1]. This introduces challenges such as adapting agile methods while keeping with the central aspects of team autonomy and balancing cross-functional teams with an efficient team size [2]. At the same time, the usage of data science in software development has expanded rapidly [1, 3], possibly introducing new challenges to agile team autonomy.

The notion of team autonomy is not new. In most agile methods the notion of enabling teams to make decisions of their own is central [4]. Such teams are often labelled as self-managing, self-organizing or autonomous. These teams should be cross-functional, consisting of the roles needed for the team to utilize their competence to deliver across roles and organizational functions [5, 6]. The assumption is that cross-functionality contribute to more empowerment and participation within the team [5, 7].

Regardless of the label, merely assembling a group of people and naming them autonomous is not enough to ensure that the group acts as a self-organizing or autonomous team [7]. Some of the identified criteria important for team autonomy

include having a common goal and direction, a trusting team climate, organizational support and efficient skill utilization [4, 8].

The increasing scale and complexity of modern software development has led to new team constellations such as DevOps and BizDev teams [4], or more recently combining and adapting agile development practices to data science and machine learning [1]. Adding increasingly more roles to the agile, cross-functional team, may come with the side effect that autonomy and self-management is difficult to maintain [2, 4]. In particular, introducing new roles such as data scientists into the cross-functional team may pose a challenge with conflicting interest and needs across team members with different backgrounds, terminologies and approaches to work. Indeed, balancing individual and team autonomy are among the key barriers to self-management [7]. We speculate that including data science roles into agile cross-functional teams will further complicate the landscape.

1.1 Machine Learning in Organizations and the Need for Data Scientists

Today, extreme amount of data is generated and stored every day. Organizations are trying to use this data to create new experience and products that are personalized for its users and to stay ahead of competitors. Leading stars in this area like Google, Facebook and Amazon, have succeeded in creating value from the data they store. The key elements of their success lie in a strong digital platform where all data generated can be easily accessed. Machine learning, a subfield of artificial intelligence, is currently the preferred method to use in order to handle the extreme amount of data. Algorithms and statistical models search for patterns, make data-driven decisions and continuously improve themselves without human interference [9]. The people possessing the skill set to work with this combination of computer science and statistics are often referred to as data scientists. This role is not to be mixed with other similar roles, such as data engineer and data analyst. Data engineers are more concerned with maintaining and building infrastructure so that the data becomes accessible [10]. Data analyst build reports and visualizations to explain what insight the data is hiding using statistical methods, but do not spend time programming advanced algorithms that the data scientist role does [11]. Often, the data scientist role in an organization can be described as a researcher trying to find meaning in the data and creating self-improving algorithms [1, 12]. The solutions they build can lead to tasks automation, personalized user content, and much more.

Recently, more traditional organizations have begun to explore how to create value from data. Fleming et al. [13] point at different factors needed to build a data-driven organization. Among these, a key point is having the right competence, such as data scientists. However, merely hiring a data scientist and expect results is likely to be wishful thinking. According to Davenport [14], a good data-driven environment where data scientist can thrive should include a focus on (1) company culture, (2) analytics capabilities, (3) data and technology capabilities and (4) individual capabilities. Here we see that there are many factors involved in order to build a data-driven organization and incorporate the data scientist role. Patil [15] also points to the importance of close collaboration between data scientist and the rest of the organization, and that to create great data products you have to build cross-disciplinary groups. One can see this as an

argument to incorporate the data scientist role into cross-functional teams. A data scientist also needs a work environment where he/she can experiment and let the creativity blossom [15]. This may indicate that they need a high degree of individual autonomy.

1.2 Research Question

While organizations are increasingly making use of both data science on the one hand, and agile methods on the other, little research has examined the interplay between the two, in particular from an agile team perspective. For instance, Larson and Chang [1] examine how agile principles can be adapted and adjusted to data science, but do not discuss how the introduction of data science roles affects autonomy in the agile team. Kim, Zimmermann, DeLine, and Begel [3] discuss the role of data scientists in software development teams, but not agile or team autonomy. As such, our knowledge about including data science roles in agile autonomous teams remains limited. In this study, we explore the following research question: *What challenges do agile autonomous teams face when introducing data scientist roles into the team?*

In this short paper, we take an exploratory approach to our research question. This first section has introduced the topics. Next, we describe our method and data collection procedures and present results from six interviews. Finally, we discuss the challenges identified from the results and suggest initial recommendations for practice.

2 Method

To explore our research question, we conducted six semi-structured interviews with an average length of 40 min. Five separate individual interviews with participants from three different organizations were conducted by the first author. The second author held a group interview with three participants from a fourth organization. Information about the respondents and their organizations are provided in Table 1. To avoid too much bias in our data by only interviewing people directly linked to the data science field, the group interview conducted by the second author was with people from an agile development team without data science experience. This gave us a more nuanced view of the data scientist role. Due to confidentiality agreements, further details about the organizations and their specific cases remain anonymous.

During the interviews, we presented our participants with our research topic and asked them to describe their experiences with data science, agile methods, team autonomy and experiences with implementing data scientist roles into agile teams. We followed the semi-structured approach, asking prepared questions but also allowing the conversations to naturally develop. During the former five interviews, detailed notes were taken, while the group interview was tape recorded based on consent and transcribed by the second author. After the interviews, the authors read through each other's notes and transcriptions, before discussing common topics that had emerged. Next, we separately coded the data. Due to the exploratory approach, the relatively short interview records, and low number of interviews we chose a holistic coding approach [16, 17]. As themes started to emerge, we discussed and resolved any disagreements in coding and interpretation.

Table 1. Data sources

Participants	Organization and team description
P1	Large private consultancy organization. Data scientist team lead, leading a team of two data scientists, 3–4 data engineers, one architect, one security architect
P2 and P3	Large public sector organization. P2 Lead for data scientist department. P3 data scientist
P4 and P5	Small start-up specialized in data scientist tasks. P4 CEO and data scientist team lead. P5 data scientist
Group interview	Three participants from a large public sector organization; one team leader, one tech lead and one interaction designer

3 Results

In this section, we present the results from the interviews, focusing on the main challenges with establishing the data scientist role in autonomous teams. Based on what our respondents told us, and discussion and analysis of the data between the two authors, six challenges and possible recommendations are summarized in Table 2.

Table 2. Main challenges identified during interviews

Challenges	Recommendation
1. Agile methods	Arrange team kick-off; allow time for the team to settle in and mature
2. The data scientist role	Provide training; Make a dictionary for key terms and expressions
3. Additional data roles	Adjust agile practices; What can be adapted to meet the needs of the DS? Adjust agile practices; What can be adapted to meet the needs of DS?
4. Creativity and freedom	Facilitate for Community of Practice (CoP)
5. Collaboration and knowledge sharing	Consider if additional “data roles” are needed
6. Data platforms and infrastructures	Build a data platform to gather data in one place

3.1 Agile Methods

Agile methods were employed to various degrees in all the respondents’ teams. Although their perceptions varied, all of them also had some understanding of what it meant to be autonomous. Many used Scrum practices such as stand-up meetings, sprints and retrospectives. According to the data scientists in our sample, the usage of sprints could be challenging. P4 explained that it can be hard to work according to a sprint schedule, since a data scientist work out from hypotheses, which not always give

something of value from a management and team lead point of view. However, as P4 stated, for a data scientist this provides insight about what is not working, and then can test other methods in the next sprint.

3.2 The Data Scientist Role

From the interviews it became quite clear that *“The definition of a data scientist has become more blurred”* (P1), and that this misconception lead to wrong expectations about what a team want the data scientist to solve, preventing the realization of the full value of having this role on the team. P3, also a data scientist, used her first months in the company working partially for different teams explaining what a data scientist is and what use-cases are suited for them to solve. In the group interview they claimed to have a data scientist on the team, but after we analyzed the interview it became clear that this role was actually a data analyst. They also explained that it could be hard to understand the terminology of the data scientist.

3.3 Additional Data Roles?

All respondents who held data scientist roles expressed that a data scientist in team should ideally be supported by a data engineer. P1, P4 and P5 explained that a data engineer’s job is to create the infrastructure, so the data becomes easily available for the data scientist. Without data, it is hard for the data scientist to do their job. They further explained that teams lacking both roles might need to increase their total numbers of members with two or more. P3 said that one additional resource to the team might not be a big deal, but when first including one more resource, it is easy to add a couple more. However, according to P3, if there are too many resources on a team it can lose its autonomy.

3.4 Creativity and Freedom

To experiment and explore the data was highlighted as important for data scientists to thrive. For example, P3 explained that an important part of her job is to test and explore different hypotheses, and if the environment she works in is too rigid, it becomes difficult for her to do her job. This is also backed up by P1, P4 and P5. Although creativity and freedom are important, they also stated that management must point out the bigger problem they are going to solve. P4 explains that data scientists need a high degree of autonomy: *“A project manager should never tell a data scientist how to do things. Just tell them what are the overlaying problem that needs to be solved and when deadline is”*.

3.5 Collaboration and Knowledge Sharing

Collaboration and knowledge sharing among data scientist were also highlighted as important. P2, who leads a data scientist lab, stated that a data scientist should not allocate all its time to a team project, but also use time working together with other data scientists. This is important she said, because it is one of the best ways a data scientist

can grow and learn. P3 explained that the time spent with other data scientist is valuable and is used to create and test different types of algorithms and make data science tools that can be reused across multiple projects. Along similar lines, the participants in the group interview expressed concern that their data analyst did not have sufficient opportunities for knowledge sharing with peers.

3.6 Data Platforms and Infrastructure

Data platforms and infrastructure was a final theme emerging from the interviews. All the data scientists expressed the need for easy access to data. Both P4 and P5 explained that a good infrastructure should be in place and that a data platform is necessary. Without it, it will be hard for the data scientist to do productive work and would use most of the time chasing data. This, they said, would not benefit neither the data scientist nor the team.

4 Discussion and Concluding Remarks

We now turn to discuss our research question “*What challenges do agile autonomous teams face in introducing data scientist roles into the team?*”.

We believe that understanding what a data scientist is and can do is key for a team to successfully incorporate the role. The confusion about the role, as seen from the group interview, can lead to sub-optimal use of the data scientist. This again can have negative effect both for the team and the data scientist. Therefore, it is important to train the team about what the data scientist can and should contribute with. It could be important that the team set a side time to manage expectations, both within the team and outwards to the rest of the organization. One way to manage expectations is to understand there may be different views of what value is. Often a team lead or manager have a different understanding of value than a data scientist. A machine learning implementation that after two weeks of work did not function as expected, would still in the data scientist eyes provide value in form of knowing what algorithms did not work so well and learn from the experience. However, a manager could see it as a failure and would struggle to find something of value. Therefore, when integrating new roles into the team, it might help the team to reflect upon the different aspects and views of value.

Also, an idea when building a data scientist environment might be to have data scientist partially in teams, and then train the teams in what use-cases is appropriate for a data scientist and help them understand the fields’ terminology. Another suggestion is to arrange team kick-offs to work with specific data scientist use-cases, so the other team members becomes familiar with the topic. In a team kick-off one should also reflect on agile practices and if they need to be adapted to the inclusion of a data scientist.

One should also be mindful about other potential expansions of a team when including the data scientist role. A variety of roles could be needed, in addition to the data scientist, for a team to become data-driven [13]. An increased team size can lead to loss of team autonomy and agility [2]. Therefore, one should think carefully about if other roles must be added to support the new data scientist role. An alternative solution could be to see if current team members can be trained to take on additional roles.

Training current team members to become data scientist might not be feasible, as the role require high expertise of both statistics and programming skills. It would likely take a lot of time and investment to retrain a person for that role. For example, a data analyst might have the necessary statistical background but lack the programming experience, while a software developer lacks the mathematical background.

Alternatively, instead of focusing on training team members to become data scientists, it might be a better solution for the organization to train its members in skills which could help support the data scientist in its work. That way a team might avoid the increase in roles when adding a data scientist to the team. Instead of adding a couple of data engineers join the team, their tasks can be done by current team members, for example building a better infrastructure and provide the data scientist with an architectural overview.

Throughout the interviews, the importance of collaboration with other data scientists were highlighted. As pointed out by Davenport and Patil [12] there is a trade-off between working in cross-functional groups versus interaction with other specialists within their own field. Towards this end, organizations can establish Communities of Practice (CoP) where data scientists can exchange and discuss ideas, develop professional skills and new ways of working. CoP's are important for knowledge sharing, coordination and decision-making in large-scale agile development projects [18, 19]. An open data science CoP could contribute both to the data scientists' skill development, but also raising the understanding of other organizational members.

The creativity and freedom a data scientist require can be seen as the need for a high degree of individual autonomy. This notion is supported by Patil [15], but he does not say anything about how it might affect self-managing teams. In Moe et al. [7] the difficulty of balancing individual autonomy and team autonomy is discussed. They explain having greater redundancy in the team can reduce this problem. This redundancy the data scientist can use to engage in CoP's. Of course, one can also debate if the creativity and freedom the data scientists talk about in the interviews is the exact same kind of freedom and creativity any role in an autonomous team need. However, given the current popularity the data scientist role has in the industry it might be that it feels more natural for the data scientists to talk about it.

Although soft skills are necessary to succeed with implementing the data scientist role into the team, we cannot get by the fact that certain technical conditions must be fulfilled as well. A data-driven platform is a critical component in order to build an organization that aim to make data-driven decisions and utilize the competence of data scientists [20]. A platform is also seen as an important prerequisite for team autonomy [21]. Therefore, if traditional industries are going to succeed with integrating data scientist and let them work autonomous, it is important that they have access to all the data through a digital platform.

As a concluding remark, this is an ongoing study and the findings are preliminary. We recognize the important limitations, such as the use of a convenience sample and single-source design [17]. The fact that several of our respondents were data scientists themselves may represent a bias in the data gathered from the interviews. As mentioned in the method section, to nuance the sample we included the group interview with agile team members working with a data scientist. Further, the recommendations in Table 2 have not been validated and are only suggestions made by the authors. Future studies

with more rigorous design and methods are needed in order to establish confidence in our findings. Future research could also include inter- and intra-team coordination, as team constellations including data scientist roles should be likely to be large-scale projects. Notwithstanding the limitations of this exploratory study, we believe more research-based knowledge on implementing data science roles in agile teams is important as organizations continue to make use of and combine data science and agile methods.

References

1. Larson, D., Chang, V.: A review and future direction of agile, business intelligence, analytics and data science. *Int. J. Inf. Manag.* **36**, 700–710 (2016)
2. Dikert, K., Paasivaara, M., Lassenius, C.: Challenges and success factors for large-scale agile transformations: a systematic literature review. *J. Syst. Softw.* **119**, 87–108 (2016). <https://doi.org/10.1016/j.jss.2016.06.013>
3. Kim, M., Zimmermann, T., DeLine, R., Begel, A.: The emerging role of data scientists on software development teams. Presented at the Proceedings of the 38th International Conference on Software Engineering (2016)
4. Stray, V., Moe, N.B., Hoda, R.: Autonomous agile teams: challenges and future directions for research. In: Proceedings of the 19th International Conference on Agile Software Development: Companion, p. 5. ACM, New York (2018)
5. Lee, G., Xia, W.: Toward agile: an integrated analysis of quantitative and qualitative field data on software development agility. *MIS Q.* **34**, 87–114 (2010)
6. Hoda, R., Noble, J., Marshall, S.: Self-organizing roles on agile software development teams. *IEEE Trans. Softw. Eng.* **39**, 422–444 (2013). <https://doi.org/10.1109/TSE.2012.30>
7. Moe, N.B., Dingsøy, T., Dybå, T.: Overcoming barriers to self-management in software teams. *IEEE Softw.* **26**, 20–26 (2009)
8. Moe, N.B., Dingsøy, T., Dybå, T.: A teamwork model for understanding an agile team: a case study of a Scrum project. *Inf. Softw. Technol.* **52**, 480–491 (2010)
9. Luger, G.F.: *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Pearson education, London (2005)
10. Provost, F., Fawcett, T.: Data science and its relationship to big data and data-driven decision making. *Big Data* **1**, 51–59 (2013). <https://doi.org/10.1089/big.2013.1508>
11. Kunis: Data Analyst vs. Data Scientist (2018). <https://www.springboard.com/blog/data-analyst-vs-data-scientist/>
12. Davenport, T.H., Patil, D.: Data scientist. *Harvard Bus. Rev.* **90**, 70–76 (2012)
13. Fleming, O., Fountaine, T., Henke, N., Saleh, T.: Ten red flags signaling your analytics program will fail. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail>
14. Davenport, T.H.: From analytics to artificial intelligence. *J. Bus. Anal.* 1–8 (2018). <https://doi.org/10.1080/2573234X.2018.1543535>
15. Patil, D.: *Building Data Science Teams*. O'Reilly Media, Newton (2011)
16. Saldaña, J.: *The Coding Manual for Qualitative Researchers*. Sage, Thousand Oaks (2012)
17. Yin, R.K.: *Case Study Research and Applications: Design and Methods*. Sage Publications, Thousand Oaks (2018)
18. Paasivaara, M., Lassenius, C.: Communities of practice in a large distributed agile software development organization – case Ericsson. *Inf. Softw. Technol.* **56**, 1556–1577 (2014). <https://doi.org/10.1016/j.infsof.2014.06.008>

19. Smite, D., Moe, N.B., Levinta, G., Floryan, M.: Spotify guilds: how to succeed with knowledge sharing in large-scale agile organizations. *IEEE Softw.* **36**, 51–57 (2019)
20. Segaran, T., Hammerbacher, J. (eds.): *Beautiful Data: The Stories Behind Elegant Data Solutions*. O'Reilly, Beijing, Sebastopol (2009)
21. Gundelsby, J.H.: Enabling autonomous teams in large-scale agile through architectural principles. In: *Proceedings of the 19th International Conference on Agile Software Development: Companion*, p. 17. ACM (2018)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

