

Requirements Engineering for Data Products: Adapting User Stories

Gina Maestre-Góngora¹, María Clara Gómez-Álvarez²

¹Department of Systems Engineering – Universidad de Antioquia
Medellin – Colombia

²Department of Computer and Decision Sciences – Universidad Nacional de Colombia
Sede Medellín
Medellin – Colombia

{gina.maestre}@udea.edu.co, mcgomez@unal.edu.co

***Abstract.** This paper proposes recommendations for adapting user stories as a tool for capturing and managing requirements that support data-based functionalities (dashboards, AI models, among others). To conclude, aspects such as collaboration between roles, verifiable acceptance criteria, integration with model metrics, and support for data science tools and algorithms are considered.*

1. Introduction

The integration of Artificial Intelligence and Machine Learning (AI/ML) models and analyses (e.g., dashboards) into software products introduces known frictions in requirements engineering: uncertainty about system behavior (data-dependent), the need for continuous validation, and tensions between business objectives, data quality, model performance, and ethical constraints [Alves 2024, Das 2021]. Recent literature converges on the view that traditional agile artifacts (user stories + acceptance criteria) remain valuable but require explicit extensions to express data requirements, performance metrics, validation conditions, and traceability; and collaboration between heterogeneous functions (software, data science, business, operations) is a critical factor that is not yet sufficiently addressed in available processes and tools [Kraut and Transche 2022, Amirian 2024].

2. Methodology

A literature review was conducted in the Scopus database, focusing on recent studies related to agile requirements engineering in data science projects. To this end, two search strategies were defined. The first (R1) combined the terms “data science projects” with “user story,” “agile development,” or “agile methods,” yielding 72 results. The second (R2) broadened the focus to include data products and Scrum, while maintaining the previous concepts, yielding 21 additional documents. The results were filtered by thematic relevance and recency (2020–2026), identifying 7 key studies on the adaptation of agile user story practices in contexts of advanced analytics and artificial intelligence.

3. Results

The main findings on adapting user stories for data products are summarized below:

Extend the classic user story (US) structure for data products—US-DP: Maintain the value-oriented core of the story, but add minimum verifiable fields for (i) data, (ii) analytics/model, (iii) artifact, and (iv) operational/decisional use [Romao et al. 2025] while preserving the story's function as a boundary object that supports coordination between business and technical teams [Kraut and Transche 2022].

Differentiate between types of data-oriented user stories (USs): Studies on Scrum applied to Data Science highlight that not all USs pursue direct functionality for the end user. Classify USs into a) data preparation USs (data acquisition, cleaning, feature engineering), b) exploration/hypothesis USs, c) modeling and evaluation USs, d) integration and consumption USs (dashboards, APIs, UI). [Cordeiro et al. 2024].

Redefine the roles involved in the user story: Make explicit in the project the collaboration between: Product Owner/Data Product Owner, Data Scientist, Data Engineer, Business Analyst, and Software Team [Alves 2024].

Incorporate verifiable acceptance criteria: Acceptance criteria should not be “works/doesn't work”. USs should be linked to model performance tests [Romao et al. 2025]. Metrics should be derived from the data product's objective and its impact on business decisions [Alves 2024].

Use user stories as dynamic and experimental artifacts: Empirical evidence in Data Science projects with Scrum shows that USs should embrace uncertainty and progressive learning [Das, 2021]. Therefore, the following should be allowed: Stories with exploratory objectives, reformulation of US after discovery sprints, and use of formalized analytical spikes as US [Amirian 2024].

Integrate ethical, explainability, and risk considerations from the user story: Recent literature on AI requirements engineering emphasizes that aspects such as bias, explainability, and responsible use should not be treated as late requirements [Halme et al. 2024].

4. Conclusions

The proposal seeks a minimal but operational adaptation of user stories to data products, seeking to maintain the communicative power of the story while introducing the verifiable elements necessary for self-contained data products in software systems. This contributes to explicitly managing interdependencies and trade-offs between software and data products. Future research could formalize integrated user story frameworks for data products aligned with MLOps and data governance.

5. References

Due to space limitations, all references cited in this paper can be consulted in our supplementary data available at <https://doi.org/10.5281/zenodo.18470923> This repository includes raw data from Scopus databases, analysis report generated by R software, document analysis, and article reference list.