



# Review report of a final thesis

**Reviewer:** Pierre Donat-Bouillud, Ph.D.  
**Student:** Bc. Michael Vrána  
**Thesis title:** Automated data analysis pipelines  
**Branch / specialization:** System Programming  
**Created on:** 5 June 2023

## Evaluation criteria

### 1. Fulfillment of the assignment

- ▶ [1] assignment fulfilled
- [2] assignment fulfilled with minor objections
- [3] assignment fulfilled with major objections
- [4] assignment not fulfilled

The thesis presents a pipeline for R as a DSL. It is able of parallel execution with GNU parallel as a backend and can restart failed tasks.

The language and runtime are well documented.

### 2. Main written part

90/100 (A)

The text is well written, although with a few spelling and grammar mistakes. The thesis is well-structured and easy to follow along.

There are some small inaccuracies about R. For instance, the order of class names in the class attribute does not necessarily represent the chain of inheritance, it can also represent multiple inheritance. Non-standard evaluation could be better presented: in Background, it sounds as if it is about laziness, although it is mainly about choosing whichever environment we want to evaluate an expression in.

The descriptions of drake and targets in the Background section should emphasize the downsides of those pipeline systems, to motivate more the creation of pipelinr and show what problems it solves.

Actually, the next section about the Design of the system states that "Pipeliner as a whole also aims to address shortcomings of drake and targets, such as the handling of failed tasks in a distributed environment" but did not describe those shortcomings in the Background section.

Sources are properly cited.

### 3. Non-written part, attachments

97 /100 (A)

The pipeline itself is written R. It heavily uses the tidyverse libraries, especially to perform non-standard evaluation, which is a good choice as it simplifies the implementation. The code is documented, with the possibility of generating a documentation from the code. It is organized as an R package.

The source code is stored on gitlab. Although it is probably open source, the repository should contain or mention an open-source license explicitly.

### 4. Evaluation of results, publication outputs and awards

95 /100 (A)

The thesis replicates one of our own pipelines. I think I could start using it in practice after a bit more of evaluation, including of the performance, as one of the stages of that replicated pipeline could not be totally run with pipelinr.

## The overall evaluation

93 /100 (A)

The thesis is easy to follow along and describes a new, useful, automated pipeline to organize data analysis for R as a DSL written in R. It is an excellent work that I hope to use for my data analysis pipelines.

However, the need of the new pipeline should have been better motivated in the thesis, by pointing out the downsides of the current state of the art of automated data analysis pipelines in R, i.e. targets. That would have given more criteria helpful to compare (more) pipelines written with pipelinr and the ones written with targets, such as the performance, the managing of failed tasks, or the ease of debugging a pipeline.

## Questions for the defense

- What is the performance compared to the targets pipeline?
- Does pipelinr have static branching?
- pipelinr assumes that stage body functions are pure. Would it make sense to statically or dynamically check this assumption?

## **Instructions**

### **Fulfillment of the assignment**

Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.

### **Main written part**

Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies?

Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 52/2021, Art. 3.

Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.

### **Non-written part, attachments**

Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.

### **Evaluation of results, publication outputs and awards**

Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.

### **The overall evaluation**

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.