

Letter of Clarification for NSF proposal "POSE: Phase II: Expanding the data.table ecosystem for efficient big data manipulation in R."

From: Matt Dowle, creator and maintainer of data.table

Affiliation: None. Employment by H2O.ai ended Dec 2022.

Relationship to Proposing Team: PI Toby Dylan Hocking is a contributor to data.table

In my previous letter of collaboration to Dr. Hocking, I wrote "I wholeheartedly support this POSE application and will collaborate with Toby's team." The point of this letter of clarification is to make it clear that, as the creator and current leader of the data.table project, I intend to participate in the proposed process to create a new administration/governance, which was outlined in Section 2.1 of the NSF proposal: "We will change the contribution and decision-making process by defining an explicit governance policy, creating new roles to facilitate PR submission/review, and better supporting the release management." In particular, I want to emphasize my agreement on the following points.

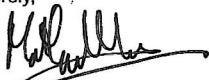
**1. To create a new model for administration/governance of the data.table project, I am open to participating in the community process outlined in the NSF proposal, and to following the new rules that will be established by the resulting governance document,** "One goal of our project is to propose a formal structure for project governance and decision-making and to create explicit written documentation of that structure...we will promote participatory design sessions with several community members to define priorities, requirements, and rules. Finally, we will produce documentation of the proposal and open it for feedback and discussion in a GitHub issue...The process of formalizing governance is anticipated to take a full year before consensus is reached and a governance document is adopted. Overall, the result of this activity will be a well-defined set of rules that the data.table developer community can rely upon to move development forward more confidently and more rapidly."

**2. I am open to the new roles for hierarchical code review outlined in the NSF proposal,** "We propose to replace the current flat leadership structure with a hierarchical structure that will offload much of the code review work to intermediate "reviewers" who are in charge of specific features/files (similar to the successful concept of sub-system maintainer from the Linux Kernel project, see Figure 7). As a part of this transition, we will create a new web page with an explicit mapping between files and who maintains them, similar to the Linux kernel maintainers list [132]. ...The result of this activity will be a hierarchical leadership structure, with intermediate reviewers who take on much of the code review work that the current author handles by himself, so the project will be able to expand to accommodate new contributors more easily."

**3. I am open to the new role of release manager outlined in the NSF proposal,** "We plan to create documentation for the data.table release process, including what to do prior to a release, and create new roles so people other than the primary author can take responsibility for each step."

I want to reiterate my support for this NSF project, and I want to emphasize that any conflicts will be resolved by public discussions, and community consensus. Overall, I believe that this NSF project will be beneficial for the long term sustainability of the data.table project, because it will result in a more de-centralized, robust, and diverse community of project members and contributors.

Sincerely,



Matt Dowle

2/6/2023