Proposal Review 3 : 1931382

Back to Proposal

Agency Name:	National Science Foundation
Agency Tracking Number:	**1931382**
Organization:	
NSF Program:	Software Institutes
PI/PD:	Roberts, Amy
Application Title:	Elements: Improving tools based on data-description standards for gigabyte-scale data sets
Rating:	Fair

Review

Summary

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

Strengths:
-- The PI proposes to enhance the kaitai-struct binary parser generator, to generate more efficient parsers by targeting the scikit-hep/awkward-array library, rather than the usual python structures.
-- The concept is not particularly innovative or transformative, but the resulting tool would make it easier to construct parsers for arbitrary binary formats (assuming they're described).

Weaknesses:
-- The utility of the proposed tool has not been clearly demonstrated. The motivating user stories all center around the problems of hypothetical users to build poorly documented libraries to read custom data formats. Yet the proposed solution û to instead use a code-generator to build parsers û makes a (huge?) assumption that the format itself is sufficiently well documented to be expressed in the kaitai-struct language. Having more real-world evidence that this is the case would be helpful. Secondly, the target audience here are users who don't know enough to write a binary parser of their own, yet who will sufficiently understand kaitai-struct concepts to express the binary data format in that language. It's not clear what is the size of this community?
-- The insufficiency (performance problems) of existing kaitai-struct parsers has not been quantitatively demonstrated.

-- The plan seems overscoped for the proposed deliverable. Developing an additional backend for kaitai-struct does not seem to require more than a 6-12 month effort.
-- The PI appears familiar with kaitai-struct software; however, no evidence is provided of the PI's experience in either leading or contributing to broad community-maintained software projects. This causes concerns when it comes to the ability of the team to build a user and developer community (which is the majority of the effort in the proposal).
-- It is not explained why the enhancement of kaitai-struct cannot be a deliverable of the SuperCDMS grant (given much of the science motivation comes from enabling SuperCDMS research).

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

Strengths:
-- The proposal would contribute a backend to kaitai-struct, which is used more broadly than just HEP
-- The proposal opens a way for undergraduate participation in coding projects.

Weaknesses:
-- The impact may be reduced by using a HEP-specific backend (awkward-array).

Please evaluate the strengths and
weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if
applicable

Science-driven:
Weaknesses: The proposal mentions the SuperCDMS experiment as a science case, but does not illustrate a specific use case where this would help.

Innovation:
Strengths
-- The proposal describes how to incorporate user feedback in iterative development.
Weaknesses
-- The transformational impact to underlying science is not demonstrated.

Close collaborations among stakeholders:
Strengths
-- The proposal builds on previous libraries, and collaborates with both HEP projects and Gateway institutes.
Weaknesses
-- none

Building on existing, recognized capabilities:
Strengths
-- The proposal builds on previous libraries. No significant CI resources are required for its execution.
Weaknesses
-- none

Project plans, and system and process architecture:
Strengths
-- The project builds on existing software.
-- Branching workflow and CI are mentioned.
-- We commend the choice of the MIT license.
Weaknesses
-- The duration of the project is not clearly justified. The actual effort required to add an additional backend to kaitai-struct appears on order of 6-12 months (depending on the expertise of the person performing the work).

Deliverables:
Strengths
-- The project describes the deliverables.
Weaknesses
-- none

Metrics:
Strengths
-- Metrics are well defined.

Sustained and sustainable impacts:
Weaknesses
-- It is not clear that having everyone learn a data definition language will win over the model where the libraries are written/provided with/for the data. Stronger justification should be provided for why the PI expects there will be a strong uptake of their library.
-- It is not clear how this will be turned into a self-sustaining community driven project. The PI does not appear to be a contributor to similar projects (or if they have, they have not mentioned it in the proposal), and experience is key in fostering a developer community.

Alignment with Directorate Specific Priorities:
Aligned.

Summary Statement

For the reasons mentioned in the discussion of intellectual merit, we recommend addressing the raised issues to improve the strength of the proposal.