

Proposal Review 1 : 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:	Good

Review

Summary

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

Processing binary data sets requires software that can load the data based on the data format. There are existing tools for loading the data, given a text description of the format, but their data structures do not support efficient processing of the data. This proposal will remedy this shortcoming by creating a system that supports run-time queries of these data sets using more efficient data structures.

Strengths

- * Lowering the bar for researchers to access data is helpful in broadening participation in research.
- * The PI is well-qualified to conduct the proposed research, having existing collaborations that provide both prior experience on related research and an existing user base for the solution.
- * Engagement with the user community is provided through a wide range of options, including annual workshops, a wiki, etc.

Weaknesses

- * The primary drawback of using Kaitai Script as is for the problems being solved seems to be that the internal Python data structures provide poor performance for data access. Does this problem also exist with other languages such as Java or C++? Why is patching Kaitai Script, which is open-source, to create a more efficient Python data structure using awkward arrays not a viable solution? In other words, why is a new project required?

* It isn't clear how the proposed research addresses all the concerns raised in the user stories. Each of the users would still need to know the data format and it isn't clear how this information is obtained in some of these cases.

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

Strengths

- * Data is useful only if it can be used, and the tools in this proposal will lower the barriers to accessing data.
- * Making data more accessible increases equity by broadening participation of those who have fewer resources available to help them overcome artificial obstacles like file formats.
- * Strong plan for outreach to the user community, including annual workshops to help users with their research and gather feedback on how to improve the system.

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

* Science-driven:

Strengths

- * Large data sets are common and efficient tools for processing them are needed.
- * Research productivity can be improved when there is a common repository of file formats so researchers don't need to reinvent the wheel to incorporate data sets into their work.

* Innovation:

Weaknesses

- * The proposal does not make clear why a patch to the Kaitai Script project to use the awkward array data structure in place of the existing data structure would not be an adequate solution. Why is a new project better?

* Close collaborations among stakeholders:

Strengths

- * The PI is well-established in the nuclear physics community. Based on these collaborations, the PI has an existing potential user community and user interest in the research.

* Building on existing, recognized capabilities:

Strengths

- * The PI has been working on related problems and has a research group that provides an excellent user base for building and testing the proposed solution.

* Project plans, and system and process architecture:

Strengths

- * Each aspect of the plan fits well with the other parts. The project plan is credible, and the user engagement is strong.

Weaknesses

- * The personnel on the budget and the budget justification appear to be inconsistent, so it isn't clear how the work will be divided. The salary for the scientific programmer is low and for less than 10 months per year, so retaining this person will be a challenge.

* Deliverables:

Strengths

- * The software development and user community building are key components to success of the proposal.
- * The software will be made available under the MIT, CC BY 4.0, or similar permissive open-source license.

* Metrics:

Strengths

- * The metrics, which focus on citations in publications and contributions from users outside the PI's group, are suitable for the project.

* Sustained and sustainable impacts:

Strengths

- * Once the software is written and tested, and some workshops have been run, minimal maintenance should be required.
- * Access to the data sets is a compelling reason for users to stay engaged in the project, supporting further improvements of the software.

Weaknesses

- * The sustainability plans are somewhat ad hoc and most likely will require someone to be engaged in the project with more commitment to have

sustainability.

* Alignment with Directorate Specific Priorities:

Strengths

- * The existing collaboration with and interest from the SuperCDMS project suggests that this would be useful for Dark Matter researchers.
- * The software tool is general and could be used by many other research groups as well.

Summary Statement

Good: A software tool that supports real-time queries of large (GB or TB) data sets while relieving users of the burden of learning new file formats would be very useful for researchers. This tool would be helpful in lowering the entry barrier for new researchers and increase equity and participation. The PI has the necessary background and skills to accomplish the project. However, the proposal has limited innovation, being primarily a combination of an existing open-source project with an efficient data structure defined in another project. The scope of the project is also quite modest for the proposed time line and level of effort.

About Services

- Account Management
- Award Cash Management Service (ACMS)
- Notifications & Requests
- Project Reports
- Proposal Status
- Public Access

NSF Award Highlights

- Research Spending & Results
- Contact
- Contact Help Desk

News & Discoveries

- News
- Discoveries
- Multimedia Gallery

Funding & Awards

- Recently Announced Funding Opportunities
- Upcoming Funding Opportunity Due Dates
- A-Z Index of Funding Opportunities
- Find Funding
- Award Search
- Proposal & Award Policies & Procedures Guide (PAPPG)

Publications & About NSF

- Publications
- About the National Science Foundation
- Careers
- Staff Directory

Feedback ▶

See all NSF social media ▶

[Website Policies](#) | [Budget and Performance](#) | [Inspector General](#) | [Privacy](#) | [FOIA](#) | [No FEAR Act](#) | [USA.gov](#) | [Accessibility](#) | [Plain Language](#) | [Contact](#)

The National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA Tel: (703) 292-5111, FIRS: (800) 877-8339 | TDD: (800) 281-8749