**EAGER: Enhanced Robust Persistent Identification of Data (E-RPID)**  
https://www.nsf.gov/awardsearch/images/common/greenline.jpg

|  |  |
| --- | --- |
| divider line | |
| **Investigator(s):** | Robert Quick rquick@iu.edu (Principal Investigator) Marlon Pierce (Co-Principal Investigator) Laurence Lannom (Co-Principal Investigator) |
| divider line | |
| **Sponsor:** | Indiana University 509 E 3RD ST Bloomington, IN 47401-3654 (812)855-0516 |
| divider line | |

**ABSTRACT**  
https://www.nsf.gov/awardsearch/images/common/bluefade.jpg  
An increasing challenge in modern research is the rapid growth in ability to collect and store research data; this growth is much more rapid than the ability to catalog, make accessible, and reuse these data. Major challenges facing the scientific research enterprise include making data discoverable, accessible, and re-usable at Internet scales; and making it possible to use data to replicate analyses done in published research. The importance of this issue is at the center of many recent initiatives in Open Science and Open Data. The objective of this project is to address these deficiencies and to enable more options for data interoperability and reusability in the current research data landscape. One aim is to demonstrate how technical solutions for data accessibility and management should be designed, implemented, operated, and usable by a broad range of researchers. Another aim is to ultimately increase the publication, discovery, and reuse of data, thus contributing to the greater integrity of the scientific enterprise.

This project aims to create a full implementation of the Digital Object Architecture (DOA) concept to improve data interoperability and reusability and potentially transform the way research in many disciplines is conducted. The approach is to integrate existing and new software components into the Robust Persistent Identification of Data (RPID) testbed which itself is a 2-year pilot project supported by NSF (Project ID 1659310). Components to be added include a Digital Object Interface Protocol (DOIP) which defines common operations performed on digital objects; and a Repository Mapping Service which would allow mapping the contents of existing repositories to the Digital Object Architecture (DOA) environment. With the new components, the ensemble "Enhanced-RPID" (E-RPID) testbed will allow managing of data objects by assigning individual identifiers toward making data FAIR (findable, accessible, interoperable, and reusable). This technical effort along with planned educational material are designed to lower the barrier for adoption of Persistent Identification (PID)-centric data management throughout the data lifecycle, from initial data collection to long-term use and reuse. The E-RPID testbed will be housed on the NSF-funded Jetstream computational resource, will be applied to a number of important scientific use cases, and will contribute to activities fostered by the Research Data Alliance (RDA). The outcomes of this project are aimed to broadly benefit disciplinary and multi-disciplinary research and to facilitate better access to and utilization of existing data repositories.

This project is supported by the National Science Foundation's Public Access Initiative which is managed by the NSF Office of Advanced Cyberinfrastructure on behalf of the Foundation.  
  
This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.

**Type-Based Automation of Scientific Data Management**  
https://www.nsf.gov/awardsearch/images/common/greenline.jpg

|  |  |
| --- | --- |
| divider line | |
| **Investigator(s):** | Giridhar Manepalli gmanepalli@cnri.reston.va.us (Principal Investigator) |
| divider line | |
| **Sponsor:** | Corporation for National Research Initiatives (NRI) 1895 Preston White Drive Reston, VA 20191-5434 (703)620-8990 |
| divider line | |

**ABSTRACT**  
https://www.nsf.gov/awardsearch/images/common/bluefade.jpg  
An approach to scientific data interoperability and reuse is through global, persistent, and uniquely identified data types that can be assembled to characterize research data sets. This project proposes to identify data types using persistent identifiers (PIDs). The PIDs resolve to records that specify the way in which metadata, such as the provenance of the data, is structured and recorded. The basic premise is that machine interpretable data is a critical goal to achieving FAIRness (findability, accessibility, interoperability, and reuse) of data as data discovery at a global scale depends on automated processing of the information in digital form. A type based approach to data interpretability that utilizes persistent IDs at the granularity of data types can overturn the Internet and stimulate an ecosystem of new tools for FAIR data. This pilot effort involves evaluating the approach through, in part, by constructing a critical mass of use cases.

This project is supported by the National Science Foundation Public Access Initiative which is managed by the NSF Office of Advanced Cyberinfrastructure on behalf of the Foundation.

This award reflects NSF's statutory mission and has been deemed worthy of support through evaluation using the Foundation's intellectual merit and broader impacts review criteria.