

Project Curation Services

Summary:

Publicly-funded grid research projects often require the research teams to design and execute effective plans for cyber security, open source Software distribution, interoperability, and data management, including sub-plans for data collection, integration, collaboration, publication, and preservation. Our Project Curation Services provide the research community with powerful solutions to meet these requirements, leveraging the results of **GRID DATA Repository** project funded by the DOE ARPA-E. The solutions are already chosen in multiple recent R&D projects.

Industry Needs:

Public-funded R&D is a primary incubator of leading-edge solutions for improving grid management. Many of these projects are collaborations between academia, labs, utilities, and private industry companies.

Sadly, most research never gets developed past the publication of a journal paper. Frequently, the originating research team is simply not in the best position to take a new technology to market. For this reason and others, it is important that we improve public access to raw research, data, algorithms, and results. The success of the open-source software and crowd-source business models shows how open licensing of data and technology can foster new innovative solutions and markets, often in unforeseen ways.

The utility industry is behind other industries in digital publication and open licensing of *data and software* artifacts of R&D. While research summaries are traditionally published as peer-reviewed papers in industry journals, it is *not yet* common practice to also publish the associated data, software, or raw results in digital, open, reusable form. This is often because there is simply no requirement to publish that data, or the mandate is loosely enforced. Industry journals also tend to limit the redistribution of the paper and related artifacts. However, the most significant challenge is that raw data and results from grid research often contain confidential, propriety, and security sensitive information with tight restrictions on public release.

When data is published, it is frequently not digitally accessible by the public:

- Data published as a simple table on paper, making reuse difficult
- Data kept on internal web servers of the research institution, effectively private
- Data that is not adequately described or searchable, effectively hidden
- Data published in proprietary or outdated formats, and/or with license restrictions
- Data sets that are incomplete and insufficient to validate or repeat results

A similar problem exists with software algorithms that are developed:

- Algorithms documented only as pseudo-code on paper, making reuse difficult
- Software kept on internal web servers of the research institution, effectively private
- Software lacking good documentation, or in a non-searchable code repository
- Software with dependencies on other proprietary commercial software
- Software with outdated libraries or compilers, or poor build instructions
- Software without clear or open license terms for reuse or redistribution

Finally, the lack of data and software *publishing* is a symptom of a larger problem with collaborative, publicly funded R&D projects in the utility industry. Existing software and skillsets do not meet emerging needs for managing data, software, and team collaboration *during the project*:

- Enterprise repository software solutions do not work well across organizations and firewalls

- Cloud-based solutions lack sufficient features and security without premium subscription levels
- Lack of experience with open source licensing, interoperability standards, and cyber-security
- Lack of neutrality to help broker and protect intellectual property rights of each team member
- Lack of experience curating raw research data and software for public release

Project Curation Services

We are currently leading a consortium to develop an open, standards-based grid model and software repository, with funding from the US DOE ARPA-E GRID DATA program. The consortium includes over 50 electric utilities, grid operators, national labs, universities, and industry software vendors, collaborating under the nonprofit organization BetterGrids Foundation, Inc. (Bettergrids.org). The GRID DATA Repository will simplify grid data sharing, R&D collaboration, and results publication to facilitate research of new grid optimization and control algorithms. A first version of the Repository will be available as a cloud-hosted 'software-as-a-service' in second quarter 2017. Support will be provided by the BetterGrids.org community.

Leveraging this Repository, and as a value-added service, we offer complete **Project Curation Services** to assist teams with data and software asset curation through the entire R&D project lifecycle:

- **Planning and Compliance:** Identifying the digital data and software assets to be used or generated during the project. Creating data management, collaboration, interoperability, cyber-security, software distribution, and results publication plans for those assets. Ensuring projects comply with publicly-funded R&D project guidelines (DOE, NSF, EPIC, etc.) and industry best-practices.
- **Secure Team Collaboration:** Creating and managing Private GRID DATA Repository instances for project teams to securely share data and collaborate during the project. We provide setup, configuration, hosting, administration, operation, backup, and security for each Private Repository.
- **Grid Data Collection & Creation:** Collecting grid models and data from project members and public sources, including those developed by the ARPA-E GRID DATA program. Creating any needed synthetic models, simulation data, or testing scenarios that are unique to project needs.
- **Interoperability & Security:** Identifying data exchange requirements and open interface risks. Selecting interoperability methods, protocol standards, and security layers. Assistance designing and building interface adapters and security frameworks. Building and managing an integration testing platform and environment. Conducting integration and penetration testing.
- **Results Publication:** Identifying data and software assets to be released into public domain to support results validation and grant publication requirements. Removing any personally identifiable, proprietary, confidential, or security sensitive information. Adding search friendly meta-data. Publishing in the GRID DATA Repository, cross referencing with online journals, and long term data preservation.

Our project curation services are already embedded in three separate industry collaboration projects selected for the DOE SunShot Initiative ENERGISE; a \$30M program to support integration of solar energy into the nation's electric grid (See <http://gridbright.com/news>). Project teams see value in leveraging the Repository for data sharing and collaboration, and relying on our curation expertise for planning, compliance, and publication. And by ensuring data, software, and results are all digitally published and preserved, the DOE can maximize their return on funding and benefits to the community.

For more information, contact info@BetterGrids.org.