

Building the Technical Ecosystem for the Data Archive (Darc) APIs, Analytics, Reflector



Daniel Gardner (LLNL), Harleen Kaur (LLNL), Sam Maphey (LLNL), Yvonne Mui (LLNL)

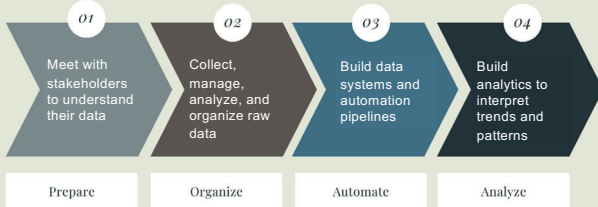
Darc

Serves as a centralized data repository for unclassified and classified test data generated by LLNL weapons programs, enabling users to efficiently search and retrieve content through advanced search capabilities.

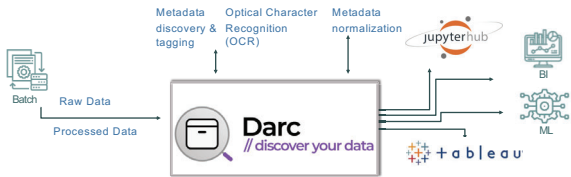
Technical enhancements have been made to **Darc**, focusing on a REST API, analytics and visualization dashboards, and the innovative Darc Reflector feature, which collectively improve data accessibility and analysis.

SD Apps' Data Lifecycle Management Team

The DLM team collaborates with data producers to understand data needs and improve current, outdated workflows. Data engineers facilitate the aggregation and characterization of legacy data and develop more efficient workflows for data producers.

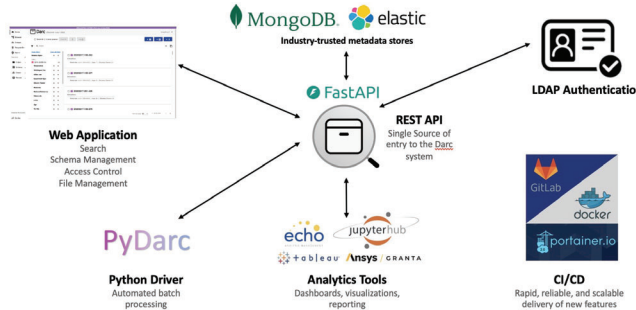


Data Workflows



Data engineers use Darc API to create automated ingestion pipelines to curate data.

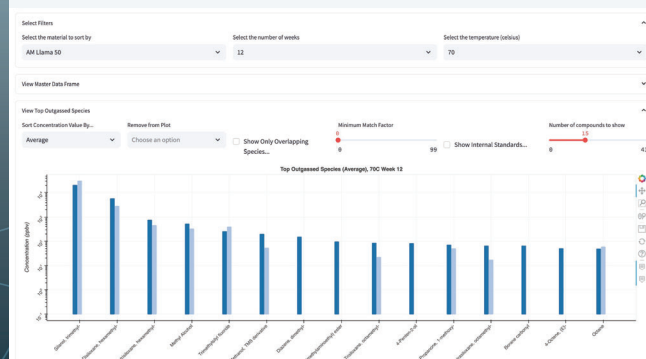
Darc Architecture



Darc REST API connects the database to web browser front-end and any number of external analytics tools.

Darc Analytics

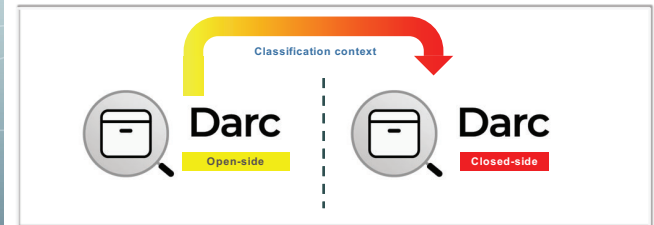
Darc has established API connections with analytics platforms such as Galileo and Tableau, creating a robust analytics engine that empowers users to derive meaningful insights and identify trends.



SMEs say the analytics: "... [Save] countless hours and [deliver] key findings to stakeholders much faster"

Darc Reflector

The latest feature, **Darc Reflector**, allows for the automated reflection of open-side data onto the closed-side environment. Maintaining unclassified data in a dedicated repo helps improve accessibility to unclassified and remote employees, and Reflector enables users to add and connect classified contextual info without recreating datasets. This initiative utilizes existing tools such as Hopper, GitLab, and Nexus to extend data analysis capabilities into the hybrid classified environment. The goal is to provide a unified system and set of tools that facilitate learning, analysis, and understanding of user data across both open and closed sides.



High-side Darc has read-only open-side data, classified data, and context and connections between records.

Darc Roadmap

- | Now | Next | Future |
|--|--|--|
| <ul style="list-style-type: none"> Advanced search capabilities Darc UI improvements Data Metrics API connections to Windchill, DOORS, etc. to enable digital thread | <ul style="list-style-type: none"> Ingest HPC simulation data Share common metadata framework using Semaphore with other SD Applications like DSpace | <ul style="list-style-type: none"> Connect sim and physical test data to enable V&V Natural language search using LLMs Leverage modern infrastructure (e.g. DataBricks) to ingest data prior to metadata curation |

Darc helps //discover your data with search, automation, analysis, and accessibility – supporting informed decision-making.



The work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC02-07NA27004.

LLNL-POST-870745

