Sandia National Laboratories

SAND2019-10558C

Citadel: Searching Metadata

Rebecca Levinson, James Mitchell

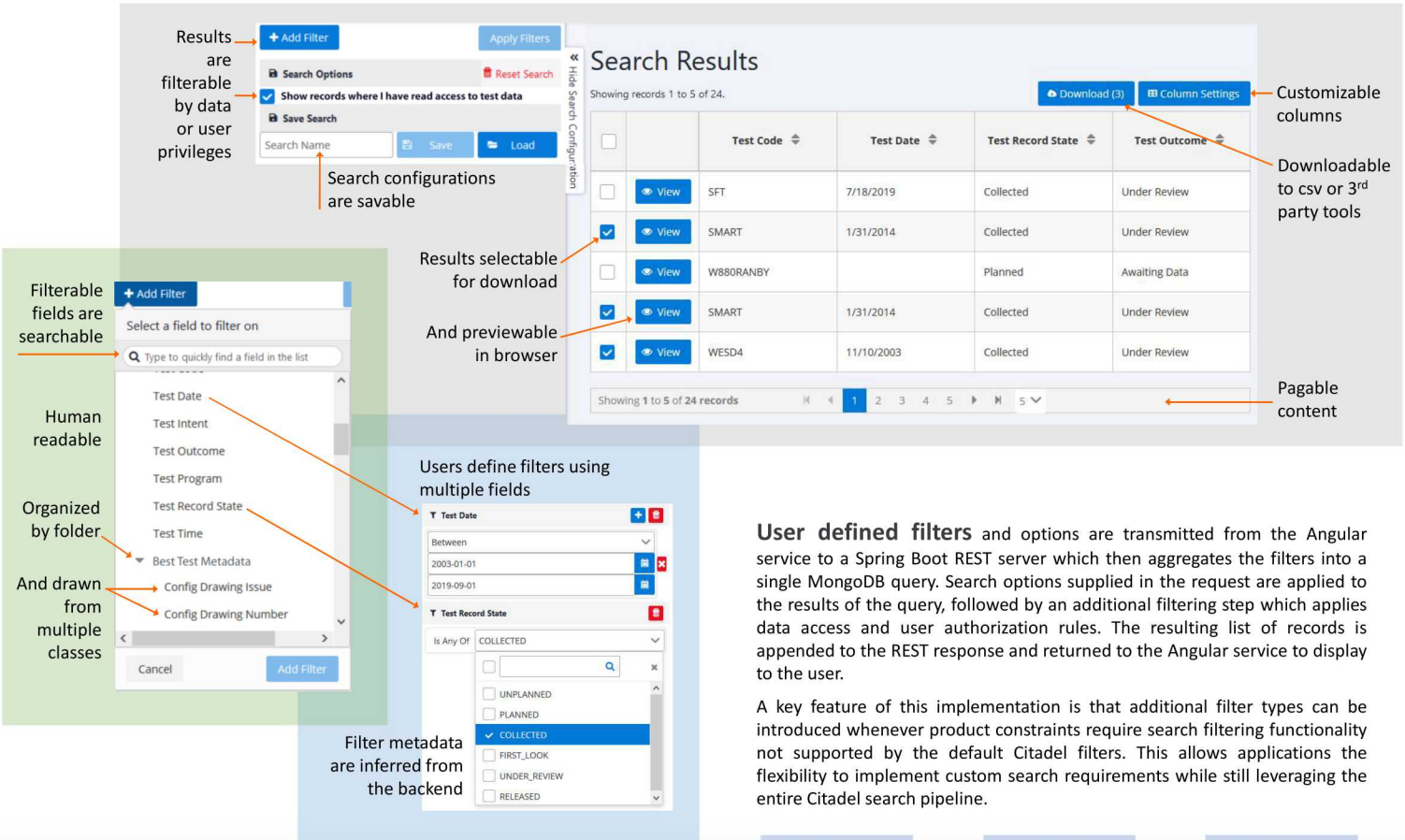
Citadel is a custom code framework designed to quickly roll out multiple data storage applications at Sandia National Laboratories. Citadel helps projects by providing features common to all data system applications. These include: enabling CRUD operations, tracking changes to the data (via a user accessible provenance database), parsing common data types into Parquet files, opening RESTful API's for third party tools to interact with the data (see posters on WASP), and establishing custom access controls to data on a granular level.

Currently the two implementations of Citadel are the SEDS and DataSEA projects. SEDS focuses on replacing a large legacy data system. It has hard-coded data structures and rules for accessing data. DataSEA focuses on quickly rolling out smaller data systems that allow the users to determine their own data structures and custom rules for storing, accessing, and manipulating data.

Citadel's search must enable a diverse set of users to intelligently access their data. Projects evolve quickly, and Citadel supports a wide set of projects, so data in records on the backend must be searchable on the front-end with as little customization as possible. Moreover the set of searchable data must include metadata fields directly in records, those connected through composition, and those associated through links in order to support robust and complex architectures such as that needed in SEDS.

The UI for selecting search fields and viewing search results must be user friendly- easy and fast to find fields of interest to filter on, intuitive to combine fields to dig down on search results, and straightforward to search with metadata of different types.

Search Results and Filter *Angular* UI

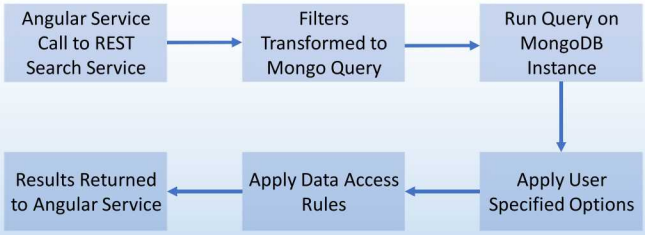


Descriptive metadata are required by the Citadel framework to properly display filters and search results in the Angular UI. Current implementations of search generate this metadata in one of two ways:

- SEDS uses Java reflection and annotations to specify field-level data such as the 'pretty' display name, field type, and dropdown options, as appropriate. In this way, all front end information for each field is retained with the field in the back end.
- DataSEA employs a Java metadata generation service to analyze Java representations of user defined schemas and build lists of field-level data. This service is made available to the frontend by a Spring Boot REST service.

User defined filters and options are transmitted from the Angular service to a Spring Boot REST server which then aggregates the filters into a single MongoDB query. Search options supplied in the request are applied to the results of the query, followed by an additional filtering step which applies data access and user authorization rules. The resulting list of records is appended to the REST response and returned to the Angular service to display to the user.

A key feature of this implementation is that additional filter types can be introduced whenever product constraints require search filtering functionality not supported by the default Citadel filters. This allows applications the flexibility to implement custom search requirements while still leveraging the entire Citadel search pipeline.



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.