

The Materials Data Curation System

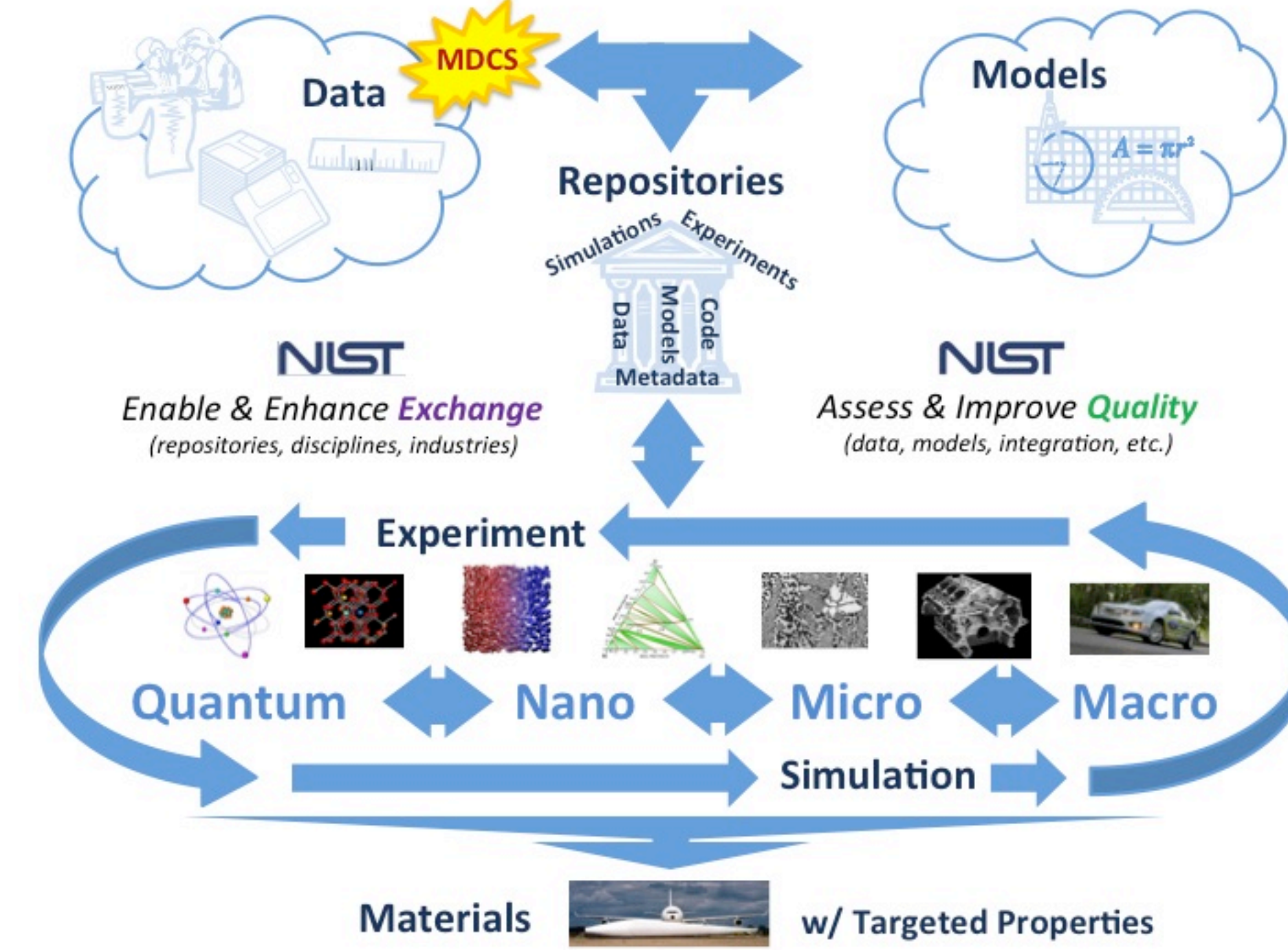
Mary Brady¹, Carelyn Campbell¹, Alden Dima¹, Philippe Dessauw¹, Marcus Newrock¹, Adele Peskin², Pierre-Francois Rigodiat¹, Guillaume Sousa Amaral¹, Zachary Trautt¹, Sharief Youssef¹

¹National Institute of Standards and Technology
Gaithersburg, Maryland 20899

²National Institute of Standards and Technology
Boulder, Colorado 80305

Introduction

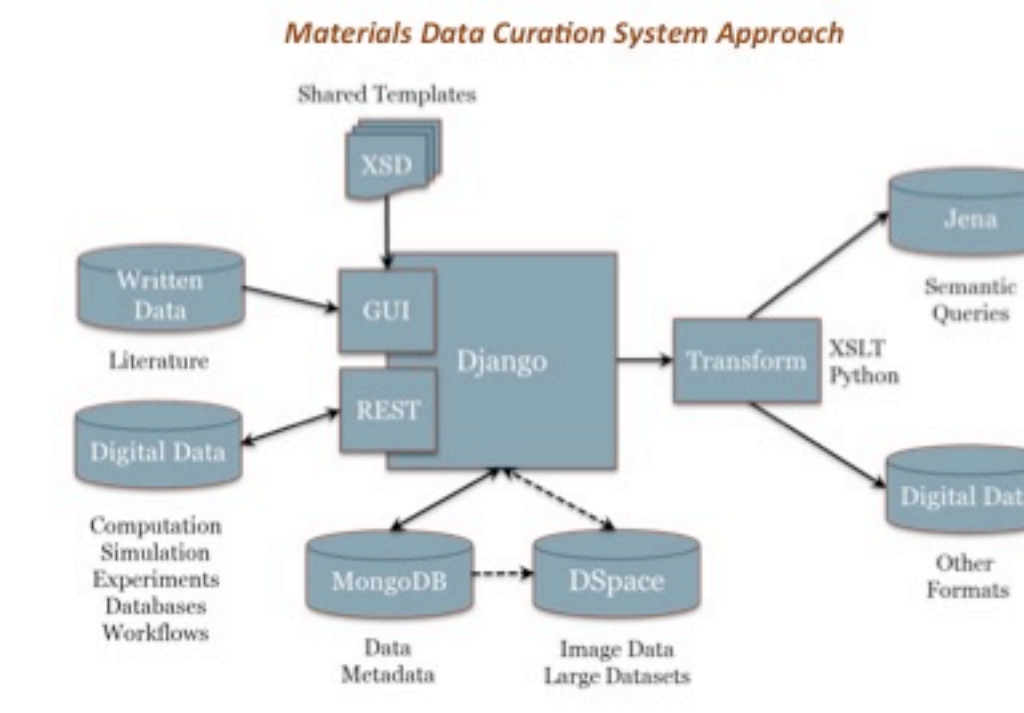
The NIST Materials Data Curation System (MDCS) provides a means for capturing, sharing, and transforming materials data into a structured format based on the Extensible Markup Language (XML) that is amenable to transformation to other formats such as those used by existing computational tools. The data are organized using user-selected community-developed templates encoded in XML Schema used to create data documents that are saved in a non-relational (NoSQL) document database. Each project, group, or organization can run as many MDCS instances as needed. Individual MDCS repositories can be interconnected for federated searches and data sharing.



How the MDCS fits into NIST's overall MGI strategy

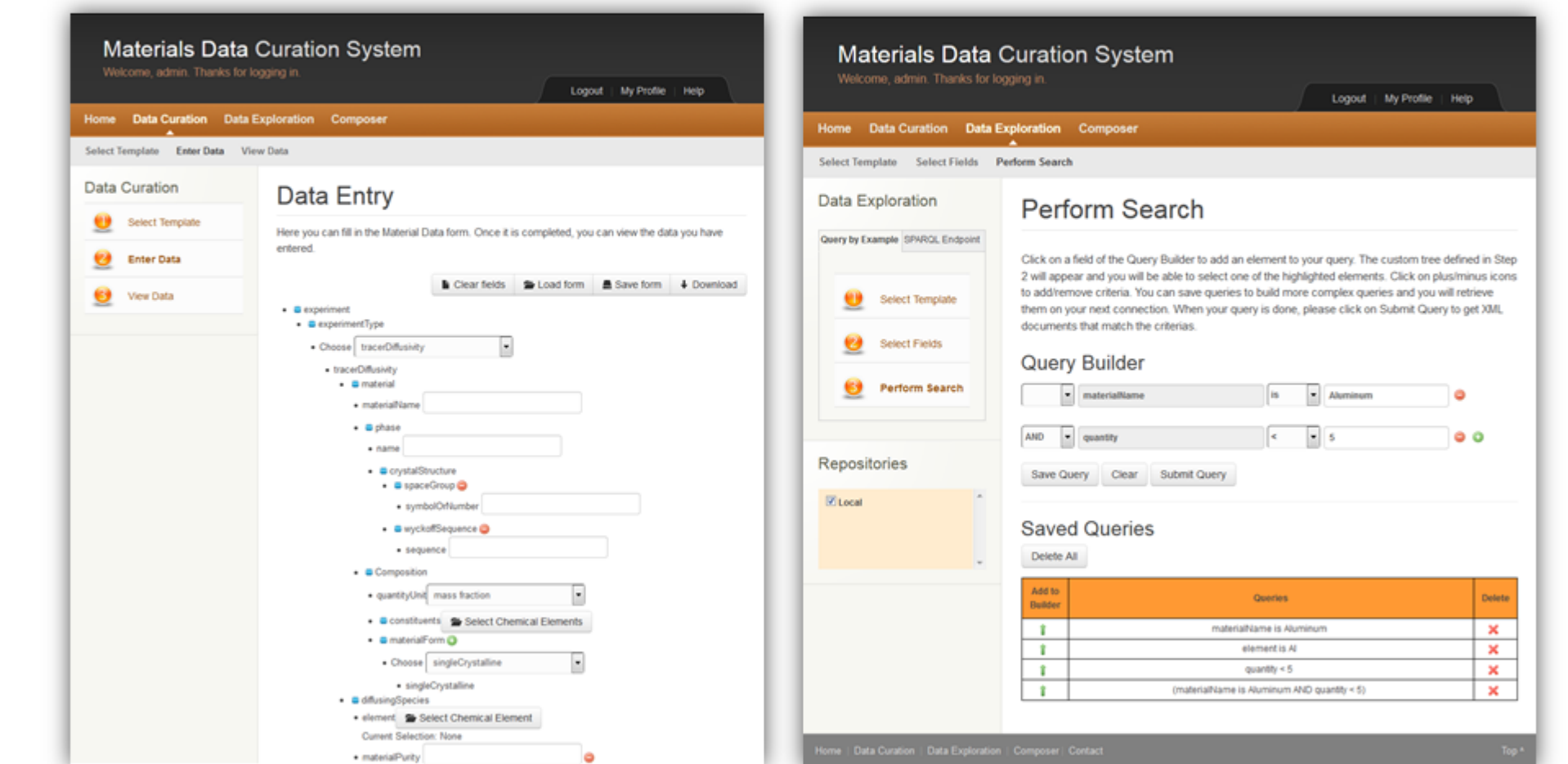
The MDCS Approach

- Web-based
 - Python/Django/MongoDB
 - REST API
 - XML-based
 - SPARQL queries
- Store data in XML-based templates
- Store, manage, & compose templates
- Spreadsheet input



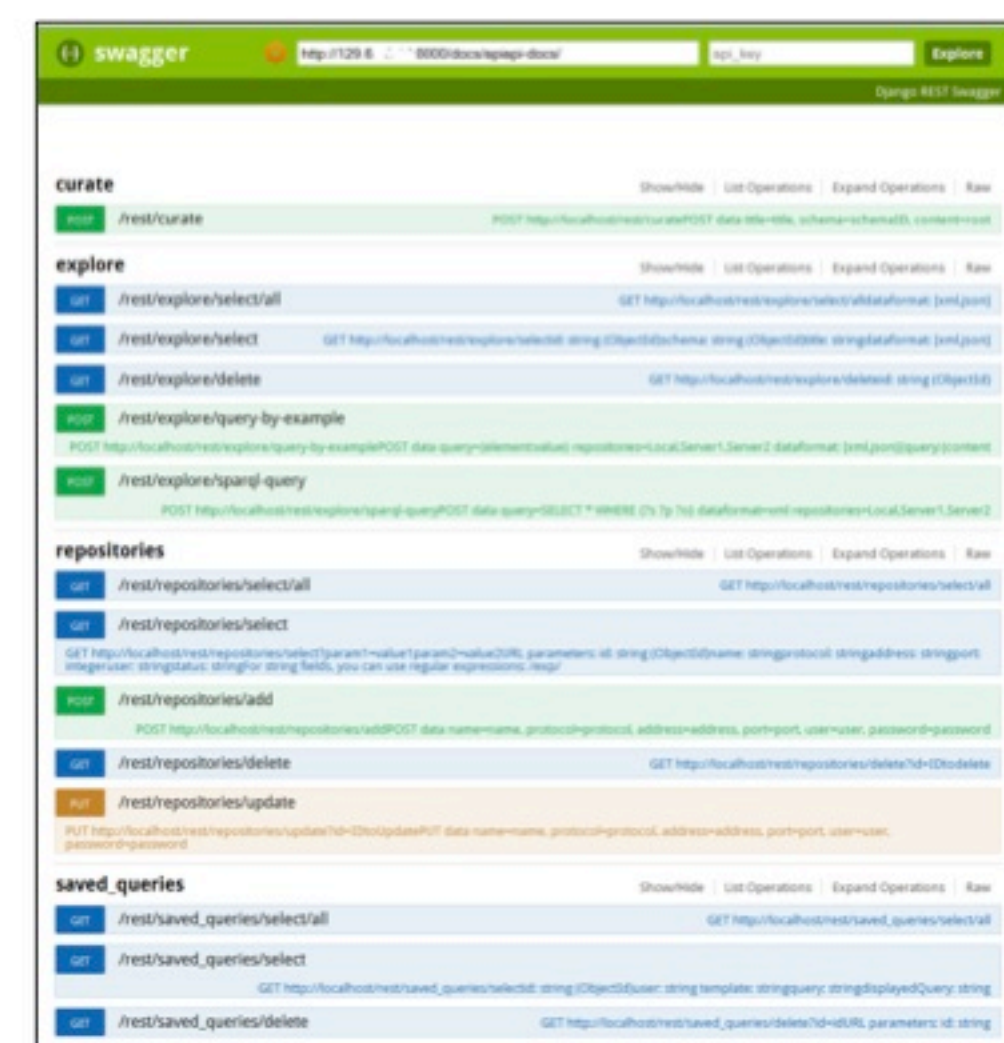
The ability of its underlying XML format to be transformed into virtually any other format using standard tools, gives the MDCS the ability to serve as a data source for a wide variety of existing materials informatics efforts that can span across projects, groups, and organizations.

Data Driven User Interface



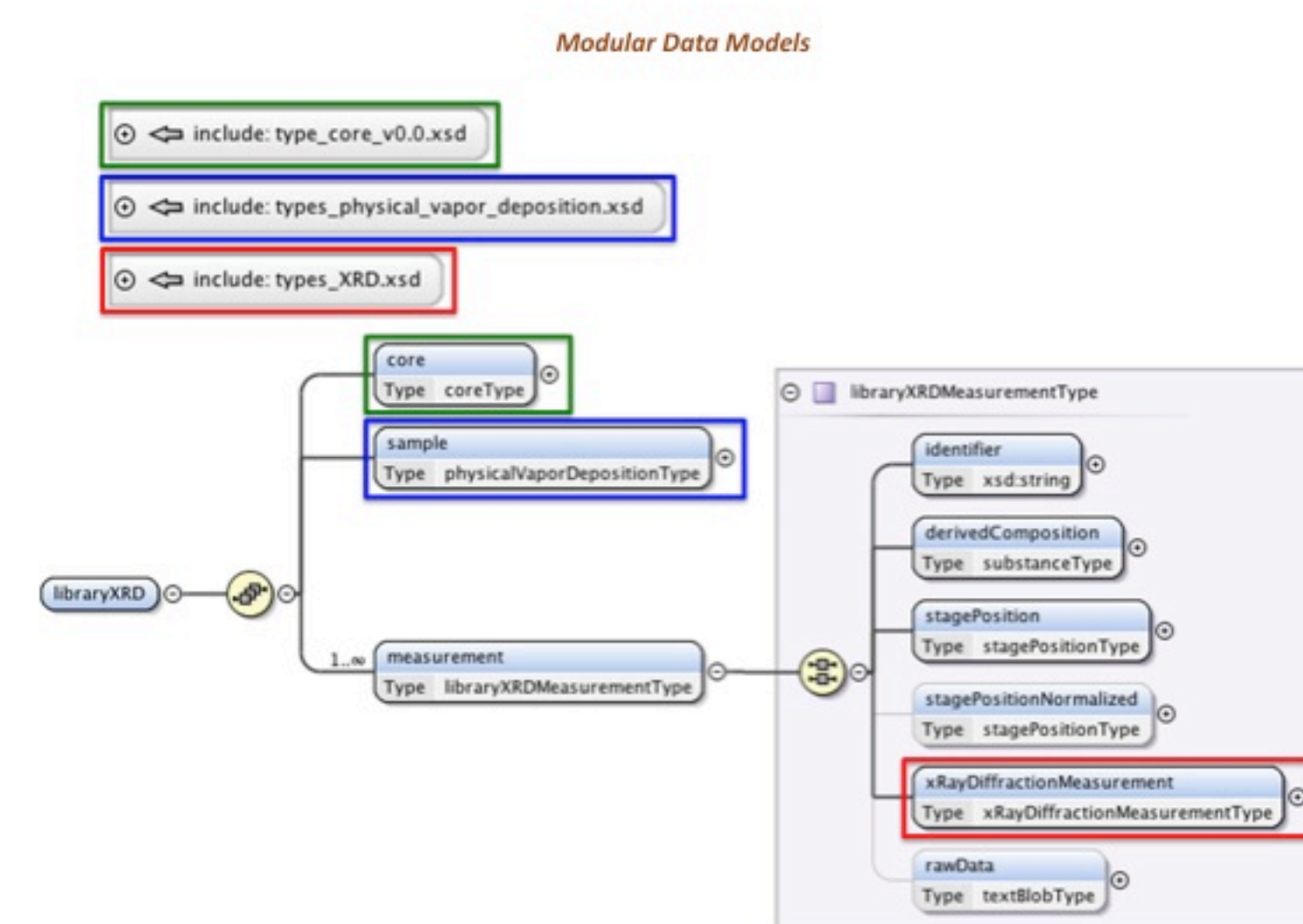
The MDCS has a flexible user interface generation system that dynamically creates data entry forms directly from the templates (XML schemas) used to organize the data.

MDCS REST API



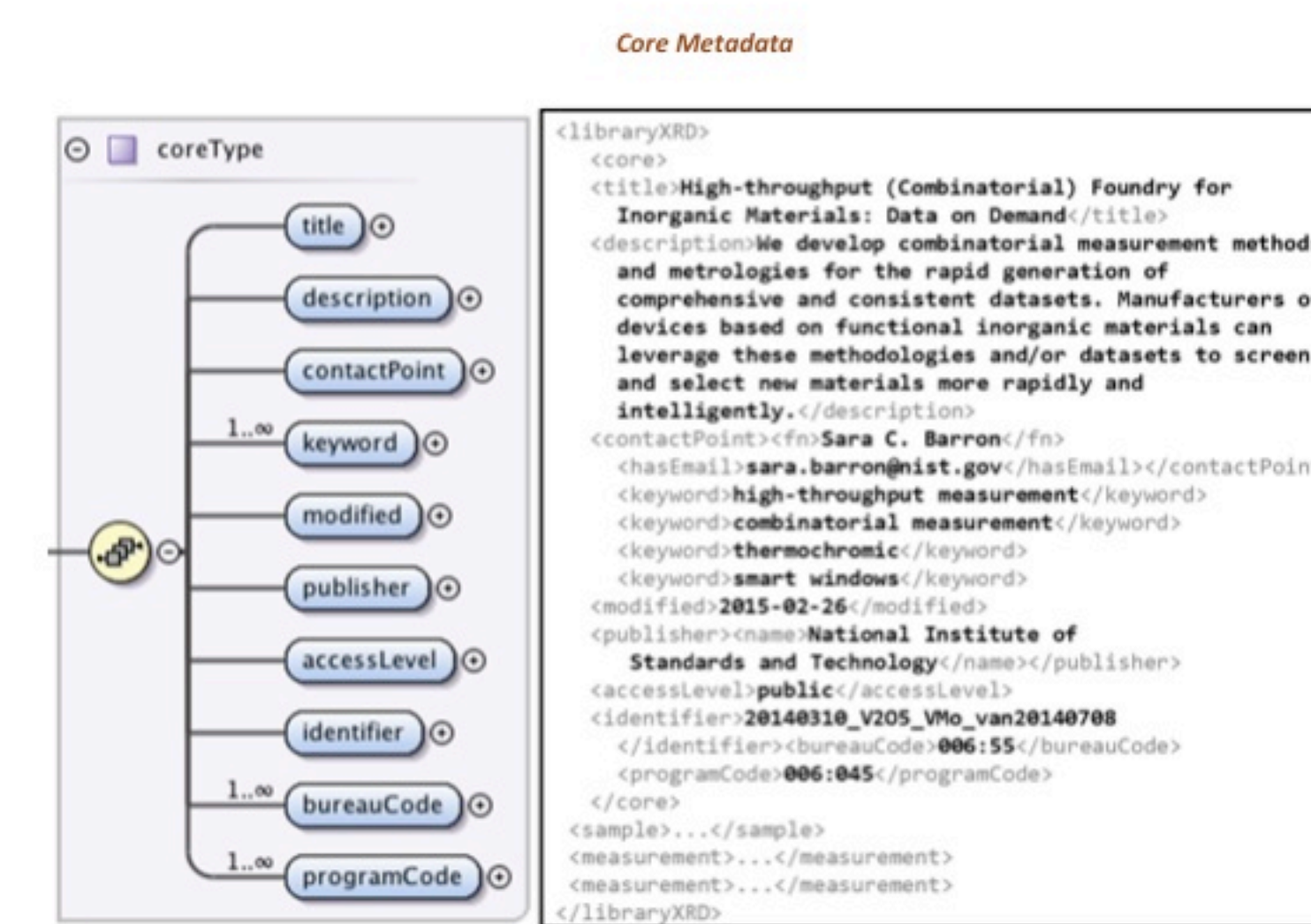
The MDCS provides a Representational State Transfer (REST) API that allows other software to directly interact with it over a network. MDCS functions are available via the API, allowing for full automation.

Enabling Modular Data Models



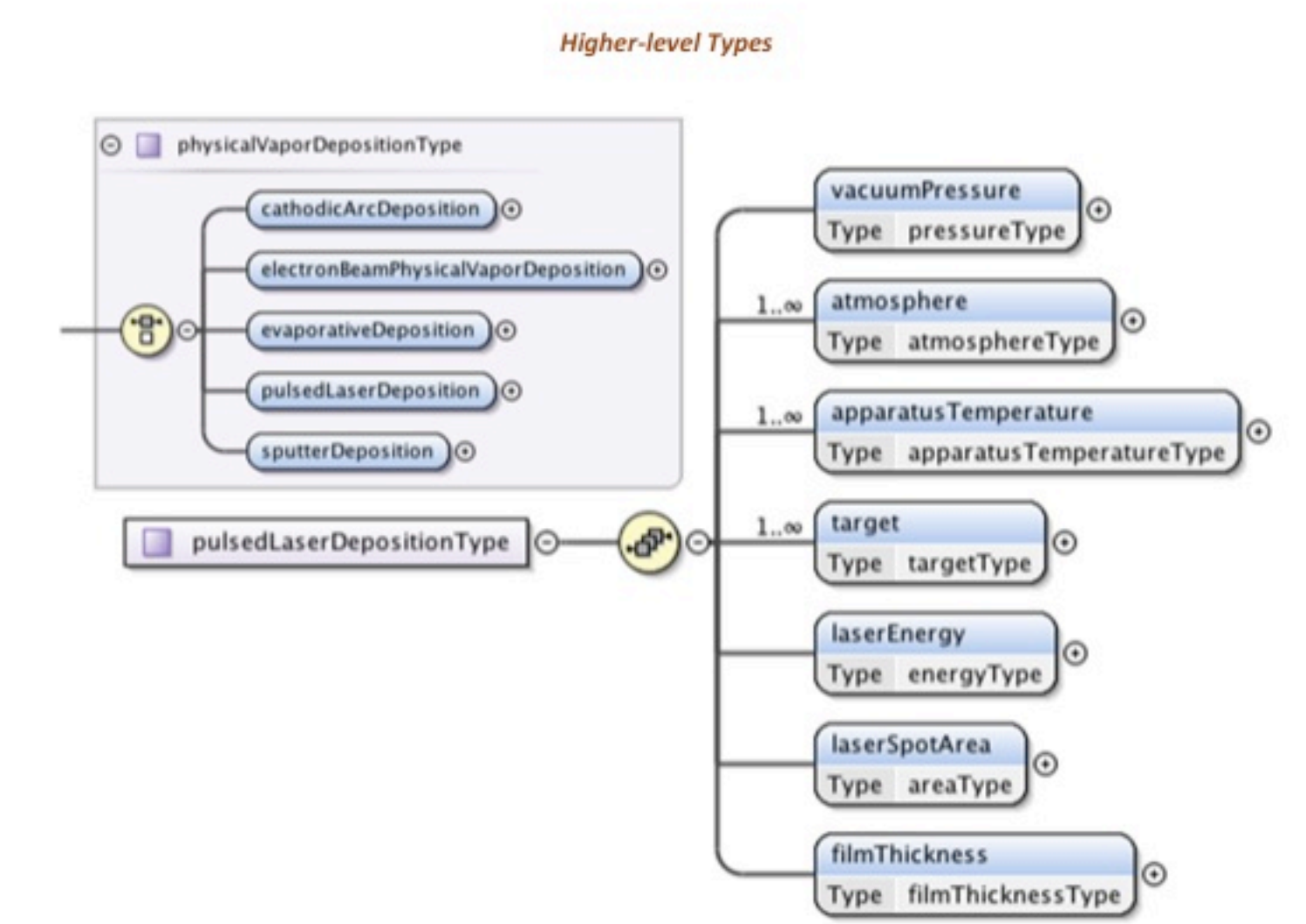
Complex data models can be composed from reusable templates allowing for groups of users and communities to develop data formats tailored to their needs. This example shows a new schema created using three schemas as types.

Capturing Metadata



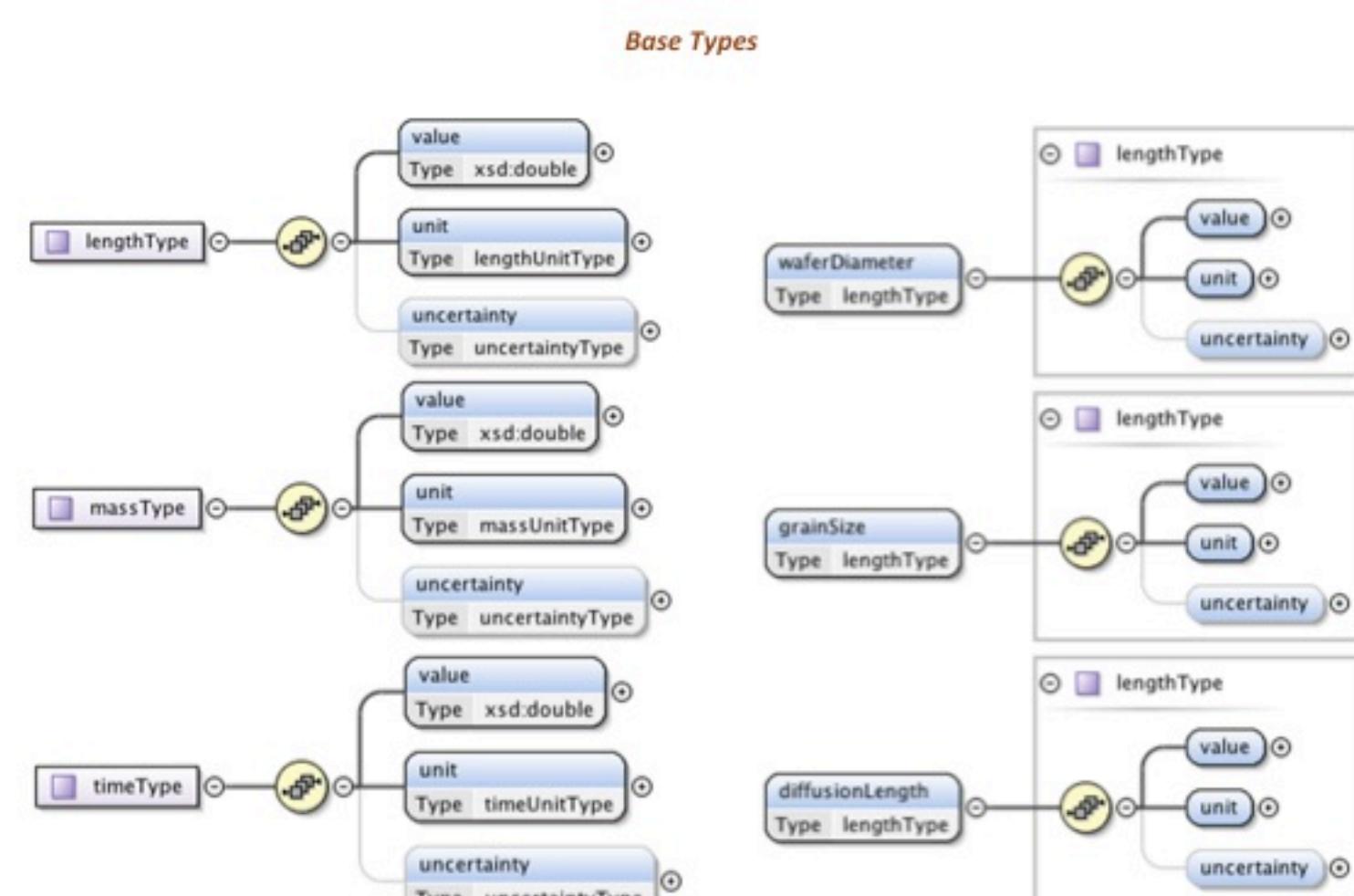
Metadata can be captured and packaged with the data allowing for the data to be fully described to enable its reuse.

Community Developed Templates



Communities can standardize templates for common use cases, leading to reusable data that is consistently formatted and described.

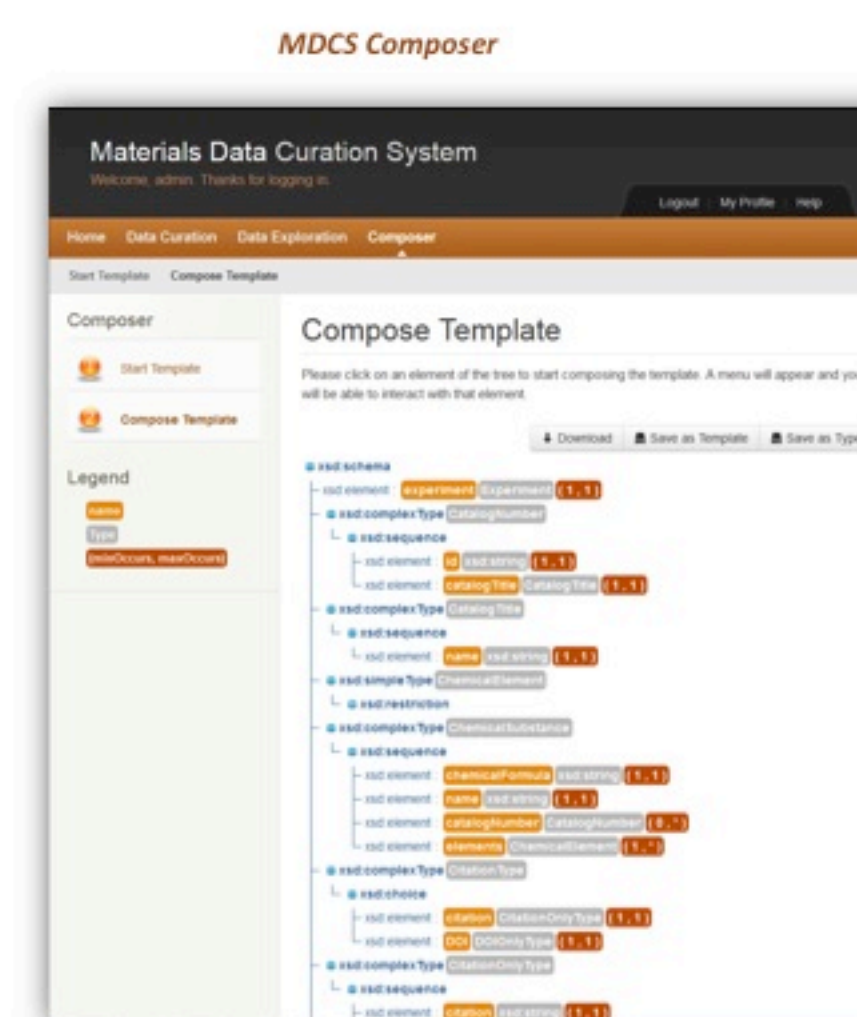
Reusable Data Types



The MDCS provides for modularity down to the lowest levels where reusable data types offer the promise of standard representations to simplify data access.

How are these Templates Created?

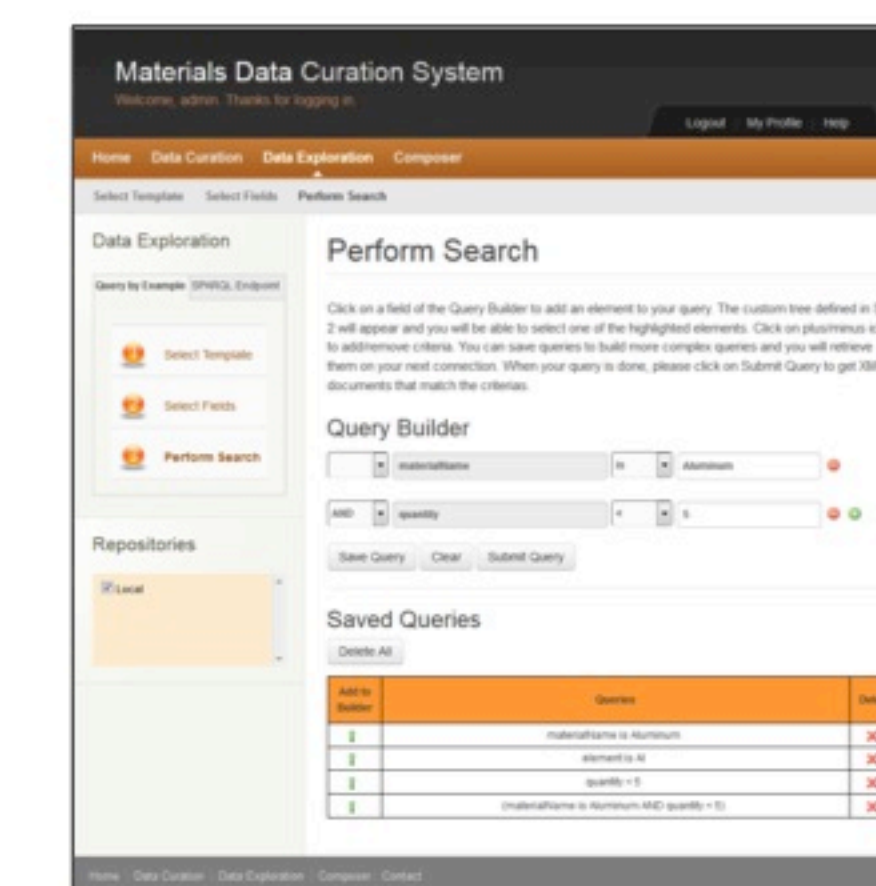
- Researchers can start with an existing template and modify it or can use existing collection of component templates to create new documents
- Future development will leverage public material registries to share templates and types across materials community



A Template Composer hides the complexity of the underlying XML Schemas allowing for users to quickly and intuitively design new templates.

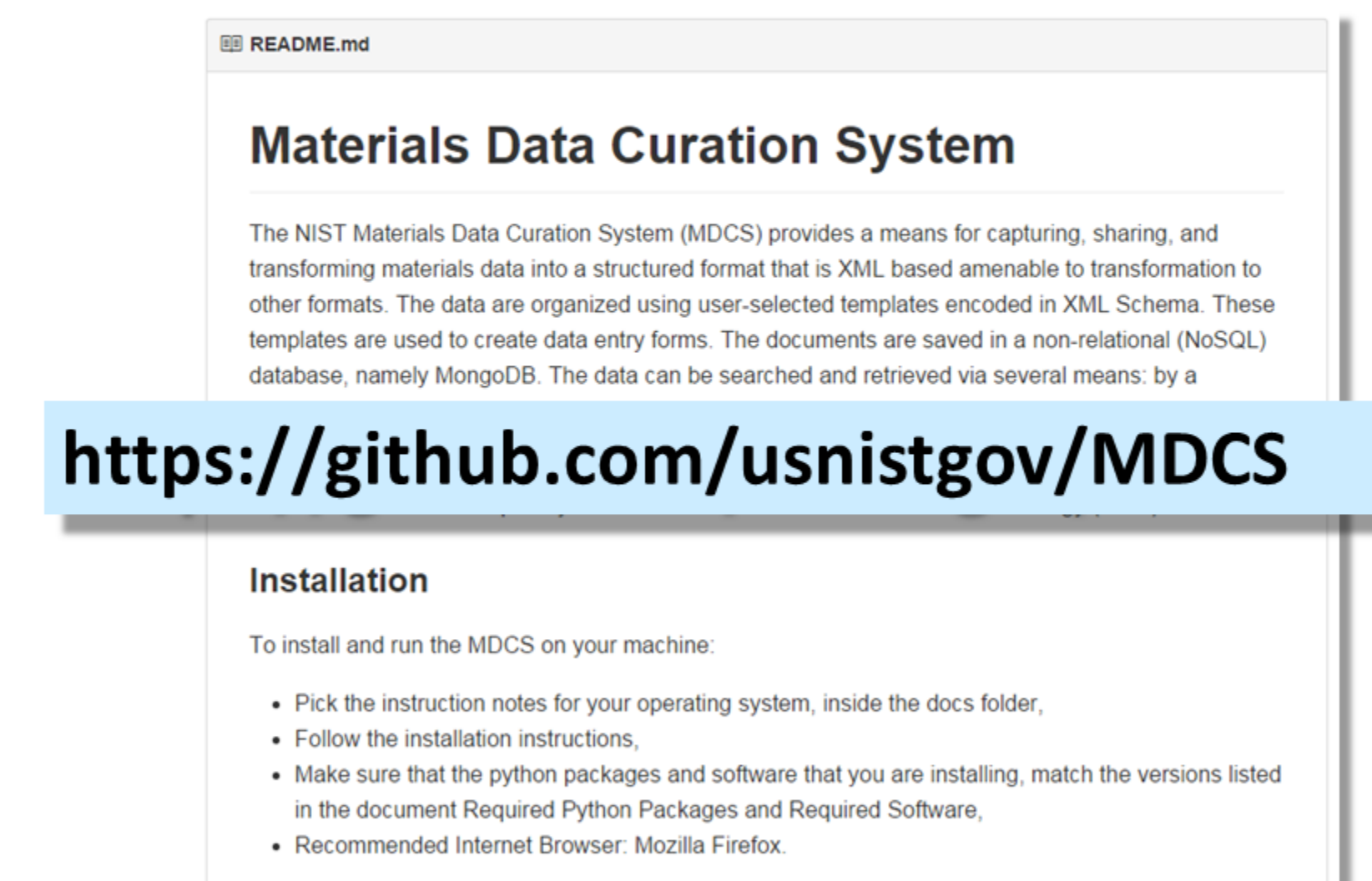
Data Exploration

- Select Template
 - Global Templates
 - User Defined Templates
- Select Fields
 - Specific fields to search against
- Perform Search
 - Query By Example
 - SPARQL Queries



The MDCS uses its templates to guide the search data. More complex searches can be accomplished using a SPARQL endpoint.

MDCS on GitHub



The MDCS is available from the NIST GitHub repository.