LA-UR-21-29682
Approved for public release; distribution is unlimited.

Title: SCHEMA ELEMENTS FOR GRANTA ANNUAL REPORT

Author(s): Brett, Jack Kevin
O'Neel, Jillian Cathleen
Schembri, Philip Edward
Waggoner, Rachel Marie

Intended for: Report

Issued: 2021-09-29

SCHEMA ELEMENTS FOR GRANTA ANNUAL REPORT

FY2021

Jack Brett

Jillian O'Neel

Philip Schembri

Rachel Waggoner

LA-UR-21-xxxxx
Date

Contents

Figures

Tables

1 Motivation

Granta: Materials Intelligence, also known as Granta:MI or Granta, is a commercial database software by Ansys, Inc. that is utilized by the Nuclear Security Enterprise (NSE) to organize and store materials data relevant to the complex. The software was first adopted by individual sites in the early 2000's, but in FY20, enterprise-wide licensing was purchased for the NSE by the Product Realization Integrated Digital Enterprise (PRIDE) program. Currently, there are 375 floating licenses available for users at all sites. There are also two shared instances of Granta; a classified production instance on the Enterprise Secure Network (ESN) and an unclassified development instance. Additionally, some individual sites such as Los Alamos National Laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), Sandia National Laboratories (SNL), Kansas City National Security Campus (KCNSC), and Savannah River National Laboratory (SRNL) host their own local Granta instances on their own networks that, in general, only their employees can access.

Table 1 provides a list of the existing local and shared Granta instances across the NSE.

Table 1. Granta:MI Instances in the NSE

	Unclassified	Classified
Local (Only accessible to site employees)		
LANL	X	X
LLNL	X	X
KCNSC	X	
SNL	X	X
SRNL	X	
Shared (Accessible to all NSE sites)	X*	X

*Development instance

Granta is a highly configurable software that allows for the creation of in-house materials databases. A 'database' is the highest organizational level in Granta and the complex has two main database types present across the local and shared instances. One database type is "Weapons Materials". These databases contain pedigree, testing, material property, material model, and risk information about materials relevant to the complex. Another database type is "Additive Manufacturing (AM)". These databases contain material batch, build, machine, and testing data for AM parts and materials.

Most sites started with Granta's stock schemas for storing general materials information and AM data; however these stock schemas are incomplete, leading sites to append them. For example, LANL has made modifications to the stock AM database to include: the addition of polymer AM attributes, the addition of Build Series organization, and the addition of materials characterization testing tables such as Differential Scanning Calorimetry (DSC) and 'Onion Skin' testing. Likewise, other sites have made their own changes to accommodate similar use-cases.

Because Granta provides many powerful and configurable options for storing data, there are multiple ways to store the same piece of data. This has led to the NSE sites modifying the stock Granta schemas in different ways, and many versions of the two database types now appear on both the shared and local Granta instances. For example, the Granta database experts at various sites have noted database schema disparities when it comes to attribute naming, attribute types (ex: drop-down menus vs. short text), link conventions, and much more. This situation is not unique to Granta; the Windchill/PDMLink product lifecycle management software is similarly configured differently at the different sites.

In addition to being incomplete, the stock Granta schemas are also undocumented. This leads to different users, groups, or sites interpreting attributes and uploading data in different ways, even when using un-modified schema.

Granta has the potential to be a very powerful resource to the complex; a universally recognized "gold standard" of materials information for the NSE. In the last few years, large programs such as the Advanced Manufacturing and Development (AMD) and Archiving & Support programs within NA-115, as well as the high explosives programs of NA-193, have recognized its benefits and have mandated that data management in Granta be included in work packages. However, having disparate and undocumented database schemas means that although NSE sites are using the same software to store the same kind of data, we cannot easily share or combine data to collaborate cross-site. Thus, we are not sufficiently meeting the intent of NNSA's data management mandates.

2 Objectives

The goal of this project is to develop and implement a workflow that can be used to align database schemas across the NSE. An aligned schema will allow the complex to easily share and transfer data between sites. It will also allow the complex to combine similar datasets for the purposes of machine learning and material modeling. Finally, an aligned schema on a shared server will allow multi-site collaborative projects, such as those in the Additive Coordination Team (ACT) AM program, to better manage and gain insight from project data.

This project proposes to achieve schema alignment through the systematic creation and documentation of common *Schema Elements* from which schema in all databases will be derived.

3 Approach

The mechanism by which the NSE will achieve schema alignment is through the creation of common *Schema Elements*. A schema element is a collection of attributes (data/metadata fields) that pertain to a particular category of data. Often these pieces of schema are reused throughout a database (like in the case of general test information such as 'Operator', 'Test Date', etc.) although this is not a requirement. All attributes in a database should belong to a schema element and all database tables should be made up of schema elements. In this way, schema elements are like fundamental building blocks. Figure 1 shows how schema elements can be combined to make the different tables needed, while ensuring the same data is being stored consistently everywhere in the database. Note that not all attributes are needed for each test type; however the attributes that they do share, like 'Test date', will be functionally identical in terms of name, type, conventions, etc. By gathering and standardizing all attributes into schema elements

and then creating tables from these schema elements, we will develop a documented, aligned NSE database schema into which all sites can map their data.

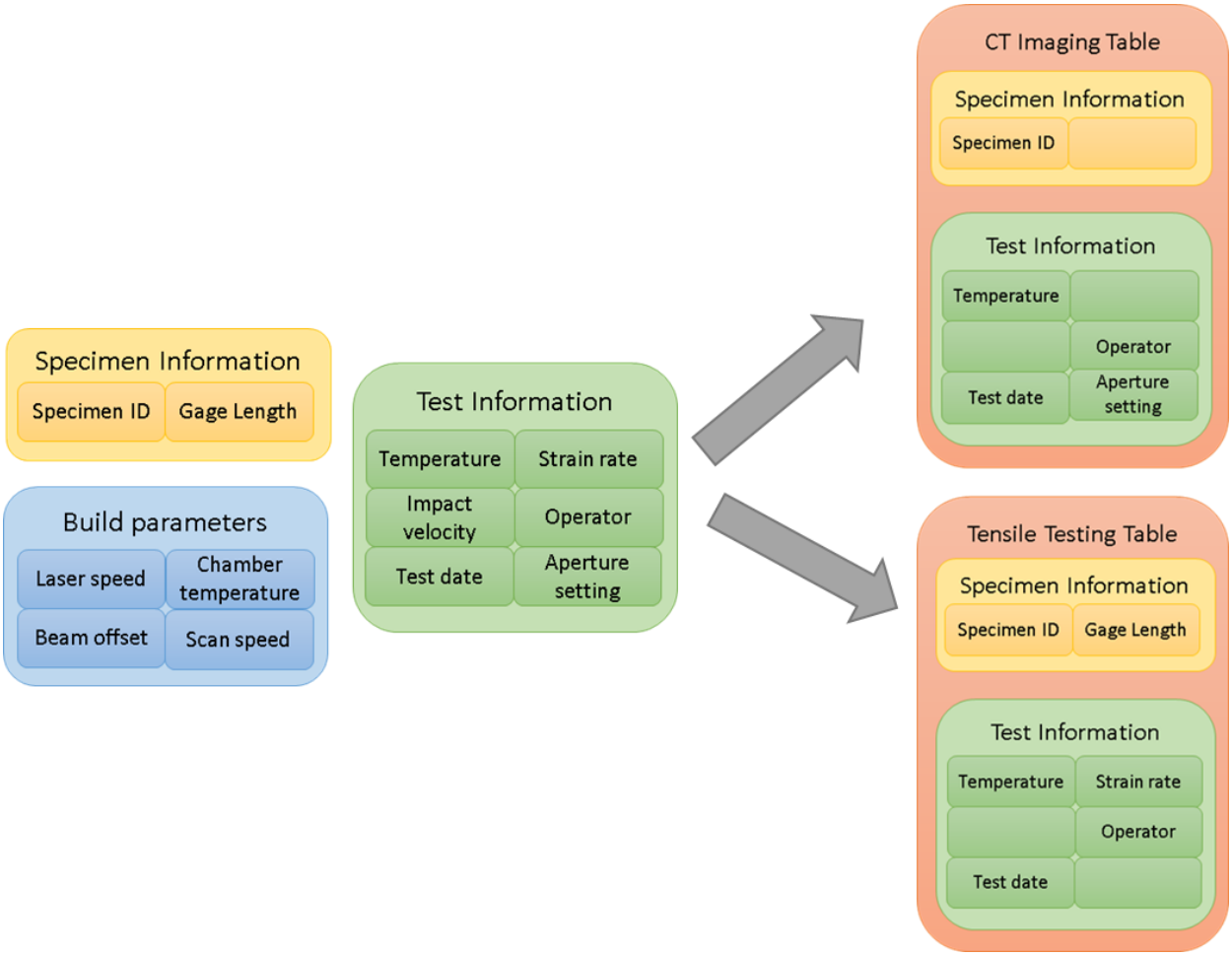

Figure 1. Three nominal schema elements containing attributes (left) can be put together to form two different testing tables (right).

Our preliminary approach for creating schema elements from disparate databases is to first gather a superset of all attributes being used and then to sort the attributes into proposed schema elements. With all attributes collected and organized, it should be more obvious where there are inconsistencies and duplication. For example, we may find that our database is using both a short text called 'Test Operator' and a discrete drop-down menu called 'Operator'; essentially the same data being stored differently depending on the database table. Once identified, these discrepancies can be discussed and a standardized, documented attribute can be created. When all attributes have been discussed, we will have a final schema element that can be used as a building block for creating tables. Finally, when all schema elements are completed and all tables are created, we will have a documented and aligned schema ready for use by all NSE sites.

4 Workflow

This section describes the workflow to be used for creating a unique set of schema elements, and tables built from them, from the disparate set of tables and attributes that exists in multiple legacy database schema. This workflow, which was the result of trial-and-error throughout the project, is not unique and can be modified as necessary. The motivation for creating and documenting this workflow is to maintain consistency in a collaborative environment, which helps to avoid confusion, incompleteness and duplication. Thus, it is important for all team members to understand and use the same workflow.

Before copying or creating any attributes or schema elements, the team should decide which schema elements and tables are the current objective since it is impractical to create all schema elements simultaneously. For example, during FY21 the team focused on creating the schema elements necessary to define thermal and mechanical testing tables. This meant that tables and attributes related to other data, such as materials availability risk and additive manufacturing (AM) builds, were not considered.

The workflow is illustrated in Figure 2 and is summarized as follows:

Step 1. Attributes are copied, table by table, from source (i.e. legacy) database schema to the Schema Elements Database to create *source tables* within the Schema Elements Database. This provides a complete list of all relevant existing attributes within a single database.

Step 2. In the Schema Elements Database, attributes are copied from the source tables to the Schema Elements Table. During this process, they are added to a layout whose name corresponds to the source table. This enables the origin of each attribute to be determined by simply sorting attributes by layout name in MI:Admin.

Step 3. Schema elements are created from the master attribute list, and these schema elements are combined to create tables. During this step, attributes may be removed or modified, and additional attributes may be added. This step also includes documentation and review, and therefore may be iterative.

There are steps subsequent to these three that address how the schema elements will be merged back into existing databases, such as the NSE Weapons Materials Database, and how change control will be addressed after schema elements and tables are released. While these subsequent processes have been discussed during this project, they have not been exercised yet and so will not be described in the current document.

Appendix A: presents the steps discussed above in further detail and also describes how layouts are used to keep track of the stages of completion of each schema element and table.

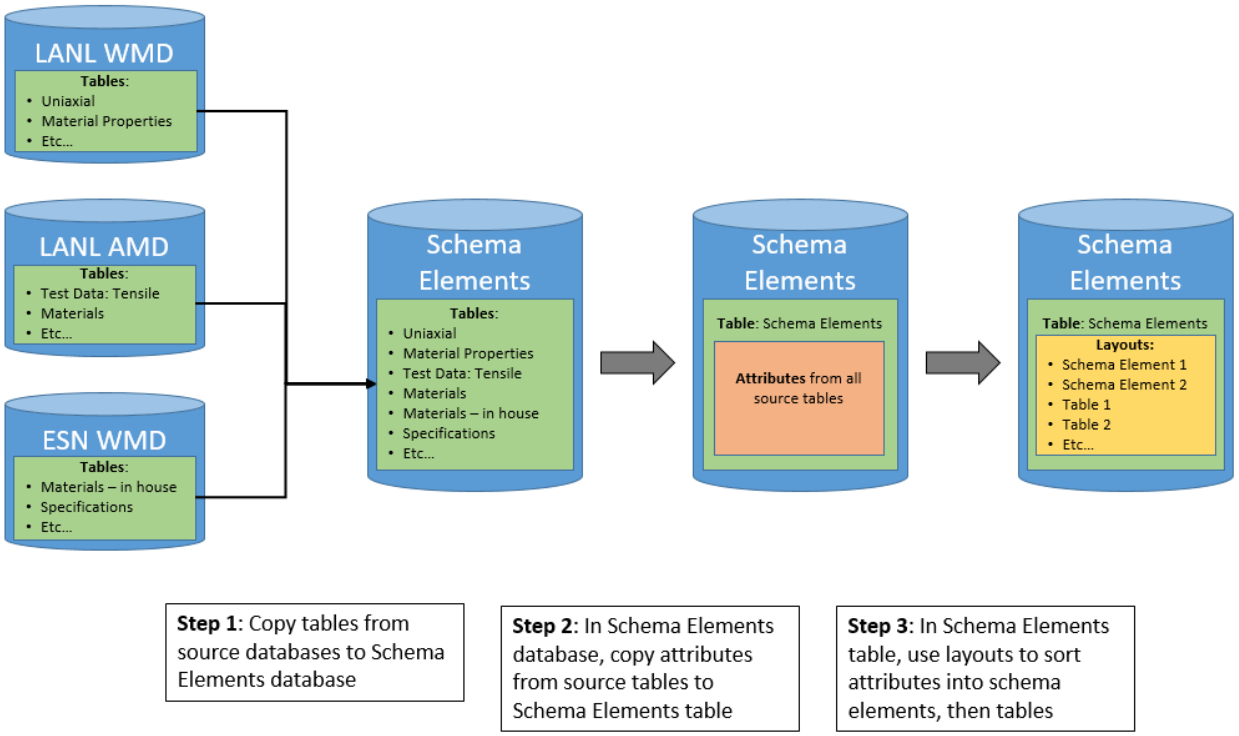

Figure 2. Diagram of the overall workflow used to create schema elements from legacy database schema.

5 Progress

The Granta schema development team has been creating and reviewing schema elements primarily related to thermal and mechanical testing. Table 2 shows the schema elements considered, their definitions, and information relating to the level of completion.

Through identifying schema elements, defining them, and placing attributes within them, the team has developed a workflow to maximize efficiency and collaboration. This is documented in Section 4. The team welcomes feedback from other sites on building on this workflow and streamlining it further.

The schema elements that are fully defined by the LANL team and are awaiting review and edits by the other sites are: Record information, Project information, and Data information. In completing these elements, the LANL team has defined the scope of the schema element, created layouts for them, assigned attributes to those layouts, and generated help files.

The schema elements that are currently being worked by the LANL team are: Test equipment, Test conditions, Test information, and Instrument parameters. These have been formally defined and drafted, and the team is currently assigning attributes to these elements and building their help files.

The remaining schema elements needed to complete existing thermal and mechanical testing tables will then be defined after the internal review of the schema elements stated above. After the LANL team has completed these schema elements, the other NSE sites will review them and make proposed changes to meet their use cases. To date, none of the newly created schema elements have been reviewed by the other sites.

Table 2. Progress for each schema element.

Schema Element	Definition	Attributes Identified?	List of Attributes	Help Files Complete?
Record Information	This schema element contains information about the record itself, including when it was created, modified, and reviewed, and who did those things. It does not contain any information about materials or data. As such, it is likely to be used the same way (all attributes and in the same order) in all tables (maybe even all databases). Some information in this schema element, like the date it was created and last modified, should be available in the Record Properties Granta objects. However, we are duplicating in these attributes because not all users have access to all record property information.	Yes	-Data sensitivity -Record create date -Record last modified date -Name of data originator -Name of creator -Name of modifier -Record owner -Record review	Yes
Project Information	This schema element contains information about the project that the record is a part of. Most of the attributes are actually links since the information about projects is contained natively in the Projects table, and we want to avoid duplication.	Yes	-Project ID -Project (link to Projects table) -Project smart link (link to Project table) -Project information -Project notes	Yes
Data Information	This is information about the data but not the data itself. Note that we include the raw and analyzed data files as being 'about' the data since they are kept for traceability purposes and are not intended to represent the trusted (reviewed) source of data to be used by data consumers. This schema element also includes information intended to characterize the 'quality' of the data in the record.	Yes	-Data quality rating -Raw data file -Corrected raw data file -Analyzed data file -Final data file -Other data file -Data information notes	Yes

Test Equipment Information	This schema element contains attributes with information about the equipment used for performing the characterization test to which the record corresponds.	In progress	-Test Equipment Information (tabular) -Test equipment ID -Test equipment (link from test table to Test Equipment table) -Furnace type -Balance gas flow -Specimen gas flow -Expansion probe type -Expansion probe diameter -Preload force -Test equipment notes	No
Test Conditions	These are attributes that specifically describe the state of the sample material (temperature, heating rate, strain rate, pressure, etc.) during the test. Note that some of these attributes may overlap with Instrument Parameter attributes. For example, Instrument Parameters may contain an attribute for Heating schedule since that is an instrument setting, while Test Conditions may contain an attribute for Temperature range, which describes the range of states the material experienced during the test and should be consistent with Heating schedule.	In Progress	In progress	No
Test Information	These are non-instrument-specific attributes describing the test such as date, time, operator, etc, (including site, location, etc., which were previously in the Source of Testing element). These attributes are likely to be shared among many testing tables, but this is not a requirement.	In Progress	In progress	No

Instrument Parameters	These are attributes that describe the setup of the instrument sufficiently for an operator to reproduce a test. This will be a very long list because it will contain attributes describing the setup of *every* type of instrument. Note that some of these attributes (e.g. Sample rate) are used by multiple instruments, so we can't have an element for each instrument.	In Progress	In progress	No
Specimen Information	In Progress	In Progress	In progress	No
Test Results	In Progress	In Progress	In progress	No
Analyzed Test Results	In Progress	In Progress	In progress	No
Calibration Information	In Progress	In Progress	In progress	No
General Material Information	In Progress	In Progress	In progress	No
Machine Information	In Progress	In Progress	In progress	No
Material Properties	In Progress	In Progress	In progress	No
Project Definition	In Progress	In Progress	In progress	No

6 Conventions and Best Practices:

In discussing attributes and curating schema elements, the team has begun to develop conventions, standards, and best practices for Granta. These conventions and best practices are intended to be utilized across all schema elements. Help files will reflect these decisions for specific attributes as they are relevant. Another best practices document will be developed by the schema elements team as a living document going forward since this list is expected to grow considerably as the project progresses. Figure 3 shows an example record in Granta, illustrating many of the conventions described below.

1. **Attribute Naming**:
   - Attributes will not have identical names if they are not intended to be used in a functionally identical way.
   - In a name with more than one word, the first word is capitalized and the rest is lowercase
2. **Notes Attributes:** Each schema element will contain a long text attribute called "*Schema Element Name" notes* (ex: Data information notes) that is intended to serve as a catch-all notes attribute for that section of the table.
3. **Linking conventions:** Links between records will, in general, consist of the objects listed below. While the schema element should contain all objects, implementation into a production database table may involve down-selecting objects. For example, either a static OR a smart

record link should be implemented based on whether the links will be one-to-one or one-to-many, but not both.

- A short text attribute corresponding to the unique 'ID' in the table that is being linked to. For example, project records contain a unique, short text *Project ID* or testing series records contain a unique, short text *Testing Series ID*. This is useful for automating static or smart link generation.
- A static record link
- A smart record link
- A tabular attribute with the previously mentioned short text ID attribute as the linking value and relevant linked columns defined.

4. **Naming of people:** To aid in searching, people's names in short text attributes must be written consistently. The following convention is proposed.
   - First name and last name as they appear in the OneID phonebook.
   - For NSE personnel, a unique identifier in parentheses following the name. This could be:
     - Z# for LANL employees
     - email address for non LANL employees since this will also show what site they are from
5. **Reviewer Convention:** Rather than using separate attributes for each different kind of record review, which isn't flexible enough, a long text attribute will be used with the format described below. Each reviewer will append with their own entry on a new line. Current review types include "Technical Review" for data accuracy, "Granta Review" for conformation to database conventions, and "Classification Review" for review by a Derivative Classifier.
   - First name and last name (unique ID) - date - review type
     - Ex: Jack Brett (338217) - 3/8/2021 - Granta Review
6. **Files in Data information section:** Files in this schema element will be stored in individual attributes rather than in a flexible, tabular 'Files' attribute. This facilitates ease of uploading.
7. **Guidelines on discrete attributes**
   - Discrete lists can be difficult to maintain; therefore, a database administrator should consider the use of a short text (with population guidance in the help file) before choosing to use a discrete type attribute
   - If a discrete list is ordered, follow Granta's documentation to list the choices in the appropriate order (i.e., best-to-worst or worst-to-best) and ensure that the "Ordered" box is checked in MI:Admin.
   - If a discrete list is not ordered, sort the values in the discrete type in alphabetical order ("Arrange by Name" in MI:Admin) and ensure that the "Ordered" box is *not* checked.
8. **Unit Systems**
   - Attributes that have units should be assigned the applicable unit in the mm-N-s consistent unit system as the default unit.
   - As units are assigned to attributes, admins should verify that the selected unit is consistently and correctly set-up in MI:Admin. This includes using "Derived from other units" whenever possible and correctly inputted "Unit System Equivalents".
   - When reviewing schema elements, units and equivalent units will also be reviewed for consistency.
9. **Help Files:** A help file is required for each attribute and must be in the standardized format.

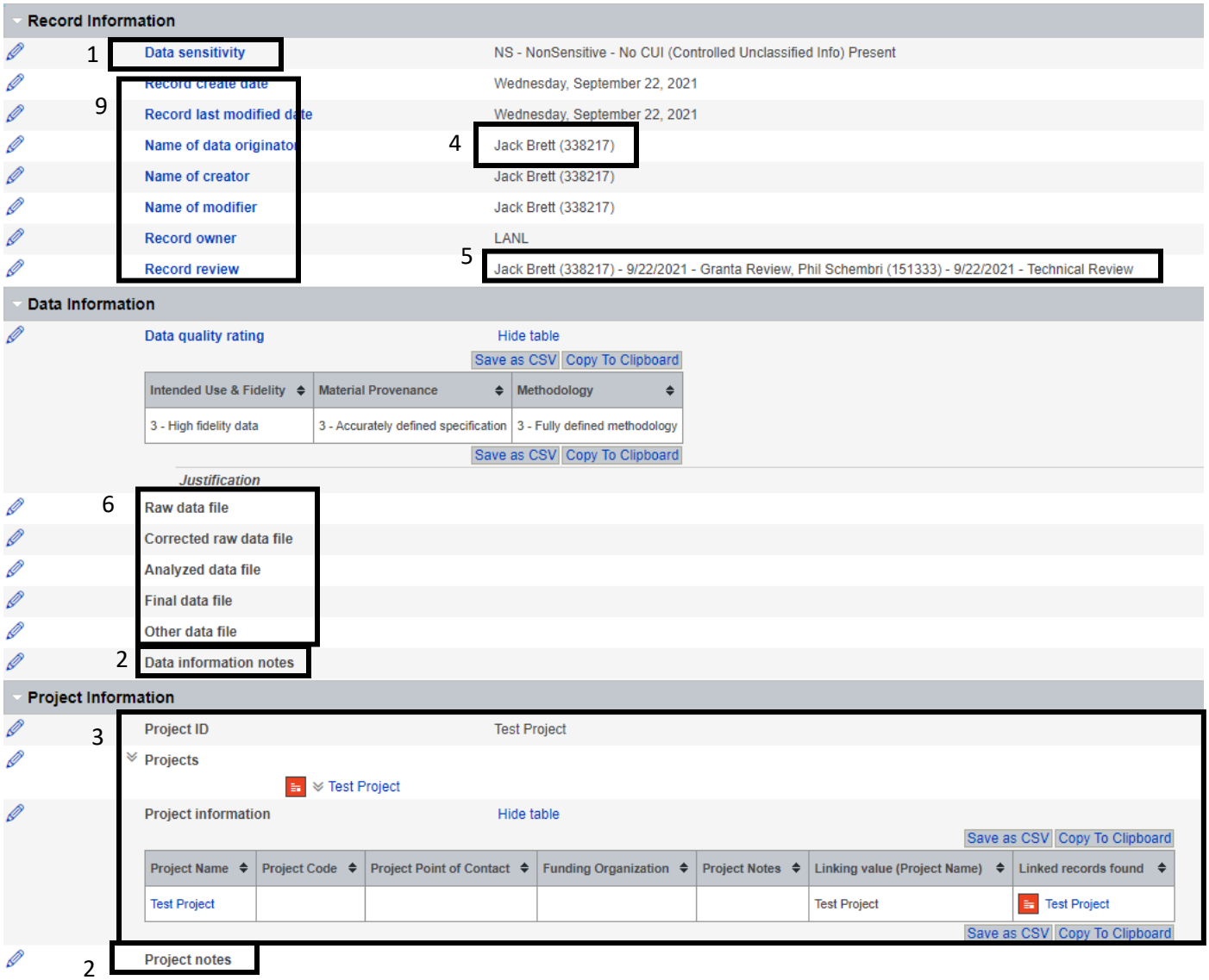

Figure 3. Example Granta record to show conventions and best practices from Section 0. The specific conventions listed below are illustrated, with numbering corresponding to the numbered conventions in Section 0:

1. Attribute Naming: The first letter of the first word of an attribute is capitalized.
2. Notes Attributes: There are notes attributes at the end of each schema element.
3. Linking Conventions: Every attribute that has a linking capability needs an ID, a link to the record, and a tabular link.
4. Naming of People: All people, in this case a LANL employee, is named using the convention: name (Z#).
5. Reviewer Convention: Reviewer names follow the convention: name (Z#) – date of review – type of review.
6. Data Information: The data information schema element was limited to multiple file attributes instead of a large tabular attribute.
9. Help Files: Attributes written in blue means that a help file has been attached to that attribute.

7 Future work

Effort in FY21 was focused on establishing the processes and workflows for creating and documenting schema elements. Future work will make use of this framework to continue defining schema elements applicable to tables in both the AM and WMD databases. Sandia National Laboratories (SNL) and Kansas City Nuclear Security Campus (KCNSC) both expect a modest level of funding for this project in FY22, and so they will both review/edit existing schema elements and add schema elements for database tables they have created locally but that don't exist in the Schema Elements Database. Furthermore, NA-193 has made the decision to use Granta for high explosives (HE) AM, property and test data, and a representative from that organization will both present use cases for additional tables and contribute as a subject matter expert to help create the necessary schema elements. The project would greatly benefit from the input of subject matter experts at all sites, so all sites are encouraged to seek some level of funding to contribute to this project, perhaps from one of the many programs that will benefit from it.

The following sections describe the objectives the project expects to accomplish over the next two years, but these will evolve as stakeholders from other sites become involved in the planning process.

7.1 Objectives for FY22

7.1.1 *Onboard new project members*

SNL, KCNSC, and NA-193 will be contributing to the project in FY22, and they will need to have access to the Schema Elements database and understand the workflow for creating and documenting schema elements. Personnel at both SNL and KCNSC obtained access to the Schema Elements database in FY21, so it remains to ensure NA-193 gains access. This report is intended to serve as the primary documentation for the schema elements workflow, and so it will be prerequisite reading for all Schema Elements team members. Feedback from new team members will be used to improve this documentation for subsequent new members.

7.1.2 *Complete schema elements for prioritized testing tables*

Schema elements relevant to the highest priority materials testing tables will be completed in FY22. Addressing testing tables (rather than tables related to AM builds, properties, material models, or material risk) should result in a larger number of completed tables since many testing tables share common schema elements (e.g. Data Information, Test Equipment Information). A proposed list of highest priority testing tables, which can be reworked based on input from other sites as necessary, is:

- Uniaxial (i.e. low-rate tensile or compression) testing
- Split Hopkinson Pressure Bar (i.e. medium-rate tensile or compression) testing
- Gun Experiment (i.e. high-rate mechanical response) testing
- Thermal Mechanical Analyzer (TMA) and Dilatometry testing
- Differential Scanning Calorimetry (DSC) and Modulated DSC testing
- Thermal Conductivity Analyzer (TCA) testing
- Fourier Transform Infrared Spectroscopy (FTIR) testing

Progress on each of these tables should consist of:

- Drafting, discussing, reviewing, and documenting all relevant schema elements
- Combining the schema elements to create tables

- Copying these tables to the classified production server on the enterprise secure network (ESN) and to the unclassified development server
- Uploading at least one dataset to test the schema

7.1.3 *Evaluate and implement Data Updater tool*

In FY21, database schema was moved between existing databases and tables either by copy/pasting individual attributes or by manually creating them. At the request of NSE customers, Ansys-Granta created a tool, called Data Updater, to automate the process of copying/synchronizing schema. Implementation of this tool would increase the productivity of the project, so its evaluation, and implementation if it meets project needs, should be included in the scope of FY22 work.

7.1.4 *Create a joint project proposal for continued work*

The schema created in FY22 will be only a small subset of the schema necessary to support AM, At-Risk Materials, HE, and other programs that generate materials data throughout the NSE. A coordinated (among all stakeholders) and appropriately funded project needs to be planned that addresses the remaining schema, along with the tools, documentation, and support needed to allow data generators to upload both legacy and newly-generated materials data. During FY22, the Schema Elements project team will draft and release a joint (among the current stakeholder programs of NA-193 and AMD) proposal outlining a multi-year project plan and justifying the necessary funding.
This project plan should contain:

- A prioritized list of all data types (i.e. databases and tables) that comprise the scope of the project
- A description of the tools and documentation needed by data generators to effectively upload to the database
- A description of the most common use-cases of data consumers to ensure the data is uploaded in a way that will provide the most value
- A plan for migrating data in existing local Granta MI instances (with disparate schema) into a centralized Granta MI instances (unclassified and/or classified) containing the newly created schema
- The level of effort and funding required at each site to achieve the objectives

7.2 Objectives for FY23

Objectives for FY23 will be better defined in the Schema Elements joint project proposal written in FY22. At this time, we expect the objectives to consist of:

- Continuation of schema element development
- Implementation of new tables on shared Granta MI instances as the necessary schema elements are created
- Creation and release of tools for uploading data to the new tables
- Migration of data from existing repositories (e.g. local Granta MI instances) to centralized Granta MI instances

Appendix A:

This appendix contains a more detailed description of the workflow steps described in Section 4.

Step 1: Copying tables and attributes from legacy databases

Once it has been determined which tables in the legacy databases are needed to create the schema elements for the current objective (e.g. thermal and mechanical testing tables), the tables and their attributes from the legacy databases are copied to the Schema Elements Database. This creates new *source tables* in the Schema Elements Database. MI:Admin is the primary tool used to copy tables from source databases to the Schema Elements Database. Copying these tables is a relatively manual process, and the following points are lessons learned during FY21 that have helped to make the process more efficient:

A. There is no facility within MI:Admin to copy a table from one database to another, so a blank table with the source table name must first be created in the Schema Elements Database.
B. Attributes are copied from the source databases into the new source tables rather than directly into the Schema Elements table to avoid MI:Admin errors when duplicate attribute names are encountered. When copying *across database servers*, duplicate attribute names cause an error message and prevent batch copy/paste actions from proceeding, whereas when duplicate attribute names are encountered during a batch copy/paste between tables *within the same database*, MI:Admin simply appends a '(2)' to the name of the copied attribute with the duplicate name and the copy process proceeds.
C. 'Simple' attribute types, like point, range, short text, and long text, can be directly copy/pasted from the source database to the source table in the Schema Elements Database. Opening two instances of MI:Admin, one showing the source database and one showing the Schema Elements Database, facilitates this process.
D. Before copying discrete attributes, the relevant discrete types should first be copied or recreated from the source database to the Schema Elements Database.
E. 'Complex' attribute types will need to be recreated mostly manually. When copy/pasted from the source table, functional and tabular attributes become an empty "shell". Functional data will need to have the parameters reassigned manually, and tabular data will need to have the columns recreated manually. The team decided to leave these attribute types incomplete during this step of the process and instead chose to add the necessary parameters/columns during Step 3, when attributes are being sorted into schema elements and discussed/reviewed.
F. If an attribute being copied is defined using a unit that doesn't currently exist in the Schema Elements Database, MI:Admin will show an error message and will not allow a batch copy to proceed. A workaround for this scenario has not yet been found – the unit must be created in the Schema Elements Database and the copy reinitiated.

Step 2: Copy attributes from source tables to Schema Elements Table

With the attributes relevant to the current objective (e.g. thermal and mechanical testing tables) copied from source databases to source tables in the Schema Elements Database, each of these attributes are then merged into the Schema Elements Table. Care must be taken to handle any duplicate attribute

names and to ensure the source table name of each attribute remains associated with that attribute. The following practices were established in FY21 to help with these issues:

A. Before a set of attributes is copied from a source table to the Schema Elements Table, every attribute in the Schema Elements Table should be included in at least one layout. Layouts in the Schema Elements Table, and their uses, are explained in more detail in the Layouts section below.
B. Immediately after a set of attributes is copied from a source table to the Schema Elements Table, those attributes should be the only attributes in the table that are *not* part of a layout and should therefore be easy to identify, e.g. by sorting attributes in MI:Admin by layout. These attributes should then be added to a *source* layout (see Layouts section below) named for their source table.
C. If the name of an attribute being copied to the Schema Elements Table is a duplicate of an attribute already existing in the table then it will appear in MI:Admin with an appended '(2)'. This naming conflict must be resolved before continuing. One option is that the attribute being copied over is functionally identical to the preexisting attribute. In this case, the preexisting attribute should remain and the appended '(2)' attribute should be deleted. The preexisting attribute should be added to the new source layout. Another option is that the attribute being copied over is meant to function differently or hold different information than the preexisting attribute with the same name. In this case, the newly copied attribute should be renamed to something unique and added to the source layout. In the event that another name is not immediately obvious, the attribute may temporarily keep its original name with an appended '- Source Table'. Note that the name and attribute type of every attribute can be re-evaluated during Step 3, so the decision made here may not be final.
D. Once all newly-copied attributes have been added to a layout, the Schema Elements Table should once again be in a state in which no attribute names are duplicated and every attribute belongs to at least one layout.

Step 3: Sorting attributes into schema elements and tables

Steps 1 and 2 are primarily required for bookkeeping, while the current step addresses the more crucial work of creating schema elements and ensuring they are consistent with other schema elements and adhere to the conventions of the database (see Section 0 for further discussion of conventions).

Recall that this workflow (i.e. these three steps) are performed with a particular objective in mind – a table or set of related tables, and the schema elements that comprise them. After all attributes related to this objective are in the Schema Elements Table, the schema elements are defined, documented, and reviewed. Some important points to consider are as follows:

A. This is partly a subjective process, although the need to adhere to conventions should reduce the number of possible solutions. Thorough discussion among the schema elements project team, as well as technical SMEs when necessary, is the mechanism we relied on in FY21 to uncover the best solutions.
B. The number and definitions of schema elements comprising a table may change during the process. For example, the schema element initially expected to be called Test Information evolved from that single schema element into multiple: Test Conditions, Instrument Parameters, and Test Information.
C. There is no requirement to keep all attributes copied from the source tables. Attribute types and names can be changed as necessary to meet the needs of the schema element and adhere to

conventions. (If existing data exists in a legacy database, the significant changing of attributes may result in challenges when mapping this data from old to new schema. However, the team in FY21 has taken the approach to defer this problem and assume that with help from computer scientists – e.g. using the python STK – it will be less difficult mapping data than dealing with inconsistent schema.)

D. Some other points the team considered when determining whether an attribute should be included in a schema element include:
   - How is this attribute being used by data consumers? E.g. is it an independent variable (like test temperature or strain rate) that someone would want to examine the effects of? If so, a simple point or range attribute might be preferred.
   - Does this fit with current conventions? If not, should the attribute be changed or the conventions?
   - Is the attribute name unambiguous, especially considering similar attributes? (e.g. **Specimen width** vs. **Specimen thickness**)
   - How will the data be uploaded? E.g., in standard Excel templates created by MI:Viewer and Toolbox, functional and tabular data are on separate sheets; will this cause problems uploading the data?

Step 3 also includes documentation and review of the schema elements and tables such that they are mature enough to be used in a production database. Documentation of a schema element consists of three types:

- **Documentation of attributes**. The primary tool for this is the html 'help file', which transforms the attribute name in MI:Viewer into a weblink that opens a page of documentation for the user. The documentation can be written to help both the consumer of data and the person responsible for uploading the data. So, for example, it can describe conventions such as those for naming materials.
- **Documentation of schema elements**. The motivation for documenting a schema element is to clearly distinguish it from similar schema elements. This makes it clear which attributes belong in it, and which don't. For documenting schema elements definitions in FY21, the team used Confluence, but this may not be practical for FY22 since there is no cross-site instance of Confluence. A Word document on the Granta SharePoint may be more practical going forward.
- **Documentation of tables**. The documentation of tables should capture standards and best practices for uploading data that are not already covered in the attribute documentation. In particular it should include description of the relationships between the table and related tables (e.g. between testing tables and the Testing Series table and between the Calibrated Material Models table and the Materials table), which are useful for those uploading data.

Review of schema elements and tables should be performed first by the principal investigator (PI), which may be an individual or a team, drafting the schema element. Following this 'internal' review, an 'external' review should be performed by all stakeholders (i.e. sites that may use the schema element or table) and any revisions incorporated. Finally, once the revised schema element or table has been reviewed by all stakeholders, it is ready for release into a production database.

The subsequent steps of merging finished schema elements back into existing databases, such as the NSE Weapons Materials Database has been discussed during this project, but has not been exercised yet and so will not be described in the current document.

Layouts

Layouts in the Schema Elements Table are used to organize attributes into schema elements. Similarly, once schema elements have been created, the schema elements layouts can be combined (utilizing the ability to copy/paste layouts using Excel) to create a layout representing a table. Thus, there are three different types of layouts in the Schema Elements Table:

- **Source**: layouts into which attributes are added to keep track of the table from which they were sourced.
- **Schema element**: layouts representing individual schema elements
- **Table**: layouts that combine schema elements, or subsets of them, to represent the attributes comprising a production database table

To differentiate these layout types, and to keep track of the stage of completion of each type of layout, the following layout naming convention is used, where the stages of completion for each layout type are defined in Table A- 1**Error! Reference source not found.**:

[Layout type] (stage of completion) layout name

Table A- 1. Stages of completion for each type of layout.

Layout type	Stages of completion
Schema element	• **draft** (attributes and documentation still being added) • **complete** (attributes and documentation ready for review by PI organization(s); review is in progress) • **internally reviewed** (PI organization(s) have reviewed the schema element; changes are complete; external review in progress) • **externally reviewed** (all sites have reviewed the schema element; changes and final review are in progress) • **released** (schema element is ready for use in tables; further changes incorporated through change-control process)
Table	• **draft** (some schema elements have been combined into a table; not all necessary schema elements are available yet) • **complete** (all schema elements have been added and the table is ready for review by PI organization(s); PI review is in progress) • **internally reviewed** (PI organization(s) have reviewed the table; changes are complete; external review in progress) • **externally reviewed** (all sites have reviewed the table; changes and final review are in progress) • **released** (table is ready for use in databases; further changes incorporated through change-control process)
Source	• **working** (attributes are actively being incorporated into schema elements by PI) • **archived** (all attributes have been considered for addition into a schema element; layout is kept for archival purposes)