

LA-UR-22-25165

Approved for public release; distribution is unlimited.

Title: Attribute types, Granta limitations, and return on investment

Author(s): O'Neel, Jillian Cathleen
Schembri, Philip Edward
Brett, Jack Kevin
Hughey, Lisa Michelle

Intended for: Communication with commercial software developer (Ansys-Granta)

Issued: 2022-06-03



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Attribute types, Granta limitations, and return on investment

Jillian O'Neel, Phil Schembri,
Jack Brett, Lisa Hughey

06/01/2022

This discussion continues a previous discussion of STXT vs. Discrete vs. Tabular; Here is where we left it...

	STXT	Discrete	Tabular
"Easy" Data Consistency		✓	✓
"Easy" User Entry	✓	✓ (if short list)	
"Good" for Searching & Reporting		✓	Sort of
Maintenance (currently)	Manual review	Admin adds new values	Create and maintain table + records

The use-case affects which type should be chosen

- Two broad classes of use-case:
 1. *Reproducibility*
 - “The metadata in a record must be sufficient for me reproduce the data in the record”
 2. *Data Analytics*
 - “I want to be able to search based on the data”
 - “I want to be able to report the data”
 - “I want to look for trends in the data”
- The use-cases result in different **requirements** in how the data is stored:
 - Reproducibility – the structure or format of the data is not critical to the use-case
 - Data Analytics – The use-case is most easily satisfied when the data is structured or formatted consistently

When to use each type

(In the absence of Granta limitations)

- STXT
 - The value set is unconstrained (i.e. value can be anything)
 - And either:
 - The use-case is 'reproducibility' (no validation needed)
 - The use-case is 'data analytics' and the value and/or format can be validated
- Discrete
 - The value set is constrained but not 'too long'
 - And the value set is relatively static (changes needed infrequently)
- Linked table
 - The value set is constrained but too long for Discrete
 - And/or metadata related to the value is necessary
 - And/or the value set changes frequently

Test Plan Document Number

TK5K0276-A-004

Example of STXT

Data Sensitivity	OUO - Export Control Info (ECI) ▼
	OUO - Export Control Info (ECI)
	Official Use Only (OUO)
	Unclassified Controlled Nuclear Info (UCNI)
	NS - Non Sensitive - No CUI (Controlled Unclassified Info) Present

Example of Discrete

Project Information				
Project Name	Project Code	Project Point of Contact	Funding Organization	Project Notes
Strength of Metals and Alloys	RL30	Brad Clements (104233), Eric Mas (165959)	Joint Munitions Program	Characterize the microstructure

Example of Linked table

Granta limitations

- Tabular data cannot be populated via text importers.
- Very limited support for **multi-table upload**
 - No support in text importers
 - Custom template development necessary for Excel importers
 - No support in Viewer or Explore
- No automated **data validation**, so either:
 - STXT can not be used for 'data analytics' use-cases, or
 - STXT data must be validated manually (big problem!)
- No **context-sensitive discrete**, so lists must be quite short
 - Even if lists are relatively static, discrete types can be too long to be practical (e.g. material classification hierarchy; test type hierarchy)
- No linking-value **prompts** (e.g. type-ahead, or searching) or **validation** in:
 - Viewer
 - Excel

Consequences of Granta limitations

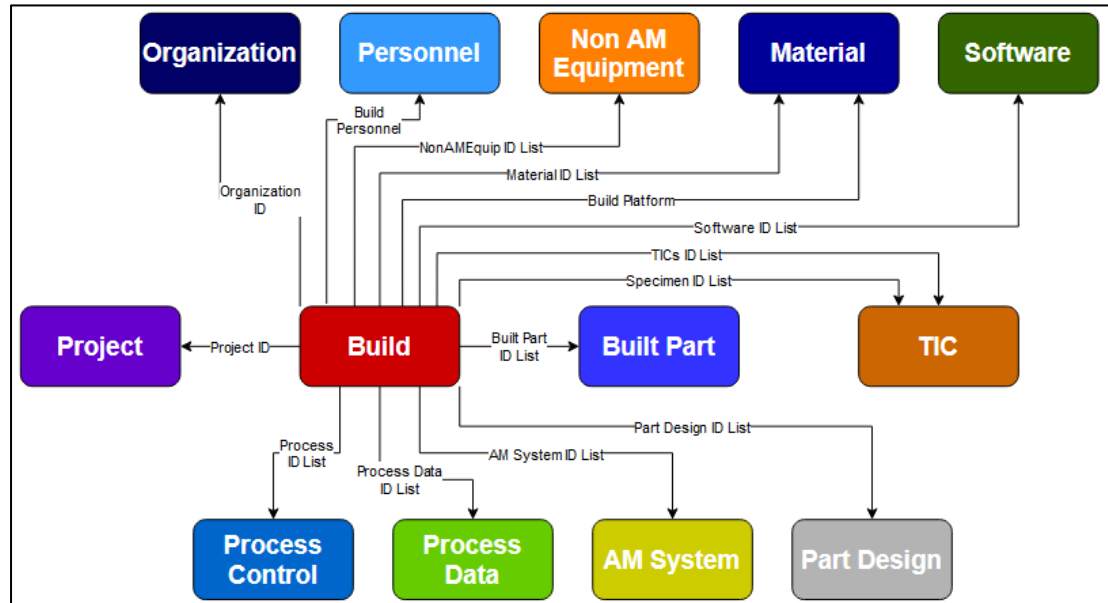
- Increased need for of training and documentation
 - Effort required to create/maintain documentation
 - Effort required by users to understand documentation
- Increased possibility of errors in data
 - Review of records by 'Granta SME' is necessary
- Poor user-experience
 - Busy users with many projects de-prioritize data management
 - Difficult to find personnel to work with Granta

These consequences result in:

1. **Slow and incomplete** database population
2. Low return on investment
3. **Decreased management support for Granta**

Example: ASTM F3490-21 (AD Data Standard)

- The 'Build Module' in ASTM F3490-21 contains many fields that link to other 'modules', including:
 - **Build Personnel**
 - **Organization ID**
 - **Software ID**
 - **AM System ID List**
 - **NonAMEquip ID List**



Links from ASTM F3490-21 Build Module to other modules (some modules not shown)

Example: ASTM F3490–21 (AD Data Standard)

- Discrete is not a good option for most of these fields:
 - Too long
 - Too dynamic
 - STXT might have been an option, except:
 - Values should be a consistent format for searching/reporting, so **data validation is needed**.
 - Granta provides no automated validation tools.
 - Manual validation (i.e. review) would be needed.
- Build Module fields include:
- Build Personnel
 - Organization ID
 - Software ID
 - AM System ID List
 - NonAMEquip ID List

Granta MI does not provide the tools necessary to populate an AM database consistent with ASTM F3490–21.

Path forward?

- Questions for Granta:
 - Do the necessary tools exist?
 - Are these use-cases addressed in other ways?
 - How do other customers populate their databases (without these tools)?
 - What are the personnel roles you assume for database population – is there intended to be a 'Granta SME' for reviewing data?
- For the Nuclear Security Enterprise, population of quality databases without addressing these issues will be almost impossible.

The Previous Discussion

Examples

- This discussion generally arises when we have data that is a “string” and we expect multiple users may want to enter in the same value(s) for multiple records. We can usually anticipate what people will want to enter, at least presently, and we expect someone might want to collect all records that have the same value entered. Some values could be used for access control.
- For example:
 - **Name of operator** (the same operator may perform many tests across many projects)
 - **Test location** (testing occurs in buildings/labs and test equipment is unlikely to change location very often, resulting in lots of test records needing the same Test Location value entered)

Our options include: Short text (STXT), Discrete (drop-down menu), or a Tabular link to a separate table for the data. Which option is chosen likely must be determined on an attribute-by-attribute basis.

Short Text (STXT)

- Pros?
 - Flexible, can accommodate any entry up to 255 characters, so write users can add new values, admin intervention is not required
- Cons?
 - Flexible, will accept any entry even if the write user is not adhering to established standards
 - Currently puts a heavy burden on the reviewers to make sure that the entries are in the correct format
 - Or puts the burden on data consumers to search on all possible variations, risks people not being able to find the records they want
- Example
 - **Name of modifier** attribute in 'At-Risk Cases' table. We guided users to enter names in the format 'First Last (DUID)', where names and DUID are from OneID phonebook

Discrete (drop-down menu)

- Pros?
 - Write users cannot change the wording/formatting of the data
 - No burden on reviewers to check formatting
 - Easier to search for records with specific values
 - Entering data is easy for write users:
 - Pick choice from the menu
 - No formatting requirements
 - No need to track down data (ex: people's DUIDs from OneID phonebook)
 - Cleaner reporting because standardized format
 - System rules, such as access control, rely on consistent values
- Cons?
 - Write users cannot change the wording/formatting of the data
 - New options must currently be added by an admin – administrative burden
 - List of options could get quite long for some attributes
 - Currently, no way to filter what options someone sees in Viewer
- Example
 - LANL used discrete for **Material Name** attribute to prevent duplicates, e.g. '304L' vs 'SS304L'
 - List got very long very quickly
 - We reverted to STXT with attribute definitions to provide guidance

Tabular link to separate table

- For the example of **Name of operator**, what if we made a database table called “People”?
- The “People” table may include attributes like **First name**, **Last name**, **DUID**, **Site**, etc.
- Each “name” in the database has a corresponding ‘People’ record

Tabular link to separate table

- Pros?
 - Simplifies review by allowing automated cross-check of linking values
 - A 'People' record only has to be completed once for each person
 - Format is semi-standardized via linking value & tabular columns
 - Possible integration with existing databases; e.g. OneID
 - Allows users to create new related records and maintain the value list
- Cons?
 - Requires the creation and maintenance of a table
 - Requires the creation, review, and release of a record for each value
 - Might not be able to be automated so potentially adds an extra step for write users
 - Attributes like **First name** and **Last name** are probably short text so people could still be entering them incorrectly when they make Name records
 - Although correcting them in one place corrects them everywhere (unless it's the linking value!)
 - Not very intuitive for casual write users in Granta's native interface, requires some understanding of database structure
 - Difficult to populate tabular attributes w/ excel and not possible with text importers
- Example
 - The 'stock' AM schema uses a similar concept for 'Machines'

Summary: STXT vs. Discrete vs. Tabular

	STXT	Discrete	Tabular
“Easy” Data Consistency		✓	✓
“Easy” User Entry	✓	✓ (if short list)	
“Good” for Searching & Reporting		✓	Sort of
Maintenance (currently)	Manual review	Admin adds new values	Create and maintain table + records