|  |  |
| --- | --- |
| Logo, icon  Description automatically generated | Enterprise Data Asset Management (EDAM) Studio  Mappings Specifications  2023-02-26 – Draft v0 (last updated: 2023-02-26) |

Table of Contents

[1.0 Mapping Work Road 1](#_Toc128302110)

[1.2 Workbooks are not Mapping Tools 1](#_Toc128302111)

[1.3 Define your ES Schema Bound Ts 2](#_Toc128302112)

[1.4 Code is Good 2](#_Toc128302113)

[1.5 Virtual Data Assets 2](#_Toc128302114)

[1.6 A T should be Describable and Verifiable 2](#_Toc128302115)

[1.7 The Failure of Graphical Only Tools 3](#_Toc128302116)

[1.8 Constructing and Deconstructing 3](#_Toc128302117)

[1.9 Regressing 3](#_Toc128302118)

[1.A Some Rumors 3](#_Toc128302119)

[1.B Conclusions 3](#_Toc128302120)

# Work in Progress

This mapping specifications will evolve in time, for now we are providing a Road Map that hopefully will be influenced by EDAM Studio contributors and others. Expect changes at any time.

# 1.0 Mapping Road Map

The EDAM Studio first implementation of its mapping capabilities included just one source one target but while revisiting the need to have a tool to help us map from one or more sources to one or more targets the need to rethink existing commercial tools impediments and limitations is needed. Nothing wrong with any commercial tool that many have been using successfully including the authors and contributors of EDAM Studio (ES).

## 1.2 Workbooks are not Mapping Tools

Business Analysts (BA’s), Data Architects (DA’s), Data Engineers (DE’s) and others (collectively the “Data Assets Technical Community” [DATC]) has struggled using a word document and mostly Excel of some other spreadsheet workbook tool to help them have the liberty they need to do schema to schema, nothing to schema or schema to nothing mappings. A specific DATC member will be referred as a DATM from herein. At the end Workbooks end up being the (short lived) source of mapping truth for a given design targeting their use cases. The problem with those is that there is no way to validate any workbook mapping since sources and targets may be illusive and intangible (meaning that those can’t be represented with an instance of any well know data instance or artifact) and since Excel and other products don’t support schema bound representation of represented items its impossible to validate those within Workbooks. Everyone has suffered for years the issues of having multiple versions of workbooks each of those only valid for a very short period of time that at the end becomes useless, difficult and mostly impossible to maintain consistently. Look for those in a drawer, a file, some hardcopy that withing a short period will be lost or forgotten and if a conscious effort has been made to keep them together somewhere it will be extremely difficult to locate anything with that repository.

## 1.3 Define your ES Schema Bound Ts

In ES every collection of data elements from any source managing a specific use case requirements set is considered a Data Asset that should always be uniquely identified by using a “Domain URI”. DATC should always be able to work with use cases mappings at their own pace while ES will be keeping the source and target schemas under the hood even on those cases that ambiguous representations are being managed including nothing to schema and schema to nothing, and optional sources or targets meaning that a source (or target) may present a mapping of some data elements that are not in the source schemas to others in the target that again may not even have a definition (from now being referred as “a (Data Asset) Thing” (T). T’s will always have a (default) schema construct that in the most extreme case of ambiguity be just a string with a Name and all other known fields of a Data Element. Consider any ES data element well or ambiguously represented a T.

## 1.4 Code is Good

Use Case Data Asset / Element Mappings should not be bounded, limited, or restricted to be represented, calculated, transformed, or consumed by any single technology, programming/scripting language (“language” for short) or development tools such as XSLT or other including those inherently in ES implementations. ES should support plugin or connectors that will allow the use of language or some external technology that will enhance ES by making the translation between something to ES T’s or ES T to something in their own external world. This means that within an ES Book – Booklet Map Item (see 9999) the code segments could be in the language it needs to be to be properly managed to produce a source or target T. The initial ES supported language is JSONata but hopefully we should see the support of others (such as Java Script, Python, and others).

## 1.5 Virtual Data Assets

While working with an elusive source or target meaning that there may be no existing schema representation for it, ES should allow its creation at any point in time in the mapping process and help DACM’s to work with those as needed in their design / mapping process alongside any existing Data Asset source or target schemas. Hopefully this will be helpful in brainstorming while working with the mappings of Conceptual to Logical to Physical backwards and forward data model mappings. At the end ES will provide a way to make those on-the-fly artifacts (schemas) to be recorded and useful while describing and validating a mapping artifact.

## 1.6 A T should be Describable and Verifiable

In ES “Data Driven” constructs should reign in where Ts are described rather than programmed, but if a simpler solution easy to maintain and manage by some DATC back-box code then just allow it. ES inputs and outputs of any T should be supported by its ES description not some code. Only on extreme cases and where it makes sense some code could be used but such code should be encapsulated, and its container needs to implement some trickery to allow gathering information of a T under ES Data Asset or Element (Data Driven) T terms. From ES perspective is just well-known verifiable T. In conclusion in ES all Ts are described and can be verified.

## 1.7 The Failure of Graphical Only Tools

ES initial implementation don’t provide any graphical representation of the Mapped Items but should support those in later versions. Using commercial tools graphical mapping capabilities failed miserably as DATM move from a very few mappings to complete mappings of everything resulting on cluttered visuals that designers and developers hate to work with and maintain them if that is possible. The end products, view and printed artifacts are unreadable by humans and therefore difficult to use in its graphical representation of the mapped everything. Commercial tools lure you into using their tool with just limited number of mapped items but at the end your projects will not be so lucky to have just a few items.

ES limit the representation of T’s with small subsets within a Map Item that will and should always be readable and could be represented with a some graphics and as soon it become difficult to read just decompose it into one or more Map Items, always keep those small and simple.

## 1.8 Constructing and Deconstructing

In ES any source and target artifact should be able to be decomposed, composed, or recomposed allowing working on the mappings at any stage. A T has been described as a “Thing” that will be used in a Map Item mapping ending in the production of a required source or target item. ES should provide a clear way to stich them up ending up with the creation of one or more source or target items using a collection of assumptions on how to compose or recompose T’s to produce the needed end product schema or instance descriptively or with code within a Map Item Booklet.

## 1.9 Regressing

ES will not stop the World production short-lived Workbooks. Use ES until a Workbook produced by the tool is needed to complete a job and hopefully at the very end or far enough on a job to allow most or (also hopefully) all the artifacts in your mapping to be fully verifiable.

## 1.A Some Rumors

The idea to work on the specification of connectors to external EDAM sources or targets like Azure Purview, AWS Glue catalogs, Collibra or others has been mumbled. Perhaps we could see some Proof of Concept (POC) of one of those in the future as time and EDAM Studio contributors’ real jobs permits.

## 1.B Conclusions

We may end-up creating Workbook mappings in a few years but in the meantime and while hope prevails contributors to the EDAM Studio project will work hard to make the above ideas into a workable product and if you are one of those DATM’s that have some time then contribute to this project in any capacity you feel comfortable with; even balcony conscious/constructive critics are welcomed.