Allotrope Foundation Comments on FDA-2017-N-2166 "Draft Standardization of Pharmaceutical Quality/Chemistry Manufacturing and Control Data Elements and Terminologies; Request for Comments"

(Docket Number FDA-2017-N-2166)

September 8, 2017

Contents

Introduction 3

Comment Overview 4

General Comments 6

PQ/CMC Data Elements 6

Summary of Allotrope Comments on Unique Identifiers 6

Summary of Allotrope Comments on Active Structure Files 6

Summary of Allotrope Comments on Specifications, Tests, Methods Metadata 6

Summary of Allotrope Comments on Impurities-Related Metadata 7

Potential Collaborations with Allotrope Foundation 7

Closing 9

Appendix: 10

1. Specification 10

2. Test 11

3. Acceptance Criteria 12

4. Batch or Lot Information 12

5. Batch Analysis Drug Substance or Drug Product 14

6. Stability Study 14

7. Nomenclature & Structure of Drug Substance 15

8. Drug Substance Characterization 15

9. Description and Composition of Drug Product 16

10. Batch Formula 17

11. Drug Substance - Control of Materials 17

13. Drug Substance Impurities 20

14. Drug Product Impurities 20

15. Analytical Methods Validation 21

Introduction

About Allotrope Foundation

Allotrope Foundation welcomes the opportunity to provide comments to FDA on the guidance FDA-2017-N-2166 entitled "Draft Standardization of Pharmaceutical Quality/Chemistry Manufacturing and Control Data Elements and Terminologies," which is termed PQ/CMC throughout this document. Founded in 2012, Allotrope Foundation (www.allotrope.org) is an international consortium of pharmaceutical, biopharmaceutical, and other scientific research-intensive industries that is dedicated to developing an advanced data architecture to help transform the acquisition, exchange, and management of laboratory data throughout its lifecycle. Allotrope aims to make the intelligent analytical laboratory a reality – an automated laboratory where data, methods, and hardware components are seamlessly shared among disparate platforms, where one-click reports and analytics can be produced based on data generated by any analytical instrument and data integrity is built-in by design from the point of data capture. Allotrope's vision of an intelligent analytical laboratory will be realized through the creation of an "ecosystem" in collaboration and consultation with instrument and software vendors and the scientific community. Our first initiative is the development of the Allotrope Framework for analytical data, consisting of a standard data format, class libraries for interfacing with applications, and semantic capabilities to support standardized, structured metadata. Our shared mission is to develop new approaches to improve data access, interoperability, and data integrity through standardization, which ultimately serves as a key enabler of data-driven innovation.

Our Approach

The development of taxonomies, ontologies and data models is enabled by working groups that are comprised of representatives from both Member and Allotrope Partner Network companies. The Allotrope Partner Network provides a forum for members of the vendor and non-profit community to interact and provide feedback to the Allotrope Foundation. The modular working groups, with representatives from Member companies and Partner Network companies, collaborate to define the terminology and data models for a particular technique and associated workflows, and feed a central governance and integration process, with a curation team and principal semantic engineer to ensure adherence to architectural and style principles consistent with Basic Formal Ontology (BFO).

Foundation Members

AbbVie, Amgen, Bayer, Baxter, Biogen, Boehringer Ingelheim, Bristol-Myers Squibb, Genentech, GlaxoSmithKline, Eli Lilly & Co., Merck & Co., Inc, Novo Nordisk, and Pfizer

---

Comment Overview

As a consortium of international pharmaceutical, biopharmaceutical, and other scientific research-intensive industries dedicated to developing capabilities to support standardized, structured metadata, Allotrope realizes the power of thoughtfully developed and implemented standards to maximize the impact of our data. Upon review of the subject PQ/CMC document, Allotrope noted that FDA did not clearly state its strategic vision for the PQ/CMC initiative, where it sees alignment opportunities with related standardization initiatives, and importantly how it plans to ensure the effort is globally harmonized to maximize value. Therefore, Allotrope urges FDA to more explicitly share its thinking in these areas to create productive strategic discussions on future goals and potential areas of collaboration with industry.

Although not explicitly stated, Allotrope infers from our review of the PQ/CMC notice that the document is an initial attempt by FDA to create a comprehensive, well-defined, and controlled vocabulary for the eCTD to support future informatics applications, and in this light Allotrope views this FDA notice as an excellent start towards that goal. Grouping of these data elements into 15 logical domains to modularize this information is also a positive step to organize the vocabularies. Similarly, efforts to create controlled vocabulary lists where possible (i.e. drop-down lists) for the data elements within each domain is also a positive design attribute. Finally, using consensus to drive acceptance of specific definitions for each data element and item in the controlled vocabulary lists is also a positive design attribute.

Allotrope views the eCTD document as one of the most critical aggregation of CMC data elements employed in the regulatory workflows for a medicinal product. Therefore, Allotrope believes that standardizing these data elements (metadata) in the eCTD provides several powerful opportunities which bring value to both industry and regulators alike through increased data integrity, data automation, data mining, and other analytics capabilities described below:

- auto-population of meta data into the eCTD from source data systems used in the pharmaceutical industry to create CMC content (e.g. laboratory information management systems (LIMS), electronic notebook (ELN), chromatography data systems (CDS), manufacturing execution systems (MES), mastered data systems, etc.),
- auto-population of meta data from the eCTD into downstream medicinal product databases such as G-SRS for the GINAS project or the IDMP database to support the lifecycle of medicinal products, including pharmacovigilance and risk management efforts, and
- highly effective data mining across this information network of source data systems → eCTD → downstream medicinal product databases, leveraging the controlled CMC metadata established throughout the information network.

Allotrope strongly believes in this investment in the future informatics potential. However, Allotrope acknowledges in the short term that modification of current data elements will have impact on existing and new eCTDs, and the longer term value of the change may not be realized until the degree of data automation envisioned has become standard practice. To avoid potential unnecessary upregulation, Allotrope encourages such standardization efforts to be very discriminating in selecting data elements that have clear informatics potential. In the sections that follow, Allotrope provides recommendations and suggestions on specific elements to help realize the long-term potential while reducing the short term impact.

Finally, Allotrope would recommend creation of a formal, machine-readable ontology using these standardized CMC data elements as critical inputs, and would be interested in partnering towards such

an effort as described later in this response. Allotrope believes we can help FDA and other contributors develop a comprehensive and robust ontology that could help enable incorporation of the PQ/CMC standardized data elements into the eCTD via enhanced data automation.

General Comments

PQ/CMC Data Elements

In the sections that follow, Allotrope has provided some directional feedback on trends we have noted in the PQ/CMC data element recommendations. Additionally, specific comments on the scope, controlled vocabulary lists, and definitions proposed for individual data elements have also been provided in the attached Appendix. Allotrope notes that our feedback is provided through three lenses (viewpoints) for FDA consideration to provide what we feel is a holistic assessment of the PQ/CMC effort in current workflows as well as potential future applications which potentially would benefit from use of the Allotrope Framework:

- Lens 1: is the CMC metadata content proposed appropriate for an eCTD today,
- Lens 2: forward-looking views of the informatics potential of standardized eCTD metadata, and
- Lens 3: opportunities to drive global CMC harmonization through Allotrope technology

Summary of Allotrope Comments on Unique Identifiers

PQ/CMC emphasizes the use of unique identifiers for manufacturing and testing sites (e.g. DUNS, FEI, CFN) and materials (e.g. UNII, INN, etc) in the eCTD. Establishment of a single code for each site, which is globally-recognized and available on a centralized web service, would increase the value of standardizing data elements and support creation of more global dossiers. Additionally, standardizing these unique identifiers will facilitate automated upload of the metadata to the eCTD, improving data integrity. Since a site or material will have different roles across eCTDs (e.g. manufacturing site vs analytical site or starting material vs drug substance, etc.), development of ontologies which recognize these differing roles would be of value to future data mining and analytics.

Summary of Allotrope Comments on Active Structure Files

PQ/CMC emphasizes the use of active structure files (e.g. SMILES, SDF, MOLFILE, InCHi File PDB, mmCIF, etc) in lieu of inactive graphics images or static pictures of structures as currently commonly used in the eCTD. Allotrope supports the intent of this innovation, which will facilitate structure-based searches across medicinal product eCTDs, as well as supporting downstream structure-based searches in databases such as those used for the IDMP and GINAS projects. Unquestionably, the ability to search using active structures is a powerful and unambiguous means to mine data for materials and leverage structure similarity algorithms. Alignment on structure formats will allow software vendors to develop the appropriate interfaces required for regulatory workflows.

Summary of Allotrope Comments on Specifications, Tests, Methods Metadata

PQ/CMC recommends a significant amount of metadata for Specifications, Tests, Methods, and Stability studies to capture elements to promote traceability such as file names, types, version, date, and status. Additionally, PQ/CMC recommends additional contextual metadata such as intended uses, procedure numbers, protocols, origins, categories, and study purposes. Allotrope supports the effort to improve the traceability of these data elements from internal company data systems to the eCTD as this will facilitate data mining, analytics, and enhanced data automation (e.g. automated upload of specification, test, or methods information into an eCTD from source data systems). However, Allotrope questions the necessity to include some of the individual metadata in these domains, as highlighted in the Appendix, unless the intent is to use eCTD to support quality audit workflows of data elements such as status, versions, dates, etc. Of note, the Allotrope Data Format is a complete record of all system metadata, so the technology will support automated transfer of all recommended metadata with 100% data integrity if the final PQ/CMC recommends including these data elements. Allotrope believes the creation of sound formal ontologies, with proper definitions and semantic relationships, as recommended in this response,

will be of great value to establish the detailed relationships between these CMC data elements data mining, and data analytics.

Summary of Allotrope Comments on Impurities-Related Metadata

PQ/CMC metadata recommendations for both drug substance and drug product impurities are a good first step to establish an initial tier of vocabularies. However, these vocabularies have significant gaps for practical implementation of a complete controlled vocabulary. Allotrope would suggest additional data elements to consider including terminology used for elemental impurities, stereoisomers, degradants, fate and purge, mutagenic impurities, and purge prediction, as well as linkages to concepts in relevant guidance such as ICH Q3A-Q3D and ICH M7). Furthermore, impurities-related information in the eCTD would be greatly enabled by creation of a formal ontology. Using a formal ontology, one can link the various roles a specific chemical compound(s) may play within or across eCTDs (e.g., raw material, starting material, drug substance, etc.) to critical process control elements, such as specifications and analytical methods.

Additionally, Allotrope acknowledges the usefulness of an analytical data file reference in the eCTD. While such a reference creates a link to the internal data systems used to generate the data, Allotrope recommends considering inclusion of selected Allotrope Data Format (ADF) files as supporting data in future eCTD submissions as a more powerful innovation than a static reference. The ADF file provides active access to a vendor-neutral version of the instrument data, including a complete record of the associated metadata and audit trail. Inclusion of an ADF file may also positively impact the ability to remotely audit eCTDs due to the implicit data integrity and completeness of the file format, which may offer value by reducing the scope of on-site CMC inspections in favor of additional remote auditing elements.

Potential Collaborations with Allotrope Foundation

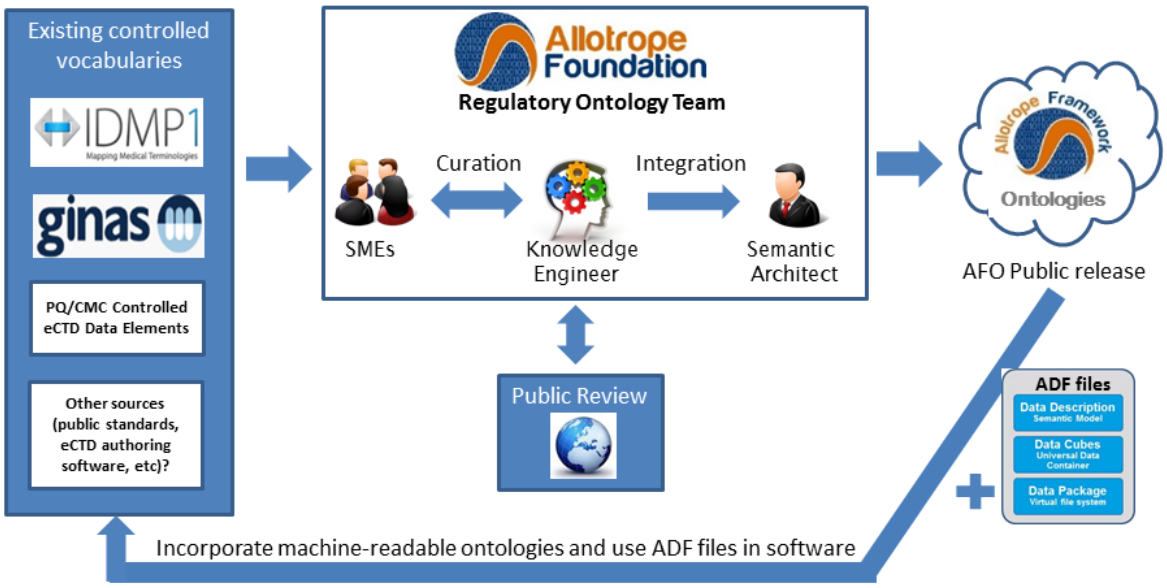
Allotrope believes our goals align well with the PQ/CMC controlled data element effort. Allotrope seeks to drive enhancements in data automation, data integrity, data mining, and data analytics through implementation of standardized data format and formal ontologies for use in workflows from the laboratory to the dossier. Through these efforts, Allotrope has developed semantics expertise and efficient workflows to develop formal ontologies and semantically-correct definitions, which are compliant with Basic Formal Ontology design principles.

Allotrope sees a significant opportunity to partner with industry and regulators to link these initial PQ/CMC vocabularies used in the eCTD with other relevant vocabulary lists from other sources, by developing them into a single semantically-correct and machine-readable formal regulatory ontology. The Allotrope regulatory ontology would be made publically available through a web interface at no cost to users along with other ontologies Allotrope is developing across pharmaceutically-relevant workstreams, actively curated via a governance process, and readily extensible as new concepts emerge.

When completed, the regulatory ontology can be incorporated into the various software systems used to author, review, search, archive, and extract regulatory information, including the downstream databases used in the IDMP and GINAS projects. The regulatory ontology will ensure that the hierarchical relationships between data elements (physical entities) are available to other software systems which employ Allotrope technology. Additionally, the regulatory ontology will also ensure the processes which connect these data elements will be accessible to software systems, creating much deeper semantic meaning and therefore much more powerful data mining and analytics capabilities.

A visual representation of this potential collaboration is provided below (Figure 1), which progresses complete controlled vocabulary lists in documents (left side) to formal machine-readable ontologies (right ride), and then reapplies these ontologies to the source computer systems for use. As noted previously, selection of Allotrope Framework technology for this effort also has the added significant benefit of association of the laboratory data in vendor-neutral format (ADF files) with the ontology, expanding potential data mining and data reuse applications.

Figure 1: Potential Project to Develop BFO-Aligned Regulatory Ontologies



Note: proposed Allotrope Regulatory Ontology Team would be composed of subject matter experts from industry, regulators, and standards bodies as required

Closing

Allotrope Foundation is grateful for the opportunity to provide comments on PQ/CMC. The Foundation hopes this feedback will assist the Agency in their efforts. Allotrope would also welcome the opportunity to further discuss how Allotrope Foundation could help FDA and other contributors develop a comprehensive and robust ontology that could support incorporation of the PQ/CMC standardized data elements into the eCTD via enhanced data automation.

If you require clarification of any of these comments, please contact the Allotrope Foundation Secretariat:

Allotrope Foundation Secretariat*
1500 K Street NW
Washington DC 20005
Telephone: (202) 230-5158
Fax: (202) 842-8465
Web: www.allotrope.org
Email: secretariat@allotrope.org
**Secretariat, Drinker Biddle & Reath LLP*

Appendix: Allotrope Foundation detailed comments on Pharmaceutical Quality + Chemistry Manufacturing and Controls Data Elements and Terminologies

Data Elements		Summary Comment
1. Specification		
1	Specification Title	Agree this element should be included. However, while this example is acceptable for a commercial product, it would be useful to add an example of a DP specification Title for products in the investigational phase, as different presentations/formulations/processes for the same product may be used. (e.g. 75 mg tablet may have different composition, shape, size and process during clinical development). To support future data mining, data automation, or quality workflows, a standardized naming format would be needed for this element for it to add practical value beyond tracking from internal to external systems.
2	Specification Type	Agree with inclusion of this element which is included in S.4.1 text and tables implicitly, but for data mining and data automation should come from standard vocabulary list else the informatics value is reduced
3	Specification Version	The added value of this element is unclear in current eCTD workflows and could be viewed as an optional element as not all companies track this level of detail in regulatory dossiers. To add informatics value to an eCTD, a standardized format would be needed to enable future data automation and data mining activities.
4	Specification Version Date	The added value of this element is unclear. To add informatics value to an eCTD, a standardized format would be needed to enable future data mining activities. In present state, this should be an optional element as not all companies track this level of detail in regulatory dossiers.
5	Specification Status	The added value of this element is unclear. The data element is potentially problematic as the FDA does not approve the specification separately from the dossier. Companies would not include an unapproved specification in the eCTD so it is unclear the value in requiring this data element in the eCTD even if automated and consistent
6	Specification Status Date	The added value of this element is unclear. The data element is potentially problematic as the FDA does not approve the specification separately from the dossier. Companies would not include an unapproved specification in the eCTD.

7	Additional Information (comments)	Tests, acceptance criteria, and methods are all part of the specification and should be included as data elements in this domain for completeness of future ontology
2. Test		
1	Test Name	To enable future data mining activities and provide greater value for this element, a standardized format would be needed.
2	Test Usage	Consider only creating release and stability terms and select both when both used rather than creating combinations as options.
3	Test Method Origin	Propose replacing term “Proprietary” with “In-house” or “Custom.”
4	Test Category	Grouping of tests by type of function may have some value in a future ontology if the function is selected from a controlled vocabulary. This grouping would allow queries such as "Show me all test results from the test category 'Assay.' The data element may be more appropriately named "Test Function."
5	Analytical Procedure	The examples provided in definition are Analytical Techniques (e.g. HPLC) and not Analytical Procedures (e.g. full method details), so suggest considering adding both terms (Analytical Procedure and Analytical Technique) with correct ontological definitions.
6	Reference to Procedure	A better title for this element may be Sponsor Method Reference. At present suggest making this element optional as not all companies embed this level of detail in regulatory dossiers, especially for USP methods, where only reference to USP may be provided. Also consider that companies may have multiple QC labs, each with its own method reference. We should consider how minor updates to these methods (e.g. typo corrections) would impact the submission and data element. However, this level of detail will be helpful for future data mining or auto-population of eCTD to ensure details from correct method version have been uploaded if Allotrope format is adopted for regulatory applications.
7	Relative Retention Time	Relative Retention time is a data element for the Method domain and not Test domain.
8	Stage Name	It is unclear if stage name is needed for tests. Suggest considering this element part of Method domain and not part of test, if included at all.
9	Stage Sequence Order	It is unclear if stage sequence order is needed for tests. Suggest considering this element part of method domain and not part of test, if included at all.
	Additional Information (comments)	There is an opportunity to clarify some definitions in a regulatory ontology. A test is a component of the

		specification which requires acceptance criteria. A test requires a method, which contains experimental and instrument execution instructions. Tests often leverage analytical techniques, but are not analytical techniques themselves.
3. Acceptance Criteria		
1	Value	As defined this element sounds like results, but this value should read "numerical portion of acceptance criteria."
2	Value Unit	No comment.
3	Literal Text	The term Literal Text is ambiguous; suggest considering alternative term.
4	Interpretation Code (text)	No comment.
	Interpretation Code (numeric)	<, >, ≤, ≥ are often used in practice but are not addressed in the controlled vocabulary list. Allotrope suggests consideration of informatics implications of text vs mathematical representations of specifications.
5	Additional Information (comments):	Acceptance Criteria also include data elements for component names which are often part of acceptance criteria, and component aggregations such as Specified and Unspecified Impurities.
4. Batch or Lot Information		
1	Batch or Lot Number (Bulk Batch ID)	This term needs to be better defined. Batch commonly refers to a manufacture of a material, while lot commonly refers to the packaged product from the batch run. A batch run in large processing equipment may result in multiple lots if too big for a single dryer; a company would typically include packaged lot number to provide genealogy, but would not include internal batch numbers.
2	Batch or Lot Number (Packaged Batch ID)	Packaged material should be referenced by lot number.
3	Manufacturing Site Name	Suggest standardizing site unique identifiers to facilitate future data mining and automation capabilities.
4	Manufacturing Site Unique Identifier	Suggest standardizing site unique identifiers to facilitate future data mining and automation capabilities.
5	Manufacturing Site Unique Identifier Type	Suggest standardizing site unique identifiers to facilitate future data mining and automation capabilities.
6	Manufacturing Date	Rationale for selecting date should be defined more explicitly for both DS and DP in definitions to drive standardization.
7	Manufacturing Date Description	Recommend standardizing definitions so that the description/rationale field is not required. If there are multiple valid choices in definitions that cannot be standardized, suggest considering if rationale be reduced to

		a controlled vocabulary list, which would facilitate future data mining and data automation capabilities.
8	Testing Site Name	Test site(s) included in S.2.1 could be populated from automated lookup of controlled vocabulary by test site FEI, DUNS, etc in future. However, there seems limited value to include test sites in S.4.4 unless multiple test sites have been used for testing and added clarity is required.
9	Testing Site Unique Identifier	Support inclusion of this element as it would facilitate future automated data look-ups and data mining capabilities.
10	Testing Site Unique Identifier Type	Support inclusion of this element as it would facilitate future automated data look-ups and data mining capabilities, but should address whether testing sites outside the U.S. for clinical phase materials have DUNS, FEI numbers.
11	Batch Size	The definition of this term should be further clarified.
12	Batch Size Unit	No comment.
13	Expiration Date	For this term to be useful, a standardized format and definition needs to be provided; also, consider different types of dating, such as MONTH/YEAR, to avoid forcing all companies to align to the same date format unnecessarily.
14	Retest Date	For this term to be useful, a standardized format and definition needs to be provided; also, consider different types of dating, such as MONTH/YEAR, to avoid forcing all companies to align to the same date format unnecessarily.
15	Container Closure System Description	Recommend this element be captured using a list of controlled vocabularies, rather than leaving free text entry, in order to better support future data automation and data mining capabilities.
16	Container Type	What is currently included may not adequately cover all materials of construction for containers (fiberboard, glass, HDPE, stainless steel, etc).
17	Closure Type	No comment.
18	Container Size	No comment.
19	Container Size Unit	No comment.
20	Container Fill	No comment.
21	Container Fill Unit	No comment.
22	Batch Utilization	The definition of "Development" is not clear - does it include non-clinical, safety and toxicology uses? Need to be able to select multiple attributes as a lot may have multiple purposes.
23	Drug Substance Lot Number	It is unclear if this is equivalent to lot number above.
24	Additional Information	It would be beneficial if data elements/terminology could also be established for PQ/CMC elements related to Manufacture and Container Closure.

5. Batch Analysis Drug Substance or Drug Product		
1	Batch or Lot Number	As above, this term need to be better defined. Batch commonly refers to the manufacture of a material, while lot commonly refers to the packaged product from the batch run. A batch run in large processing equipment may result in multiple lots if too big for a single dryer; a company would typically include packaged lot number to provide genealogy, but would not include internal batch numbers.
2	Specification Version	The added value of this element is unclear. To add value to an eCTD, a standardized format would be needed to enable future data mining activities. This should be an optional element as not all companies track this level of detail in regulatory dossiers.
3	Test Date	No comment.
4	Test Category	Grouping of tests by type of function may have some value in a future ontology if the function is selected from a controlled vocabulary. This grouping would allow queries such as "Show me all test results from the test category 'Assay.' The data element may be more appropriately named "Test Function."
5	Results	No comment.
6	Conformance to Criteria	No comment.
7	Testing Site Name	Mention test sites only in S.2.1
8	Drug Substance Product Indicator	No comment.
9	Drug Substance Lot Number	No comment.
10	Release Date	Only manufacturing date is provided in current eCTD.
11	Additional Information (comments):	Specification, acceptance criteria, tests are also part of the batch analysis tables; standard formatting is critical to enable data automation of batch data tables.
6. Stability Study		
1	Study Name	To add value to a future eCTD and support future data mining, data automation or quality workflow capabilities, a standardized naming format would be needed.
2	Study Design	It is unclear if this element is the same as a stability protocol. A standardized tabular format instead of text would be beneficial and associated terminology to describe content (e.g time point, test, acceptance criteria, method, result, etc.) would also need to be defined. It may be possible to derive the study design from the metadata in Specification, Test, Acceptance Criteria.
3	Storage Conditions	Please consider photostability conditions.

4	Protocol Indicator	The term "stability protocol" first needs to be adequately defined to include this term (protocol identifier). A standardized format would be needed for the element to add value in a future eCTD and to support future data mining, data automation or quality workflow capabilities.
5	Study Identifier	To add value to a future eCTD and support future data mining, data automation or quality workflow capabilities, a standardized naming format would be needed.
6	Study Type	No comment.
7	Container Orientation	No comment.
8	Study Purpose	This element would benefit from a standard vocabulary. It is unclear if this element would include terms like long term stability, accelerated stability, site specific stability. The distinction between study type and study purpose should be clear in the definitions of these elements.
9	Additional Information (comments):	Terms such as stability protocol, tests, acceptance criteria, compound names, results, etc. all missing from stability taxonomy.
7. Nomenclature & Structure of Drug Substance		
1	Chemical Name	No comment.
2	CAS Number	No comment.
3	INN	Not consistently added to S.1.
4	USAN	No comment.
5	IUPAC Name	No comment.
6	UNII	Not consistently added to S.1.
7	Company Code	No comment.
8	Substance Structure Graphic	No comment.
9	Chemical Structure Date File	This is not typically included in S.1, but there may be benefits to having an active structure file in dossiers.
10	Chemical Structure Date File Type	No comment.
11	Chemical Structure Date File Origin	Data Element missing from Nomenclature & Structure Table
12	Additional Information (comments):	Inclusion of established codes for compounds from controlled vocabularies and active structure file formats in dossiers supports future data automation and data mining capabilities and is strongly encouraged.
	8. Drug Substance Characterization	
1	Chemical Name	Not consistently added to S.3.1.
2	USAN	Not consistently added to S.3.1.
3	UNII	Not consistently added to S.3.1.

4	Drug substance Method Type	This term needs to be better defined. The definition provided cites the technique used to elucidate structure. However, a technique (e.g. HPLC) is not a method (experimental conditions).
5	Analysis Graphic	No comment.
6	Analytical Instrument Data File	It is unclear what the expectation is for this element. Is it to reference raw data file name? This is typically not done; currently, the sample is identified but raw data file(s) are not included. However, use of a standardized data format like the Allotrope Data Format (ADF) as discussed earlier in our General Comments could help to enable this capability.
7	Analytical Instrument Data File Type	It is unclear what the expectation is for this element. Is it to reference raw data file name? This is typically not done; currently, the sample is identified but raw data file(s) are not included. However, use of a standardized data format like the Allotrope Data Format (ADF) as discussed earlier in our General Comments could help to enable the capability to include raw datasets in the dossier.
	Additional Information (comments):	This block should include analysis technique, analysis results (peak table); in the future ADF could enable presentation of this data.
	9. Description and Composition of Drug Product	
1	Product Proprietary Name	No comment.
2	Product Non-proprietary Name	No comment.
3	Dosage Form	No comment.
4	Strength	No comment.
5	Strength Unit of Measure	Formatted examples in the definition would be helpful to avoid confusion.
6	Overage Percent	No comment.
7	Overage Justification	No comment.
8	Drug Product Description	No comment.
9	Product Component Name	No comment.
10	UNII	No comment.
11	CAS Number	This value can be linked from S.1.
12	Drug Product Composition Function	No comment.
13	Amount per unit	No comment.
14	Content (%)	No comment.
15	Quality Benchmark	Suggest renaming this element Quality Standard for consistency with ICH M4Q.
16	Drug Product Component Additional Information	No comment.
17	Diluent Description	No comment.

18	Diluent Volume	No comment.
19	Diluent Unit of Measure	No comment.
20	Diluent Container Closure Type	No comment.
21	Diluent Component Name	No comment.
22	UNII	No comment.
23	CAS Number	No comment.
24	Diluent Component Function	No comment.
25	Amount Per Unit	No comment.
26	Content (%)	No comment.
27	Quality Benchmark	No comment.
28	Diluent Component Additional Information	No comment.
29	Diluent Component Supplier Name	No comment.
30	Diluent Component Supplier Address	No comment.
31	Diluent Component Manufacturer Name	No comment.
32	Diluent Component Manufacturer Address	No comment.
	10. Batch Formula	
1	Amount	No comment.
2	Amount UOM	No comment.
3	Batch Formula Additional Information	No comment.
4	Product Component Name	No comment.
5	Component Amount Per Batch	No comment.
6	Quality Benchmark	No comment.
7	Component Additional Information	No comment.
	11. Drug Substance - Control of Materials	
1	Specification Version	The added value of this element is unclear. To add informatics value to an eCTD, a standardized format would be needed to enable future data mining activities. In the current state, this could be an optional element as not all companies track this level of detail in regulatory dossiers.
2	Specification Version Date	The added value of this element is unclear. To add informatics value to an eCTD, a standardized format would be needed to enable future data mining activities. In the current state, this could be an optional element as not all companies track this level of detail in regulatory dossiers.

3	Specification Status	The added value of this element is unclear. The data element is potentially problematic as the FDA does not approve the specification separately from the dossier. Companies would not include an unapproved specification in the eCTD so it is unclear the value in requiring this data element in the eCTD even if automated and consistent
4	Specification Status Date	The added value of this element is unclear. The data element is potentially problematic as the FDA does not approve the specification separately from the dossier. Companies would not include an unapproved specification in the eCTD.
5	Substance Component Name	The name is misleading if the scope of this element is intended just for raw materials; suggest using term like "raw material" and add another term like "starting material."
6	Quality Benchmark	It is unclear what a “company standard” is. Is this just an in-house method?
7	UNII	No comment.
8	CAS Number	No comment.
9	Source Type	Proposed controlled vocabulary list is acceptable. Source Organism, Source Organism Subsource, Source Organism Subsource should only be applicable if material of biological origin is selected.
10	Diluent Component Supplier Name	It is unclear what diluent refers to and this may be in the wrong location. The assumption is this element is only applicable to large molecule APIs, vaccines and blood products.
11	Diluent Component Supplier Address	It is unclear what diluent refers to and this may be in the wrong location. The assumption is this element is only applicable to large molecule APIs, vaccines and blood products.
12	Diluent Component Manufacturer Name	It is unclear what diluent refers to and this may be in the wrong location. The assumption is this element is only applicable to large molecule APIs, vaccines and blood products.
13	Diluent Component Manufacturer Address	It is unclear what diluent refers to and this may be in the wrong location. The assumption is this element is only applicable to large molecule APIs, vaccines and blood products.
14	Source Organism	See comment on Source Type.
15	Source Organism Subsource	See comment on Source Type.
16	Source Organism Country of Origin	See comment on Source Type.
	Additional Information (comments):	Test, acceptance criteria, method, component names, etc. should also be included as part of specification; consider also adding terminology for starting materials and synthesis of starting materials (description and pictorial representation); diluent terminology is confusing and is only

		applicable to large molecule APIs, vaccines and blood products.
12. Drug Product - Control of Excipient		
1	Specification Version	The added value of this element is unclear. To add value to an eCTD, a standardized format would be needed to enable future data mining activities. This should be an optional element as not all companies track this level of detail in regulatory dossiers.
2	Specification Version Date	The added value of this element is unclear. To add value to an eCTD, a standardized format would be needed to enable future data mining activities. This should be an optional element as not all companies track this level of detail in regulatory dossiers.
3	Specification Status	The added value of this element is unclear. The data elements is potentially problematic as the FDA does not approve the specification separately from the dossier. Companies would not include an unapproved specification in the eCTD so it is unclear the value in requiring this data element in the eCTD even if automated and consistent
4	Specification Status Date	The added value of this element is unclear. The data element is potentially problematic as the FDA does not approve the specification separately from the dossier. Companies would not include an unapproved specification in the eCTD.
5	Drug Product Component Name	No Comment.
6	Quality Benchmark	No Comment.
7	UNII	No Comment.
8	CAS Number	No Comment.
9	Source Type	No Comment.
10	Drug Excipient Component Supplier Name	No Comment.
11	Drug Excipient Component Supplier Address	No Comment.
12	Drug Excipient Component Supplier Manufacturer	No Comment.
13	Drug Excipient Component Manufacturer Address	No Comment.
14	Source Organism	Suggest making this element optional as not all companies track this level of detail in regulatory dossiers.
15	Source Organism Subsource	No Comment.
16	Source Organism Country of Origin	No Comment.
17	Test Category	No Comment.
18	Analytical Procedure	No Comment.

19	Reference to Procedure	No Comment.
	13. Drug Substance Impurities	
1	Drug Substance Impurity Name	A compound may be both an impurity and a degradant so it would be useful to be able to assign either one or both roles within the terminology.
2	UNII	No comment.
3	Impurity Classification	It is suggested that this element be expanded to include metals, mutagenic impurities, PGI's, organics, etc. Also, would organic impurity classification be done according to the hazards classification scheme given in ICHM7?
4	Chemical Structure Data File	/nclude .cdx format.
5	Impurity Structure Graphic	No comment.
6	Drug Substance Impurity Method Type	Methods are generally described in S.2.3, S.2.4, or S.4.2 not S.3.2. This element also needs to better defined; the definition indicates it is the technique used, but the technique (HPLC) is not the same as a method (experimental instructions how to use HPLC, make samples, etc. to execute a test properly).
7	Analysis Graphic	Analysis Graphic not typically in S.3.2; representative chromatograms may be provided in S.2.3, S.2.4, or s.4.2 as required.
8	Analytical Instrument Data File	It is unclear if providing the file names or types adds value to the eCTD unless active access to a common format like ADF is provided.
9	Analytical Instrument Data File Type	It is unclear if providing the file names or types adds value to the eCTD unless active access to a common format like ADF is provided.
	Additional Information (comments):	Categorization of impurities is missing elemental impurities, stereoisomers, degradants, and mutagenic impurities; terminology around fate and purge missing; terminology from ICH M7, ICH Q3A, Q3C, and Q3D missing (both as a control strategy reference and also to include the needed sub-terminology such as Options 1-4, purge prediction, etc). Significant additional vocabulary would be needed to build complete ontology, but this is a good start. Consider revising the Inorganic Impurities definition to include components and container as additional sources per ICH Q3D.
	14. Drug Product Impurities	
1	Drug Product Impurity Name	A compound may be both an impurity and a degradant so it would be useful to be able to assign either one or both roles within the terminology.

2	UNII	No comment.
3	Impurity Classification	It is suggested that this element be expanded to include mutagenic impurities.
4	Chemical Structure Data File	/nclude .cdx.
5	Impurity Structure Graphic	No comment.
6	Drug Product Impurity Method Type	Suggested this be deleted. Methods must be suitable for their intended use and certification as YES or NO is not value added. Additionally, the definition indicates this is the technique used, and technique (HPLC) is not the same as a method (experimental instructions how to use HPLC, make samples, etc to execute a test properly).
7	Analysis Graphic	No Comment.
8	Analytical Instrument Data File	A standardized data format like the Allotrope data format could, in the future, enable a new more interactive way to share data if desired.
9	Analytical Instrument Data File Type	It is unclear if providing the file names or types adds value to the eCTD unless active access to a common format like ADF is provided.
15. Analytical Methods Validation		
1	Compendial Method Verification Indicatior	The intent of this element is unclear. Is this a statement of whether the compendial method is suitable for the material? However, a method would not be used unless it was suitable, so in that case, capturing this element does not add value.
2	Validation Title	Capturing the title of a figure could aid future data mining if the format is standardized and sufficiently informative.
3	Test Name	To enable future data mining activities and provide greater value for this element, a standardized format would be needed.
4	Report Number	This is not typically done, but to enable future data mining activities and provide greater value for this element, a standardized format would be needed. There is question around method validation as to whether the agency is changing its expectations with regards to submission of validation data for Phase 1 INDs.
5	Report Date	This is not typically done, but to enable future data mining activities and provide greater value for this element, a standardized format would be needed.
6	Validation Parameter	Instead of including repeatability, intermediate precision, and reproducibility in the definition of precision, these terms should be separated, defined vocabulary terms.
7	Test Usage	No comment.
8	Test Method Origin	No comment.

9	Test Category	Grouping of tests by type of function may have some value in a future ontology if the function is selected from a controlled vocabulary. This grouping would allow queries such as "Show me all test results from the test category 'Assay.' The data element may be more appropriately named "Test Function."
10	Analytical Procedure	The examples provided in definition are Analytical Techniques (e.g. HPLC) and not Analytical Procedures (e.g. full method details), so suggest considering adding both terms with correct ontological definitions.
11	Reference to Procedure	This is not typically included in the eCTD, but to enable future data mining activities and provide greater value for this element, a standardized format would be needed.
12	Validation Acceptance Criteria	This is not typically directly included in the eCTD, but if included, a standardized tabular format is recommended.
13	Batch or Lot Number	The lot number is more appropriate; a validation may also be conducted on a representative laboratory sample, but it should still be possible to assign a lot number in same format to this material.
14	Reference Material Standard	Suggest considering creation of a parent term "Reference Material" with controlled subterms such as primary reference standard, secondary reference standard, and working standard, etc.
15	Validation Results	Substantial defined, subterminology would be needed for this element to support validation concepts used in S.2.3, S.2.4, and S.4.3 for DS (e.g. QbD terminology, statistics terminology, etc)
16	Additional Information	No comment.