A Framework for Extending the Science Traceability Matrix: Application to the Planned Europa Mission

Laura Jones-Wilson
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr.
Pasadena, CA 91109
Laura.L.Jones@jpl.nasa.gov

Sara Susca
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Dr.
Pasadena, CA 91109
Sara.Susca@jpl.nasa.gov

Abstract— **One of the most critical functions of the systems engineering requirements process for a large multi-instrument science-driven space mission is to successfully communicate customer expectations into a comprehensive and traceable science requirements flowdown. These requirements are essential to communicating the constraints on the scope of the science investigations and clarifying how multiple instruments contribute to a given science goal. They also provide insight into how the science goals of the whole mission are affected by design choices. There is little specific guidance available on best practices for developing this science-driven flowdown. A unified Science Traceability Matrix (USTM) contains a significant amount of information that can be leveraged for that purpose, but the USTM was not designed to directly produce a complete science requirements flowdown. Thus, starting with the principles codified in a USTM, the authors propose a framework that directly maps into the requirements flowdown and supports broader systems engineering processes while retaining its meaning to the science team. This Science Traceability and Alignment Framework, or STAF, defines a set of common definitions and valid relationships to structure communication across the project. In addition, STAF populates a network of information that can be useful to support complex mission analysis activities such as fault protection. This work discusses the highest-level implementation of the STAF, the project-domain or P-STAF, which describes an approach to decomposing customer requirements into science requirements. The planned Europa Mission is used as a case study for the implementation of this framework and its potential benefits to a project.**

TABLE OF CONTENTS

1. INTRODUCTION

Systems engineers serve a range of critical functions on NASA missions, but perhaps one of the most important is their role as translators between different disciplines and technical communities. A successful project team communicates most effectively when using a common language, and systems engineers are key in not only developing this dialect, but also leveraging it to achieve a shared technical understanding across the project. Requirements are powerful systems engineering communication tools that constitute the formal sentences in this shared language: they follow a particular syntax to provide specific meaning to all team members. On science-driven space missions in particular, clearly conveying the science needs via interpretable requirements is essential to the integrity of the project's requirements structure. If properly posed (i.e. if written in terms that are meaningful to both scientists and engineers), these science requirements can then be decomposed into a multi-level set of requirements that define the functional and performance needs across the system elements (such as the payload, mission design, etc.) that are necessary to support the mission's science objectives. Collectively, this chain of interrelated information (the "science requirements flowdown" discussed in this paper) forms a powerful network of information that can be leveraged to 1) clearly link engineering designs to science needs and complete the system's requirements traceability, 2) understand the sensitivity of the mission objectives to particular engineering aspects of the design, and 3) (for multi-instrument missions) reveal relationships among instruments that illustrate how they collectively support different science questions. This flowdown is thus clearly a valuable product of the systems engineering process.

The process of building this science requirements flowdown, however, is challenging for more reasons than the fact that the information that must be conveyed in the flowdown is complex, including the facts that: 1) the science performed (especially with multiple instruments) is often highly interrelated, making it difficult to parse out the stronger relationships from the weaker ones, 2) the requirements must be interpretable across diverse technical communities, and 3) existing systems engineering process guidelines do not provide enough detail to address this specific class of requirements.

To develop this product, it is first important to understand how the stakeholders in the flowdown – the project scientists and engineers – approach this problem. Engineers work best when handling distinct categories that clearly identify a set of members, because these categories and sets enable clean demarcations to be made. However, scientists may struggle

978-1-5090-1613-6/17//$31.00 ©2017 IEEE

to describe their investigations in these terms because subtle connections can be lost in that binning process. Thus, any sufficiently simple scheme used to identify different types of science, in terms engineers can use, will inherently capture only an approximation of the true set of science interconnections that exist. The challenge is thus to identify the appropriate level – the relative strength of the relationships – at which science categories are created in order to support both requirements traceability and engineering comprehension without losing meaning to the scientists.

To further complicate the process, scientists and engineers work within different technical spheres that maintain overlapping but inconsistent technical vocabularies. Terms such as "dataset" are often overloaded, causing confusion when a formal project definition does not exist (and sometimes, even when it does). Here, the systems engineer must work to build up a common technical vocabulary that can be used meaningfully in requirements, such as those in the science requirements flowdown. Because these requirements exist at the boundaries of science and engineering stakeholders, they serve as a bellwether for misunderstandings and other communication issues that threaten a project's successful implementation.

Given these complexities and the important role that this flowdown has in ensuring and proving the mission's success, it is perhaps surprising the authors found little published work directly addressing this specific category of requirements. Most guidance provided by standard systems engineering sources [1] [2] [3] is too vague to be useful in addressing the complexity of the topic, or is so specific to other categories of requirements that it is not applicable. None of these sources call out science-derived requirements as a topic of special interest, and most of the available advice on how to manage sets of requirements only provides broad discussions, such as the value of sorting requirements into "established categories" that are left undefined. [3] As a result, each project tends to create its own approach to developing these requirements with varying degrees of heritage, applicability to other projects, clarity, information transfer, and documentation.

In view of these issues, there is clearly a gap in available information on best practices for implementing a framework that can organize the science requirements flowdown in a meaningful way. This paper offers a method of bridging this gap by describing the Science Traceability and Alignment Framework (STAF). STAF propose a structured language and element hierarchy that addresses the challenges that otherwise make it difficult to create a science requirements flowdown. Much like the musical staff, the framework serves as an organizing standard that coordinates the vocabularies of many different project elements and can thus be used to make the conversations (and requirements) accessible to both scientists and engineers. Perhaps the most powerful aspect of STAF is the fact that it has been developed in the crucible of an active flight project, with the input of dozens of individuals (scientists, engineers, and managers on NASA's planned Europa Mission). This input has refined STAF to the point where it is flexible enough to speak across elements of the project and detailed enough to address the idiosyncrasies and logistical issues with its implementation on a large flight project.

Because the STAF was developed during the Europa Mission, its evolution mirrored the requirements structure of this project. The Europa Mission requirements hierarchy has two levels in its science requirements flow: the science requirements and their children the measurement requirements. Thus, the STAF has been divided into two different domains: 1) the project-level domain, or P-STAF, which addresses the development of the science requirements, and 2) the measurement-level domain, or M-STAF [4], which addresses the development of the measurement requirements. This paper describes STAF as a concept, but focuses on the P-STAF level implementation, specifically the decomposition of the customer's level-1 (L1) requirements into science campaigns and associated science datasets from which the science requirements can be developed. The companion paper, [4], focuses on the development of the measurement requirements which serve as the direct parents of the engineering subsystem performance requirements. The specifics of the requirements structure, although used as examples in this work, are less important than the unifying concepts of the STAF as a whole; namely, developing a common project vocabulary within a set of linked concepts that can be used to trace engineering designs to the science objectives that drive them.

This paper first addresses how the STAF builds on and extends the existing science traceability matrix (STM) tool. From here, we provide a detailed description of the vocabulary and taxonomy of the framework. We then explain how to implement and use the P-STAF using examples from NASA's planned Europa Mission to illustrate the process and highlight its value.

Science Objectives	Measurement Objectives	Measurement Requirements	Instruments	Instrument Requirements	Data Products
Science Objective 1	Measurement Objective 3	• Requirement 1 • Requirement 2	Inst 1	• Requirement A • Requirement B	DP 1
Science Objective 2	Measurement Objective 1	• Requirement 3 • Requirement 4 • Requirement 5 • Requirement 6	Inst 2, 3	• Requirement C • Requirement D	DP 2,3
	Measurement Objective 2	• Requirement 1 • Requirement 4 • Requirement 7	Inst 1,3	• Requirement E • Requirement F • Requirement G	DP 4,5,6
	Measurement Objective 3	• Requirement 1 • Requirement 8	Inst 4	• Requirement H	DP 7
	Measurement Objective 4	• Requirement 9	Inst 2,4	• Requirement I	DP 8, 9

Figure 1 The structure of a science traceability matrix as described in [5]

2. EXTENDING THE SCIENCE TRACEABILITY MATRIX

The concept of decomposing the science objectives into discrete elements and then tracing those elements to the necessary measurements and instrument specifications is not new. In fact, a tool known as the Science Traceability Matrix (STM) [5] has been available to projects for nearly a decade, and often forms an important part of a mission or instrument proposal to NASA. Because it has already been adopted by segments of the space science community, it is a familiar tool that can guide the conversation on the distinctions among different science objectives and how different measurement classes, observations, or instruments work together to accomplish them.

The STM, a generic example of which is shown in Figure 1, is a two-dimensional representation where the science objectives are on the rows, associated measurement objectives populate the subrows, and the column headings identify the associated measurement requirements, instruments, instrument requirements, and data products as they map to those science and measurement objectives. When laid out in this structured way, the relationships between science, measurements, instruments, and data are much easier to follow. The STM is particularly effective at highlighting cases where multiple instruments are needed to collect certain measurements. The row-and-column structure of the STM provides a path from the mission or science objectives down to the science data products simply by tracing along a row. In fact, the STM can be thought of as a highly-interconnected network where science objectives and instrument measurements are nodes and the connections amongst them are shown in the structure of the matrix rows. Organizing information in this way enables systems engineers to open trade spaces between science needs and system designs [5]. By all counts, the STM is a valuable tool that brilliantly represents the linkages it was designed to illustrate – how science objectives can be traced down to measurements, instruments, and data products.

Knowing that the STM is so powerful, it is natural that systems engineers seek to leverage the information captured in it to develop other products across the project; especially the requirements flowdown. Yet, when pursuing this line of inquiry, it becomes clear that the STM excels at the purpose for which it is designed, but is difficult to extend further for application to the requirements. For one, the STM is often written by and for scientist stakeholders exclusively, making it difficult for engineers to parse and leverage in the requirements flowdown. In fact, a project may not update their STM after the proposal stage due to resource constraints (as noted by Weiss et. al. [5]). Even if a project does take this step (as did the Europa Mission), the resulting Unified Science Traceability Matrix (USTM) might not be a governing document and, hence, never be seen by the bulk of a project's engineering team.

Assuming that the project produces a USTM, in its traditional form the tool is not designed in such a way to enable its application to a full requirements flowdown development. The primary reason for this is that, as scientists create the USTM, they are often interested in capturing the full breadth of the linkages between the science objectives and instrument measurements so that the full impact of a given investigation and the full complement of achievable science is clear. Thus, the USTM may preserve even relatively subtle relationships in the science along with the strong, requirements-driving relationships. Because the USTM does not provide a relative weight to the connections it illustrates, the densely connected network is difficult to query, write requirements against, and use to gain insights into the system sensitivities.

Another difficulty in using the USTM in a science requirements flowdown is that it does not specifically disallow many-to-many mapping relationships across its columns. For example, the "science objectives" in the first column of most USTMs may be written so that they link go many of the L1 requirements. This type of mapping makes it difficult to understand which measurements contribute to which L1 requirement directly, and can overestimate the sensitivity of the L1 requirements to any given measurement. Subsequent analyses may then degenerate into a problem where the loss of any measurement immediately impacts all of the customer requirements, which may not be realistic in the context of engineering trades.

Finally, although the USTM captures some subset of the key and driving requirements in its columns, the tool was not designed to provide an exhaustive way to generate requirements or categorize the relationships among requirements listed in any given column. The USTM was instead designed to provide information on the relationship between instruments, measurements, and the science objectives in the row space of the matrix. Yet it is clear that these requirements are likely to be sortable in ways that could be useful to leverage in a complete science requirements flowdown.

The STAF seeks to formally address each of these challenges by extending the USTM into a tool that can be better leveraged to create a science requirements flowdown by:

1) Defining a formal vocabulary of terms that can be understood in both the science and engineering communities and tying that taxonomy to the requirements flowdown so that it can be used in governing project documents to constrain elements at the appropriate project level.
2) Limiting the relationships formally codified in the framework to those that are strong enough to be captured as requirements. More subtle connections among the measurements and science objectives can still be codified in a non-governing STM structure, but are not formally tracked by STAF to limit the extent to which weaker connections drive the engineering trade space.
3) Enforcing a structured one-to-many mapping among most elements of the framework in order to provide better understanding of the sensitivities in the system.
4) Expanding on the categories and organization of items within the "column space" of the USTM, which supports cross-checking and completeness evaluations across requirements

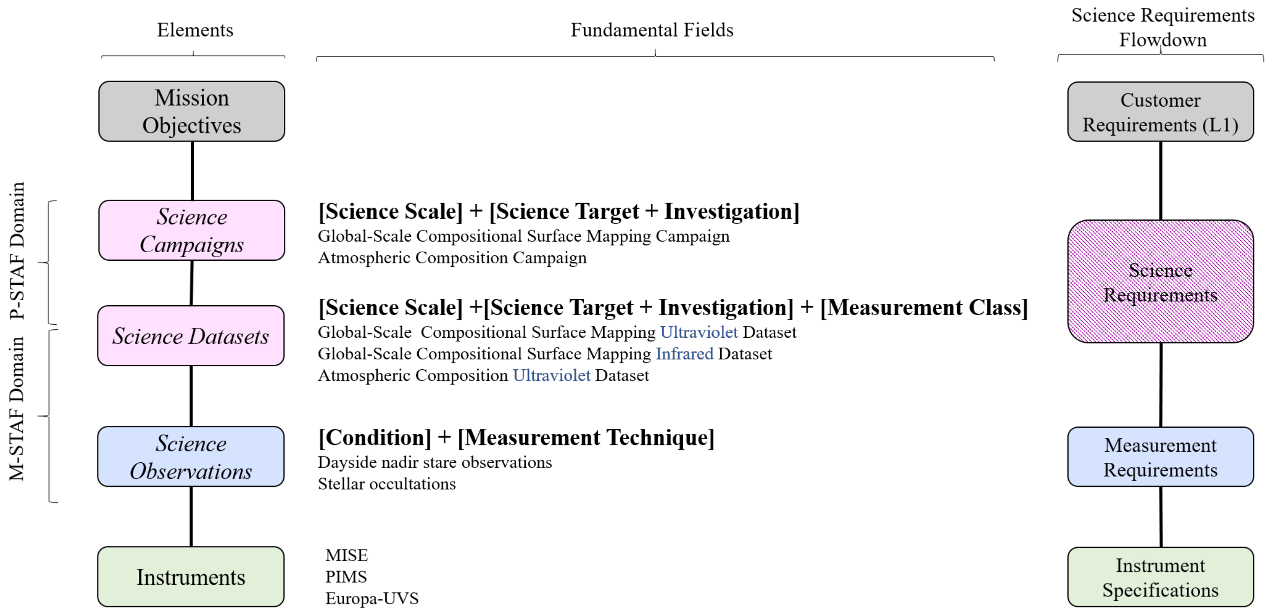

Figure 2 The basic taxonomy of the STAF and its relationship to the science requirements flowdown

All of these benefits come from the same product – a common language with a defined set of elements and structured relationships between those elements that can be mirrored in existing engineering and science team processes via the requirements flowdown and USTM respectively. Collectively, this taxonomy is the Science Traceability and Alignment Framework or STAF. This self-consistent language can be used by all stakeholders when interacting with one another – a device known as a pidgin in social science circles. [6]

3. Framework Description

STAF Hierarchy of Elements and Fundamental Fields

STAF uses the organizational concept of "elements" to describe aspects of the science requirements flowdown that can be meaningfully constrained (and are therefore able to serve as the subject of a requirement). These elements are written hierarchically from the most specific elements in STAF – the instruments – to the most broadly scoped elements – the mission objectives. Each of these elements can pair to a level in the science requirements flowdown, as shown in Figure 2.

The key elements in the STAF are: *mission objectives, science campaigns, science datasets, science observations,* and *instruments*.

The *mission objective* element is codified by the associated L1 customer requirements, which are taken as inputs to this framework. Although similar to the "science objectives" suggested by the first column of the USTM, the mission objectives here are directly tied to the language of the L1 requirements and the mission success criteria. This is an important distinction because the project must report on its progress toward meeting its L1s and success criteria. The USTM "science objectives" are often developed in the proposal stage of a project and therefore were created before the L1 requirements were ever negotiated. Thus, they do not necessarily derive from the L1s in such a direct way, meaning that it is possible for the first column of the USTM to not immediately map into the requirements structure for the project. Even more to the point, the customer requirements may not follow traditional science categorizations depending on the priorities and focus of the customer. Thus, "science objectives" that are unmoored from the ultimate customer requirements may not mirror the idiosyncrasies of a given L1 set of requirements. This reality makes it somewhat cumbersome to go back and attempt to reshuffle a set of science objectives to enforce a one-to-many relationship with the L1s. STAF sidesteps this issue by initiating the framework with an element that directly mirrors the customer's requirements and thus the project's ultimate measure of success.

At the other end of the framework, STAF defines the *instrument* as an element. Any requirements written on the instruments (or other subsystems on the spacecraft) are outside the domain of the STAF requirements flowdown, but they often drive the needs of the requirements at higher levels and so are integral to the understanding of the STAF. The fact that the instrument is the lowest-level element in the STAF is important because it highlights one of the common sources of confusion in the science requirements flowdown. Instruments are the design realization of a need for a given type of measurement; strictly speaking, the measurement and science requirements can be written independently of a given instrument. It can be difficult to enforce this level of abstractness in higher-level requirements because the instruments have typically already been selected when the flowdown development process gets underway. When we write about an "ultraviolet" measurement, it is clear in the context of the Europa Mission that we are talking about measurements collected by the only ultraviolet instrument on the spacecraft, Europa-UVS. It can seem unnecessarily obtuse to write requirements like "the ultraviolet observations shall" rather than skipping that step and writing "Europa-UVS shall". The distinction between them is subtle, but the latter representation implies that the instrument team will verify that requirement at a lower level of the project, which may not be appropriate given the broadness of a measurement or science requirement. Similarly, the latter representation can obscure trade options in the case where instruments have redundant capabilities and either one could make a given observation.

In between these bounding elements (*mission objectives* and *instruments*) in the STAF, the three remaining elements require more detailed explanation because they form the core of the framework. They will be described in the context of the "fields" proposed by the STAF.

STAF uses the concept of "fields" to categorize types of information that distinguish between different elements of the same type. For example, a dayside image and a nightside image are both science observation elements, but they are taken under different lighting conditions and are therefore distinguished by the "condition" field. The fundamental fields proposed by the STAF represent a minimum set of distinctions that are necessary to properly serve all of the stakeholders of the science requirements flowdown, although this concept could be expanded to capture other ways of delineating between elements when such distinctions are important to capture in the requirements flowdown.

The five fundamental fields in STAF (**science scale, science target + investigation, measurement class, measurement technique**, and **conditions**) are those that identify important distinctions among the main three elements in the STAF hierarchy, as shown in Figure 2. Each of these elements will be described in more detail in subsequent sections, showing how the fundamental fields identified here help distinguish between the different members of a given element.

Table 1 Fundamental fields in the P-STAF and their basic description

Science Scale	When identified as relevant, [science scale] is used to distinguish between science investigations that apply over different geographical distances and therefore study processes that act over different ranges on the target. Although most observations have a specific scale or resolution, the science scale is tied instead to the scale of science questions that are being answered: Global-scale science studies effects that have hemispheric or larger implications, Regional-scale science focuses on unique features, landforms, or specific topography Local-scale science examines the target properties on a human- or lander- scale
Science Target + Investigation	Used to identify the different physical objects under study and the science purpose of that investigation. For example, surface geology vs. surface composition, vs. atmospheric composition.
Measurement Class	Used to distinguish between science investigations that collect unique kinds of information and those that may provide robustness because they collect the same type of measurements. For example, radar vs. magnetic vs. ultraviolet measurements.

As noted in the introduction, the STAF can be split into two domains, the P-STAF and the M-STAF, that differ in the project level at which they are implemented and the stakeholders involved in their development. This paper focuses on the higher-level implementation of the framework, the P-STAF, which includes the fields of **science scale**, **science target + investigation**, and **measurement class** to define the elements of *science campaigns* and *science datasets.* Table 1 shows the fundamental fields of the P-STAF domain. Our companion paper, [4], starts with the assumption that the *science datasets* are established in the P-STAF domain and then describes the fields of **measurement technique** and **conditions** in order to define unique *science observations* that relate to the measurement domain of the framework, or M-STAF. When implemented together, the more strategic value of the P-STAF can be combined with the more tactical value of the M-STAF to provide a rich common language and syntax that both the engineers and scientists can navigate and leverage to benefit the project.

Science Campaigns

In the elements and fields comprising the P-STAF, the framework defines an element below the mission objectives called *science campaign*. A *science campaign* is a construct used to group together related science investigations that study similar hypotheses or related scientific features/targets. Each science campaign must address a single mission objective (and thus trace to a single L1 requirement). A given mission objective or L1 requirement, however, many link to multiple science campaigns, in order to provide more resolution in the science being studied.

The term "campaign" is overloaded and can carry many conflicting meanings across the science and engineering communities, so it is important to be clear about this use of the term. In some communities, the word campaign refers to a specific operations plan or sequence that is designed to investigate a given science target, but this is not how the term is used in this framework. STAF instead uses the term in the sense of a strong science "theme" – a categorization of different hypotheses or features that as a whole the mission intends to study.

The STAF proposes that the science campaigns need to capture several important distinctions that may matter to the scientists and the subsequent flowdown of requirements. These distinctions include science scale (e.g., "global-scale" science may differ in some important ways from "local-scale" science), the science target (e.g., "surface composition" and "atmospheric composition" are distinct targets for compositional science), and the type of hypothesis being tested (e.g., "plume search" tests a different hypothesis than "plume characterization" for the same target). This information is codified into two STAF fundamental fields: the **science scale** and **science target + investigation**.

The **science scale** is included in the science campaign when it is necessary to distinguish between science performed over different extents or geographical ranges. For example, if the mission is interested in studying processes or scientific hypotheses that affect a hemisphere or larger of a planetary body, but also wants to study as a distinct investigation the properties of the surface on a scale to assess landing sites, these two investigations can be distinguished by a science scale. This distinction is particularly valuable when the scale of the science necessitates different types of observations or measurement qualities. It is important, however, to make a distinction here between the science scale and the pixel scale of an instrument or measurement. The **science scale** field is intended to capture the scale over which the processes or hypotheses apply, not necessarily the pixel scale of any given measurement. Studying processes that act on regional scales, for example, may require measurements with much finer resolutions that may not be classified as "regional-scale." A diagram explaining how the different scales can be defined is shown in Figure 3. It is important to consider both vertical and horizontal scales in these definitions, and in some cases a time scale may be appropriate.

Given this definition of **science scale**, it is worth noting that it is not always a part of the definition of a science campaign. Although science scale often serves as a second dimension to the science investigation being developed – such that nearly every hypotheses or process being studied could in theory be categorized into specific geographical or dimensional scale – it is not always valuable to make this distinction. For example, the science campaign may be "surface geology", rather than the distinct "global-scale surface geology" and "local-scale surface geology," depending on the focus of the investigation and the specific customer objectives.

The second field in the science campaign, the **science target + investigation**, defines a science target (such as the "surface" or "ice shell"), and a short descriptor of the type of science being performed or hypothesis being tested (such as "search" or "mapping"). These distinctions are grouped together in the fundamental field because it became clear that the scientists often think of the target as closely linked to the hypothesis under test, and separating out the two pieces of information is not helpful. Other projects may find value in identifying these separately, at least when first trying to construct the list of science campaigns. Selecting a list of appropriate **science target + investigation** is highly dependent on the specifics of the mission; more information on how to develop this list is discussed in Section 4.

Science Target	Scale Range and Definition	
	Depth Scale	Surface Scale
External Environment	Above Hill Sphere	[]
Exosphere	B km to Hill Sphere	Scale 2
	A - B km	Scale 1
Surface	0 km	Global Regional Local
Subsurface	0 - X km	Near Surface
	X - Y km	Shallow
	Y - Z km	Deep

Figure 3 An example of a notional scale definition table suitable for a planetary science mission

When these two fundamental fields are combined, the name of the science campaign can be constructed. For example, "global-scale surface mapping" or "atmospheric composition" or "active plume search" are all examples of science campaign names. Regardless of any specific naming convention, however, it is most important that the scientists understand what types of investigations fall into each science campaign (possibly clarified by having the project scientists create short descriptors of each campaign) and that the science campaign names are descriptive enough to be clearly understood by a generally knowledgeable scientist or engineer on the project.

Although any given science campaign should only address a single customer L1 requirement, an L1 requirement may be broken into any number of contributing science campaigns. Preserving this one-to-many mapping enables greater insight into system analyses later in the mission. The science campaign list and its associated mapping to the mission objectives should originate from the project science office, although steps for guiding that process are described in Section 4.

Science Datasets

At a level below the *science campaigns*, STAF proposes an element called a *science dataset*. The science datasets are groupings of observations or measurements collected by a single instrument that can collectively be used to address a given science campaign. *Science datasets* support a given *science campaign*. The word dataset is yet another term that is often used in different ways by different communities, so it is important to be clear about what data populate a science dataset. The group of data from any type of *science observation* generated by a single **measurement class** that contributes to a specific *science campaign* makes up the *science dataset*. Because observations may support many different science campaigns, the science datasets of a given instrument may include overlapping information. Collaborations across instruments are represented in the science campaign level of the taxonomy, where many science datasets can support a given science campaign.

Science datasets are essentially the science campaign with added distinguishing feature: the **measurement class**. The measurement class, the third STAF fundamental field, is a generalized way of framing an instrument's type of measurement. For example, an ultraviolet imaging spectrograph generates ultraviolet imaging spectroscopy measurements (or ultraviolet measurements for short). The measurement class for that instrument, depending on the scientist's preference for brevity, may be "ultraviolet imaging spectroscopy" or simply "ultraviolet." Therefore, one might have a science dataset called the "atmospheric composition ultraviolet dataset," or an "atmospheric composition plasma dataset", or a "global-scale surface composition visible dataset." This naming makes it clear that the same kind of science is being supported by three different classes of measurements.

Here it is worth noting that the term dataset simply means a grouping of data. One might choose to define a variety of different ways to group data to meet different needs. For example, it is possible to define "measurement class" datasets (i.e., "the ultraviolet dataset") that encompasses all measurements collected by a given measurement class. The STAF taxonomy explicitly identifies a science dataset, where measurements are grouped by the science campaign they support, as a distinct category from other equally valid data groupings.

When categorizing data this way, it is possible (perhaps even likely) that the same observations support many different science campaigns. This science observation to science dataset mapping is the only allowed many-to-many mapping relationship in the STAF, and it is allowed because, while science campaigns or science datasets are somewhat artificial (but useful) ways of grouping information, the measurements, by their nature, can support many different science categories. This mapping relationship simply reflects a reality. Thus, the STAF is designed to accommodate this reality and can easily recommend ways to codify this relationship in the requirements.

The advantage in separating the measurement class field from the instrument element is not always obvious, since the measurement class may map one-to-one with the selected instruments on some missions. It may be tempting to write the names of instruments into higher-level requirements rather than using the measurement class to make the distinction. However, this level of abstraction does have uses that are not perhaps immediately obvious. For example, the measurement class enables the definition of science-useful measurements that are not necessarily collected by a science instrument. For example, the "gravity" measurement class may be defined if there is a unique type of measurement of this type. However, these gravity measurements may derive from the telecommunication antennas and not a specific scientific instrument. Abstracting out the type of measurement rather than writing high-level requirements with respect to a specific instrument avoids these unique scenarios where the "instrument" is in fact an engineering subsystem. Beyond this special case, however, the concept of a measurement class also enables the project the expose any redundancy that comes with the selected payload. For example, it is possible to write requirements that say "The observations made by Europa-UVS shall…" rather than "The ultraviolet observations shall…" In the first case, a second instrument sensitive to the ultraviolet part of the spectrum would not be able to provide redundancy to the requirements set. Similarly, if the payload includes both a narrow-angle visible camera and a wide-angle visible camera, the high-level requirements do not necessarily need to specify which camera collects exactly which kind of observations – it may be a trade that can be better performed at lower levels. Thus, the measurement class field allows the requirements to specify the kind of measurement needed without unnecessarily over-constraining the option space at lower project levels. Table 2 shows the mapping between the selected instruments, investigations, and their proposed measurement class for the Europa Mission.

Table 2 Example mapping between selected instrument and measurement class for the Europa Mission

Selected Investigation	Selected Instrument	Measurement Class
Gravity Science	(none selected)	Gravity
SUDA	SUDA	Impact Mass Spectrometry (IMS)
MISE	MISE	Infrared
ICEMAG	ICEMAG	Magnetic
MASPEX	MASPEX	Neutral Mass Spectrometry (NMS)
PIMS	PIMS	Plasma
REASON	REASON VHF REASON HF	Radar
ETHEMIS	ETHEMIS	Thermal
Europa-UVS	Europa-UVS	Ultraviolet
EIS	EIS NAC EIS WAC	Visible

Science Observations

The science datasets are populated by individual *science observations* taken by a single measurement class. These *observations* can be thought of as the individual data unit that any given *instrument* generates. The observations can also be called any number of unique names: images, cubes, scans, groundtracks, or simply measurements. Some missions may find it useful to make a distinction among these terms, but the STAF does not.

Observations are uniquely defined by the last two STAF fundamental fields (**measurement technique** and **condition**) but are out of the scope of this paper. More information can be found on these distinctions and the M-STAF domain of the framework in the companion paper. [4]

4. P-STAF IMPLEMENTATION AND USES

The implementation of the P-STAF domain of this framework can be broken into four steps. Each step essentially captures a unique kind of conversation and consensus that must be established before moving to the next phase. The first step is to develop master list of science campaigns. The second step is developing a master list of science datasets. In the third step, the science campaigns and science datasets are linked to their appropriate level-1 requirement and placed in a matrix which shows this information in a convenient visual format. The fourth step involves writing the text of the science requirements. These four steps are described in detail below.

Step 1: Writing the Science Campaign Master List

As the first element of the STAF, the *science campaigns* must be created first. Because a primary purpose of STAF is to link to the L1 requirements and the customer mission objectives, it is essential to use them as a starting point. Key words in the L1 requirements may suggest the basis for a list of apparent targets and hypotheses that can be codified as **science target + investigations**. First, look for words that identify the object under investigation (the "target"). For planetary science, this target may be a specific body, if the mission covers multiple bodies, or a specific feature of that body – the surface, subsurface, exosphere, etc. If the target is missing from a requirement or a proposed campaign name, probe to see if adding a target distinction is useful. For example, if a mission objective is to search for thermal anomalies, and someone proposes a "thermal anomaly search campaign," it is useful to establish if the intent of the customer and the science team is to look for thermal anomalies in both the surface and the exosphere. It is possible that they intended to limit the scope to one unspoken target and that adding the name of the target (i.e., "surface thermal anomaly search") clarifies the scope of the campaign. Alternatively, it may mean that there are in fact two different campaigns ("surface thermal anomaly search" and "exospheric thermal anomaly search") because the techniques or measurement classes that support the investigation of those different targets are different enough to merit a distinction. Or, if the scientists and/or customers do not see value in making such a distinction, then the campaign name remains ("thermal anomaly search"). It is also worth noting that some investigations imply a target and that adding it may be redundant. For example, geology investigations necessarily deal with the surface features of the body, so "surface geology" may not be a useful distinction to carry in the name of the campaign. Once establishing the target of a campaign, one must also identify what kind of activity is being performed, or what discipline in science is most applicable. For example, a "search" activity may be distinguished from a "characterization" activity, and "geology" can be distinguished from "chemistry." These terms may also show up as key words in an L1 requirement, or they may be implied by the hypotheses under study by the mission objectives. In some cases, it may be worthwhile to make a distinction that implies the way in which the hypotheses will be studied. For example, the distinction between "Active Plume Search" and "Inferred Plume Evidence" is that the measurements contributing to the former campaign will involve regularly-planned scans of the disk and activities designed to seek out plume activities. The latter campaign captures measurements that may imply something about plume activity, but are serendipitous and not directed specifically to look for plumes. This distinction allows the team to separate the mission's primary "plume hunters" from the instruments that may still contribute to plume science but do not plan to direct activities to the search.

Ultimately, when selecting an appropriate list of **science target + investigations**, the terms should be short and clear enough to convey a specific type of science and purpose, but not so simple as to cause confusion among the scientists. Explaining that these terms are simply a shorthand for divisions in the science that can be explained in detail elsewhere may help ease concerns about the necessary brevity of category titles. (More detailed descriptions may be contained in a concept description in the system model, or can be captured in a project memo). For example, the term "Subsurface structure and dynamics" campaign can be difficult to use in requirements, if only because of its length. A requirement may say something like "For the subsurface structure and dynamics radar campaign, the VHF sounding measurements shall…" If the scientists are comfortable shortening the name to "subsurface structure" campaign, it will lead to requirements that are less verbose and perhaps more clear as the STAF is implemented at lower levels in the flowdown.

The list of **science target + investigations** is best developed collectively by the project science group (PSG and Project Science Office), or the science management team in collaboration with the scientists on each investigation team. Because these categories have implications throughout the science requirements flowdown, earning consensus on these categories early on in the process significantly smooths the implementation of the rest of the STAF. With this guidance as a starting point, the PSG and the Project Scientist can refine the list by considering that this list of **science target + investigations** should contain distinctions that can provide insight on how the mission as a whole is performing when an engineering analysis of the current design is executed. So, as long as the list can be sorted into a one-to-many mapping with the L1 requirements, the framework is quite flexible to the specific naming conventions and distinctions they choose to capture.

Another element to consider in developing science campaign names is that many science investigations can be can be performed on a different **science scale** (like that shown in Figure 3). By calling it out as a separate dimension of the science campaign, the scientists involved have the option of distinguishing the science by scale or grouping all scales together into one investigation category. There are four considerations that drive this decision:

1) Do the customer requirements describe different scales? If so, the science campaigns should also reflect this distinction.
2) Does the science management team see value in understanding analyses broken out by how certain science scales are affected? If for example radiation faults were to strongly affect local-scale geology but were not found to significantly affect global-scale geology, would that distinction influence their decisions about how to address the issue versus if they were told that simply "geology" science was affected? If that resolution in the reporting is valuable, the science campaigns where this applies should include a science scale.
3) Do different instruments work together to address science at different scales? If the near-surface science involved a significantly different cast of instruments than the deep-subsurface scale science, the use of science scale in the campaign should be strongly considered.
4) Do the distinctions between these science scales necessitate significantly different science observations or measurement requirements? If so, identifying that unique (and perhaps driving) scope as a separate science campaign may be valuable in trades as the project progresses.

Once the **science scale** is added appropriately to the **science target + investigation**, the campaigns master list should be revised as a set in order to homogenize the scope across the science campaigns in the list. If one science campaign seems too detailed, or is only supported by one instrument, that particular science campaign name should come under more scrutiny. Ensure that it really is necessary to call out that type of science separately from broader, more inclusive categories. Similarly, if an L1 has only one supporting science campaign with many instruments that contribute to the science campaign, it is worth investigating if that science campaign is not specific enough and if there is value to the science management team in reporting out engineering analyses in a finer resolution. The goal of this exercise is not to forcibly homogenize the mission science list into a handful of short phrases, but rather to ensure that the set of science campaigns is developed intentionally and collectively, and not driven by or defined only by one instrument or investigation team.

Step 2: Writing the Science Dataset Master List

Starting from the agreed-upon science campaign master list, the science dataset master list is constructed using the list of **measurement classes** (such as the one in Table 2) and mapping which measurement class supports which science campaign in <u>requirements space</u>. This distinction is important because while many investigation teams will identify a contribution in the USTM, they may not want to have that contribution derive requirements on the system for any number of reasons. Thus, ensuring that the science campaign to science dataset mapping is mapped only on connections that should appear in requirements is a critical difference between this P-STAF and the USTM. In other words, STAF asks "should this connection be captured in the requirements" (or, alternatively, "should this connection be used to make engineering decisions about the system?"), and any relationship that does not meet that threshold is set aside to be captured in other forums (such as the USTM). This process pares down the links to reveal the stronger relationships among the nodes of the science information network (Figure 4).

However, the relationships between instruments can be far more complex than just contributing to the same science campaign. It is possible, in fact, that one instrument can achieve better performance if it can use data from another instrument (measurement class) to interpret its own. This can be captured in P-STAF using a "support" dataset. For example, visible camera data can be made into a digital terrain map that can be used to improve the quality of radar data by allowing the radar team to model the returns generated by the terrain. In this case, the visible imager is providing a unique set of data to the radar team to improve the quality of the return. This set of data may require a separate resource allocation such as data volume or energy and should be tracked at the same level as other science datasets. An example of these support datasets in the science information network can be seen highlighted in Figure 4.

Before constructing the P-STAF matrix, the science information network can be examined as a whole to ensure traceability and remove invalid connections. As shown in Figure 4, the STAF framework is designed to ensure that each science campaign contributes to only one customer requirement, and that each science dataset supports only one science campaign. This system then becomes analyzable in profound ways. For example, because of this one-to-many mapping in the P-STAF domain, it is possible to identify which customer requirements contain the most scope. It is also possible to understand which science is affected if one instrument / measurement class for example becomes unavailable. Engineers can report back fault and reliability analyses can rank specific science campaigns for their vulnerability to specific types of faults (say, radiation-induced faults). Similarly, science reliability requirements can be written on an appropriate level of the project. For example, writing a requirement on the confidence level of returning 80% of a science dataset is different than returning

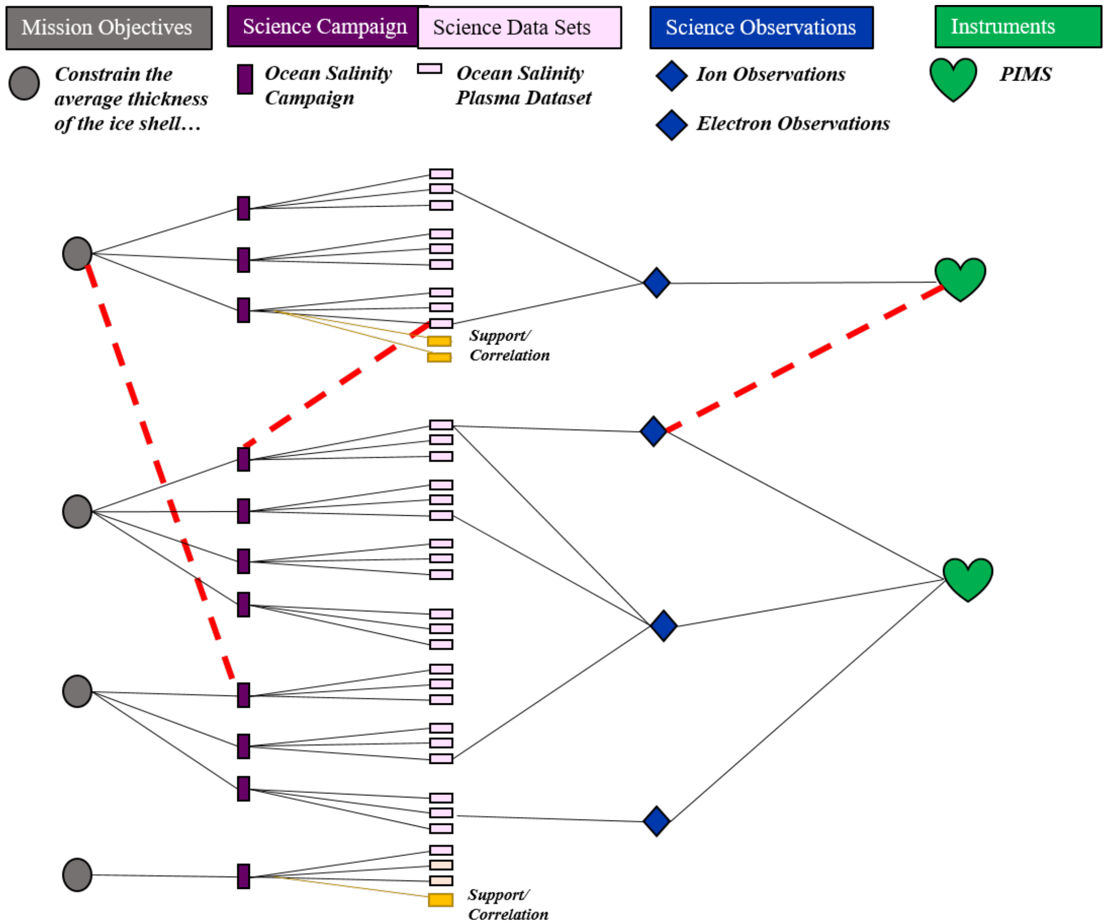

Figure 4 An example of a notional STAF science information network with valid (solid) and invalid (dashed) connections highlighted that emphasize a one-to-many relation between all elements except the Science datasets and science observations.

80% of a given set of science observations or 80% of the data contributing to a science campaign. The project systems engineering team can now make intentional calls about which element to constrain in order to best achieve their intent.

Step 3: Constructing the P-STAF Matrix

Given these master lists of science campaigns and science datasets, it is now possible to place it in a matrix form to reflect its STM roots. This new P-STAF matrix modifies the row-and-column meaning slightly to capture the links to the L1 requirements and preserves a one-to-many mapping as it flows from left to right. An example that was developed as part of a case-study involving the Europa Mission can be found in Figure 5. The first two columns show the candidate mission success criteria and the candidate L1s. The content in both these columns is an input to P-STAF. The main P-STAF contribution is in establishing the mapping between L1s, the science campaigns, and the science datasets which are shown in the last two columns of the matrix. Thus, the P-STAF matrix shows at glance the traceability from mission success to each measurement class for the selected payload. This mapping is further extended in the idea of the M-STAF, which focuses on how the given science datasets list can be categorized by measurement class and further developed to show relationships among the science observations and requirements imposed on them.

By comparing P-STAF with the STM one can see how the link established between science objectives and instrument in the STM is made more general by the link between science campaigns and measurement class but more importantly, how the idea of science objectives is expanded into science campaigns that are generated by a structured dialog between a large and diverse group of stakeholders.

Step 4: Writing Science Requirements

Given the STAF taxonomy, the project now has the freedom to write requirements in the science requirements flowdown to constrain any of the elements it deems appropriate. For P-STAF in particular, this leaves the question of how to address the requirements that come hierarchically after the L1s – the science requirements. The STAF does not prescribe the subject of these requirements, but does imply that there are two options: science campaigns and science datasets. Both have a unique profile of advantages and disadvantages.

<u>Science Requirements Written on Science Campaigns</u> – Because science datasets are constructed by identifying the contribution of a specific measurement class to a science

	Candidate Mission Success Criteria	Candidate Baseline Level 1	Candidate Science Campaigns	Candidate Measurement Class Supporting Science Campaign
Ice Shell & Ocean	Confirm and constrain the depth to the subsurface ocean, and determine processes of surface-ice-ocean exchange.	Map the vertical subsurface structure beneath ≥50 globally distributed landforms to ≥3 km depth[, to understand the distribution of subsurface water and processes of surface-ice-ocean exchange].	Deep Subsurface Structure	Radar (with Visible Support)
			Shallow Subsurface Structure	Radar (with Visible Support), Thermal
		Constrain the average thickness of the ice shell, and the average thickness and salinity of the ocean, each to ±50%.	Ice Shell Properties	Gravity (with Visible and Radar Support), Magnetic, Plasma, Radar (with Visible Support), Visible
			Ocean Properties	Gravity (with Visible and Radar Support), Magnetic, Plasma, Visible
Composition	Identify the composition and sources of key non-ice constituents on the surface and in the atmosphere, including any carbon-containing compounds.	Create a compositional map at ≤10 km spatial scale, covering ≥70% of the surface [to identify the composition and distribution of surface materials].	Global-Scale Compositional Surface Mapping	Infrared, IMS, Ultraviolet, Visible
		Characterize the composition of ≥50 globally distributed landforms, at ≤300 m spatial scale [to identify non-ice surface constituents including any carbon-containing compounds].	Landform Composition	IMS, Infrared, NMS, Radar (with Visible support), Ultraviolet, Visible
		Characterize the composition and sources of volatiles, particulates, and plasma, with sensitivity sufficient to identify the signatures of non-ice materials including any carbon-containing compounds, in globally distributed regions of the atmosphere and local space environment.	Atmospheric Composition	IMS, NMS, Magnetic, Plasma, Radar, Ultraviolet
			Space Environment Composition	IMS, NMS, Magnetic, Plasma, Ultraviolet
Geology	Produce a ≤100-m spatial-scale map over ≥30% of the surface, and determine the three-dimensional characteristics of major landform types at higher resolution.	Produce a controlled photomosaic map of ≥80% of the surface at ≤100-m spatial scale[, to map the global distribution and relationships of geologic landforms].	Global-Scale Surface Mapping	Thermal, Visible
		Characterize the surface at ≤25-m spatial scale, and measure topography at ≤15 m vertical precision, across ≥50 globally distributed landforms[, to identify their morphology and diversity].	Landform Geology	Radar (with Visible support), Thermal, Visible
		Characterize the surface at ~1-m scale to determine surface properties, for ≥40 sites each ≥2 km x 4 km.	Local-Scale Surface Properties	Infrared, Radar, Thermal, Visible
Current Activity	Search for current activity.	Search for and characterize any current activity, notably plumes or thermal anomalies, in regions that are globally distributed.	Active Plume Search	Thermal, Ultraviolet, Visible
			Inferred Plume Evidence	IMS, Infrared, Magnetic, NMS, Plasma, Radar Thermal, Visible
			Surface Thermal Anomaly Search	Infrared, Thermal
			Surface Activity Evidence	Infrared, NMS, Thermal, Visible

Figure 5 Example of P-STAF Table based on the Europa Mission

campaign, and science campaigns are built to capture the synergy among different instruments, it is likely that the mission will have more science datasets than science campaigns. Thus, constraining the science campaigns in the science requirements (e.g. "The atmospheric composition campaign shall…") will generate fewer requirements, perhaps making them easier to manage. However, one potential drawback of this approach is that the science campaigns likely span multiple investigation teams, and thus any requirements written to constrain them would require the buy-in of more than one Principal Investigator. If the requirement generating process is extended to the co-Investigators, it means that the consensus has to be built amongst a large group of people. But this is not the only challenge.

This approach to science requirements necessitates understanding early on in the requirements development process how the different measurement classes relate to each other, but it can, as a result, seed important conversations between the science and the engineering team about the true high-level science needs. For example, if a mission has an ultraviolet and an infrared measurement class, they may both contribute to a global-scale surface compositional mapping campaign but given their different spectral range, can potentially identify a different set of species. For clarity, assume that the presence of species group A can be detected by ultraviolet measurements and the presence of species group B can be detected using infrared measurements. If A and B are not identical, then the project team must determine if the global-scale surface compositional mapping campaign needs to have the ability to detect 1) $A \cup B$, 2) $A \cap B$, 3) A only, or 4) B only. The answer to this question has to be determined via a conversation among the science team, the principal investigators, and the project scientist, and should focus on which measurements are strictly necessary to include in requirements space to address the hypotheses and activities addressed by this campaign. Understandably, it may be difficult to establish which approach is required – obviously, the more species one can map on the surface of the body, the better the science return is. But choosing which set combination will have implications for V&V (verification and validation) of the science requirements and for mission success. For example, if the right answer for the global-scale compositional surface mapping campaign is to constrain the ability to detect species in $A \cup B$, then both the ultraviolet and infrared instruments are needed to verify requirements on this science campaign, and the only redundancy in the campaign is where A and B overlap. This relationship has a different risk profile than a case where either instrument could suffice to address the L1.

Once these issues have been addressed, it is relatively straightforward to write requirements that constrain the science campaign scope. These requirements should generally constrain the spatial or temporal mapping needs of campaign that will ensure sufficient data to draw conclusions related to the campaign hypotheses and the broad qualities of that data such as necessary resolution or scale, accuracy within a range, type of features that should be studied (different landforms, or different species), etc. Here are a few examples of what these requirements might be:

Example of L1 Requirement

1. Create a compositional map at <10 km spatial scale, covering >70% of the surface to identify the composition and distribution of surface material.

Example of Select L2 Requirements on Science Campaigns

- 1.1. The global-scale surface composition mapping campaign shall cover > [TBD] % of the surface.
- 1.2. The global-scale surface composition mapping campaign shall address spatial scales < [TBD] km.
- 1.3. The global-scale surface composition mapping campaign shall be able to detect the following species [TBD] when present at concentrations > [TBD]

Where the TBDs are populated based on the decision of the PSG on how to represent in requirement space the contribution of two or more instruments to that campaign.

Although this approach may require more negotiation across investigation teams and may generate a more abstract set of constraints, it allows the scientists to specifically constrain the kind of information necessary to address the hypotheses and investigations codified in the science campaigns.

Science Requirements Written on Science Datasets – Another approach to the science requirements may be to instead constrain specific science datasets (e.g. "The atmospheric composition ultraviolet dataset shall."). As noted previously, requirements on science datasets will likely form a larger set because there are more science datasets than science campaigns. Yet by constraining science datasets rather than science campaigns is that they are, by design, limited to a single measurement class and thus typically only need the buy-in of a single Principal Investigator. This fact may make it easier to draft the requirements and may make the requirements themselves more relevant and specific to the flowdown because ultimately the measurement requirements will need to be class-specific. Using our example from above, if the science team decides that the science campaign only needs to see species set A to address its hypotheses, then the requirement on the campaign would only be written to constrain A – and the ultraviolet measurement requirements can trace to it as a parent. However, it does not address the infrared requirements, leaving that team to self-generate a requirement at lower levels because they still need a parent to drive the spectral bandpass of their instrument. Writing requirements instead on the science datasets sidesteps this issue because both measurement classes are constrained at this level. On the other hand, one downside to this approach is that it tends to lead to a scenario where all science datasets linked to a science campaign appear to be equally necessary, and the subtleties in risk profiles described previously are not as easy to establish.

These requirements should address the same kinds of topics that those written on the science campaigns may cover, but will instead have one requirement for each class in the campaign, allowing constraint to vary with measurement class. Below we provide an example of a L1-L2 requirement flowdown when the L2s constrain science datasets and both the ultraviolet class and the infrared class contribute to it:

Example of L1 Requirement

2. Create a compositional map at <10 km spatial scale, covering >70% of the surface to identify the composition and distribution of surface material.

Example of Select L2 Requirements on Science Datasets

- 2.1 The global-scale surface composition mapping ultraviolet dataset shall cover > [TBD] % of the surface.
- 2.2 The global-scale surface composition mapping infrared dataset shall cover > [TBD] % of the surface.
- 2.3 The global-scale surface composition mapping ultraviolet dataset shall address spatial scales < [TBD] km.
- 2.4 The global-scale surface composition mapping infrared dataset shall address spatial scales < [TBD] km.
- 2.5 The global-scale surface composition mapping ultraviolet dataset shall be able to detect the following species [TBD] when at concentrations > [TBD]
- 2.6 The global-scale surface composition mapping infrared dataset shall be able to detect the following species [TBD] when at concentrations > [TBD]

Where the TBDs are populated based on each investigation separately.

So while this approach may obscure some subtleties in risk profile and intra-class interactions, it allows individual investigation teams to individually draft specifically meaningful science requirements that link down toward their own measurement requirements.

No Science Requirements – At the beginning of this section we stated that science datasets and science campaigns were both valid options as subjects of the science requirements. However, it is the prerogative of the project systems engineering team to implement a third option, not necessarily recommended by the authors, in which there are no science requirements at L2 and the requirements at the level below the L1s directly constrains the science observations; in other words, that the requirements flowdown skips the P-STAF domain and just implements the M-STAF. If this approach is chosen, it certainly removes the more abstract concepts from the requirements flowdown, perhaps making it easier to draft the requirements because those that constrain observations are in fact so concrete. Even in this case, it is still valuable to construct a P-STAF matrix because the process of developing this matrix requires discussions that will reveal the stronger relationships between investigations and mission objectives, which leads to better insight into the project science as a whole. The strength of the STAF approach is its analyzability from the science observations through the mission objectives, so as long as the science observations can be categorized in meaningful ways to link to the L1 requirements, the specifics of the science requirement wording are less critical. However, systems engineers can recognize the value of codifying these relationships in requirements space if they are going to remain a meaningful way of analyzing the system. Although some of this information might be still gleaned from the USTM, the resulting science information network is not be nearly as analyzable as the one provided by the STAF.

5. Links to M-STAF and Mission Assessments

Once the P-STAF matrix is completed, and the requirements are written at L2, for each measurement class we have a list of science datasets mapping directly to the L1s. If these datasets are then used as starting point of M-STAF, the STAF is fully implemented and its full benefits can be realized [4]. In the flowdown, at this point, the next natural step is to generate measurement requirements that can be further decomposed into engineering performance requirements on spacecraft, mission design, instruments, etc. Together, P-STAF and M-STAF ensure full traceability of the requirements from the L1 down but, even more importantly, generate a network of connections that can be used to provide objective information to the scientists to evaluate the mission design and assess mission margins with respect to the L1s. In Section 5 of our companion paper [4], we explain how it is possible to greatly automate the assessment of a mission point-design with respect to the measurement requirements because, given their nature, they constrain concrete quantities such as amount of surface coverage. On the other hand, the L1s as well as the science requirements are higher-level constraints and require the scientists-in-the-loop to assess a particular mission design against. In our companion paper, we explain at length how implementing both P-STAF and M-STAF allows the engineers to provide richer, more meaningful analysis that highlight science sensitivities to engineering and environmental issues. In turn, the team can work together to understand if they have enough redundancy in the payload to meet the L1s in the face of the many uncertainties on the mission. This information can guide decisions about reliability in the flight system. In other words, P-STAF and M-STAF successfully bridge the cultural gap between scientists and engineers not only in writing the requirements but also, and perhaps more valuably, in the analysis of the mission as a whole.

6. Conclusion

One of the most critical functions of the systems engineering team on a project is to bridge cultural and technical gaps on the project. One of the most important tools a systems engineer has to accomplish this feat is the requirements decomposition process. Requirements in the science requirements flowdown serve as meaningful pidgins that transverse the science-engineering divide to codify a shared understanding of the mission's science needs and its engineering implementations. Yet, few examples are available in published literature to guide a systems engineering team in architecting an efficient and meaningful set of science and measurement requirements.

In this paper we introduce STAF (Science Traceability and Alignment Framework) and its project-level implementation in the P-STAF domain. STAF is a framework that starts with the Science Traceability Matrix (STM) tool, but proposes a language and taxonomy that facilitates more structured communication between the scientists and engineers. As part of its taxonomy, P-STAF defines the concept of science campaigns and science datasets which are used to build the flowdown from the L1s to the science requirements and are the subjects of the science requirements themselves. The flowdown can be then shown in the P-STAF matrix, which forms the basis of the M-STAF domain of the framework. Throughout the paper we provide practical examples from our case-study in the planned Europa Mission.

The process of developing this case study provided invaluable insight into the specific subtleties of codifying science needs into requirements language. The power behind the STAF is in the structure of its common language and defined relationships, and how it can be used to enable more meaningful discussions among the science and engineering teams, not in the specifics of any formula or terminology. Its implementation calls for finesse and flexibility, and a recognition that these guidelines must not supersede the conversations that they are designed to encourage. Ultimately, these conversations can ensure that a mission generates the best science return possible – a goal that everyone on the team can aspire to.

Acknowledgements

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

The authors would like to thank the Europa Mission science team, project systems engineering team, and mission design team for helping in the development, refinement, and ultimately improvement of the STAF concept. In particular, they would like to thank Dr. Robert Pappalardo, Dr. David Senske & Dr. Haje Korth, and Brian Paczkowski (Europa Mission Project Scientist, Deputy Project Scientists, and

Science Manager) for their patience and insightful discussions. They would also like to thank Brian Cooke and John Day (Europa Mission Project System Engineer and Deputy respectively) for their support and insight. They would like to thank Jan Ludwinski, Ben Bradley, Eric Ferguson, Brent Buffington, and Kelli McCoy for their thoughts on how to use the STAF to address critical analyses and assessments across the project.

Finally, the authors would like to thank the Europa Mission payload office, including Valerie Thomas, Valerie Duval, and Kari Lewis for their enthusiasm and support of this concept.

REFERENCES

[1] IOC/IEC/IEEE, "International Standard for Systems and Software Engineering -- Life Cycle Management -- Part 4: Systems Engineering Planning," *ISO/IEC/IEEE 24748-4,* pp. 1-73, 15 05 2016.

[2] ISO/IEC/IEEE, "International Standard - Systems and software engineering -- Life cycle processes -- Requirements engineering," *ISO/IEC/IEEE 29148-2011,* pp. 1-94, 01 12 2011.

[3] NASA, "NASA Systems Engineering Handbook," Washington, D.C., 2007.

[4] S. Susca and L. Jones-Wilson, "A Framework for Writing Science Measurement Requirements and its Application to the Europa Multipl Flyby Mission," in *IEEE Aerospace Conference*, Big Sky, MT, 2017.

[5] J. R. Weiss, W. D. Smythe and W. Lu, "Science Traceability," in *IEEE Aerospace Conference Proceedings*, April 2005.

[6] P. Galison, "Trading zone: coordinating action and belief," in *The Science Studies Reader*, M. Biagioli, Ed., New York and London, Routledge, 1999, pp. 137 - 160.

BIOGRAPHY

Laura Jones-Wilson *earned her B.S. in Aerospace Engineering at Virginia Tech and went on to study at Cornell University, where she obtained her M.S. and Ph.D. in Aerospace Engineering specializing in Dynamics and Controls. She then joined JPL in 2012 as a guidance and control systems engineer, where she served as the project systems engineer for STABLE. She now serves as the PI a Mars Sample Capture technology development effort, the co-manager of the SmallSat Dynamics Testbed, and the instrument engineer for Europa-UVS as a member of the Europa Mission payload team*

Sara Susca *received her PhD in Electrical Engineering from UCSB in 2007. She spent three years at Honeywell Aerospace developing new technology for GPS denied navigation. She has been with JPL since 2011 where she covered various roles including Project Manager for STABLE (a balloon-borne sub-arcsecond pointing demonstration). She is currently member of the Europa Mission Payload team as the instrument engineer for EIS and PIMS instruments.*