DUNE Framework, Fermilab National Accelerator Laboratory (FNAL)

DUNE Stakeholder Requirements

(Approved, removed, or ready for approval)

Produced by Kyle Knoepfel, 2024/12/03

TABLE OF CONTENTS

1   DUNE-DUNE_STKH-1 Algorithm Decomposability

Requirement	The framework shall allow the execution of multiple algorithms.
Notes	This is ID #01 from the original DUNE document.
Status	Approved
Tags	General, Original

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-111 Algorithm Communication Via Data Products (Split Into)
- DUNE-DUNE_STKH-110 Algorithm-Data Separability (Split Into)

Comments:

2024/07/05 Kyle Knoepfel	Devs response: Suggested rewording: "The framework MUST allow the execution of decomposed units of processing (algorithms) that exchange transient data objects which are separable from those algorithms." !decisionneeded Re: @DUNE-DUNE_STKH-1
2024/07/05 Kyle Knoepfel	DUNE response: we accept this change

1.1   DUNE-DUNE_STKH-111 Algorithm Communication Via Data Products

Requirement	The Framework shall mediate communication between algorithms via data products
Status	Approved
Tags	General, Reproducibility, Provenance

Upstream Relationships:
- DUNE-DUNE_STKH-1 Algorithm Decomposability (Split Into)

Relationships:

Comments:

2   DUNE-DUNE_STKH-2 Data Product Representation

Requirement	The framework shall separate the persistent representation of Data Products from their in-memory representations as seen by algorithms.
Notes	This is ID #02 from the original DUNE document.

Status	Approved
Tags	Original, General, Accelerators

Upstream Relationships:
- DUNE-DUNE_STKH-7 Multiple persistent data formats (Related to)

Relationships:

Comments:

2024/06/12 Saba Sehrish	Devs response: The framework MUST not prevent reading and writing multiple persistent data formats.Re: @DUNE-DUNE_STKH-7
2024/06/12 Saba Sehrish	DUNE response: We should re-open a discussion internally about whether we need the FW to be able to convert between different persistent formats or not.
2024/06/12 Saba Sehrish	Devs response: Note: if the framework can read and write data using format A, and it can read and write data using format B, it can then read data using format A and write it using format B (and vice versa).
2024/07/26 Chris Green	Saba and Chris believe that this requirement is redundant when considered with @DUNE-DUNE_STKH-2, @DUNE-DUNE_STKH-79, and @DUNE-DUNE_STKH-110, and should be removed.!decisionneeded
2024/10/01 Kyle Knoepfel	After discussion with DUNE, the proposal is to remove this requirement after specifying the relationships to the stakeholder requirements in the previous comment.
2024/08/02 Amit Bashyal	Question: For algorithms that run in accelerators, is the data product that persist going to be accelerator friendly or the in memory one? Would the later give more flexibility in terms of design according to the task and compute power of machine? Re: @DUNE-DUNE_STKH-2
2024/06/11 Saba Sehrish	!question Devs response: Are you saying you want algorithm writers to not worry about the persisted representation of the data required by the algorithm? Or are you saying that the framework should be able to convert the persisted data representation to the in-memory

	representation that is optimal for the algorithm?Re: @DUNE-DUNE_STKH-2
2024/06/11 Saba Sehrish	DUNE response: We think yes to both. However we don't necessarily think the in-memory representation has to be optimal for the algorithms, and a possible solution would be a generic in-memory format that the algorithms (or perhaps support services) would have to deal with as appropriate.
2024/07/02 Chris Green	Devs response: Algorithms MUST depend only on the in-memory representation of their data and MUST not depend on any persistent representations. Algorithms MAY be used to transform data from one in-memory representation to another more suitable for further algorithm use.
2024/07/11 Peter van Gemmeren	This requirement is related to persistence
2024/07/26 Chris Green	After discussion, Saba and Chris believe that the above comments raise the question of whether a data product can/should have multiple in-memory representations, or whether they should be separate data products. However, because this impinges on the question of whether and how algorithms can access _some_ data from _many_ data products, that this matter is not yet ripe for disposition into specific requirements.!issue
2024/07/05 Chris Green	Devs response: The art framework (the UPS product "art") itself does not depend on ROOT or any other I/O system. Only the art-compatible ROOT I/O subsystem ("canvas_root_io") introduces that dependence. Is the desire to extend this independence to higher-level packages, containing DUNE reconstruction code and DUNE data product definitions?Re: @DUNE-DUNE_STKH-79
2024/07/05 Chris Green	DUNE response: The desire is to extend this independence to higher-level packages. Although we acknowledge that a component of this might be setting guidelines/policies for algorithms that run in the modules.
2024/07/05 Chris Green	Devs response: What IO technology independence does this imply other than what

	has already been requested in requirement 2 [@DUNE-DUNE_STKH-2]?
2024/07/05 Chris Green	Devs response: We suggest rewording into two requirements: "The framework's I/O subsystem MUST provide for schema evolution of data products." "The framework's I/O subsystem MUST provide for schema evolution of ancillary data." Note: The implication of this requirement with requirements 2 [@DUNE-DUNE_STKH-2] and 52 [@DUNE-DUNE_STKH-79] is that a schema evolution approach independent of the IO technology is necessary to support backwards compatibility. Is this what you're intending to ask for, or do we misunderstand the scope of these requirements? Re: @DUNE-DUNE_STKH-76
2024/07/05 Chris Green	DUNE response: We don't object to the requirements you proposed, but feel like our requirement was more narrowly focused on the need to support old data in conjunction with new data which may evolve over time. Perhaps we should discuss whether your suggested rewording meets that need, we think you may be drawing a distinction that we aren't fully understanding so want to check in. (We should add that if it became necessary to refer to a conversion process on old files to pass it to new versions of the framework, that would be acceptable, if not desirable)
2024/07/05 Chris Green	Devs response: We suggest rewording into two requirements: "The framework's I/O subsystem MUST support backward compatibility of data products." "The framework's I/O subsystem MUST must support backward compatibility of ancillary data." We should further discuss this in a meeting.

3   DUNE-DUNE_STKH-3 Modularity of Services

Requirement	The framework MUST be sufficiently modular in design to allow re-use of framework services (a) Some services must be able to be used directly by algorithmic code in a lightweight way, such as messaging and profiling, which can be controlled centrally across an entire framework job.
Notes	This is ID #03 from the original DUNE document.
Status	Removed
Tags	General, Original

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-60 Elapsed time information (Split Into)
- DUNE-DUNE_STKH-61 Framework-independent message logging (Split Into)
- DUNE-DUNE_STKH-59 Monitoring global memory use (Split Into)

Comments:

2024/06/12 Saba Sehrish	!question Devs response: We do not assume that a "service" will be an essential ingredient of the framework. Rather the specific request seems to be that the framework should provide logging facilities that can be used by algorithms, and that the framework should provide time- and memory-profiling information for those algorithms. Did you have something more general in mind? Can you reword this accordingly?Re: @DUNE-DUNE_STKH-3
2024/06/12 Saba Sehrish	DUNE response: Perhaps "facilities" would be a better term than "services" here. We want to ensure that useful features and facilities can be utilized by code that lives outside the FW. E.g. if an algorithm is typically run by the FW and utilizes various features (e.g., logging), it would be desirable for it to still be executable outside the framework.
2024/06/12 Saba Sehrish	Devs response: Even the term "facilities" is too general to form a meaningful requirement, as facilities vary widely in scope and purpose. Requirements must be placed on explicitly enumerated facilities. For logging, we propose the following: "The framework MUST adopt a

	logging solution that can be used in an algorithm without the algorithm explicitly relying on the framework." Please see requirements 65–67, which (given your approval) will replace this requirement 3. Note: requirement numbers 65-67 correspond to the original spreadsheet. Please see the associated requirements.
2024/07/03 Kyle Knoepfel	Original requirements 65-67 correspond to @DUNE-DUNE_STKH-59, @DUNE-DUNE_STKH-60, and @DUNE-DUNE_STKH-61.

4 DUNE-DUNE_STKH-7 Multiple persistent data formats

Requirement	The framework MUST not prevent reading and writing multiple persistent data formats.
Notes	This is ID #04 from the original DUNE document.
Status	Removed
Tags	General, Original, Persistency

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-110 Algorithm-Data Separability (Related to)
- DUNE-DUNE_STKH-79 Backend-agnostic data I/O API (Related to)
- DUNE-DUNE_STKH-2 Data Product Representation (Related to)

Comments:
2024/06/12 Saba Sehrish	Devs response: The framework MUST not prevent reading and writing multiple persistent data formats.Re: @DUNE-DUNE_STKH-7
2024/06/12 Saba Sehrish	DUNE response: We should re-open a discussion internally about whether we need the FW to be able to convert between different persistent formats or not.
2024/06/12 Saba Sehrish	Devs response: Note: if the framework can read and write data using format A, and it can read and write data using format B, it can then read data using format A and write it using format B (and vice versa).
2024/07/26 Chris Green	Saba and Chris believe that this requirement is redundant when considered with @DUNE-

	DUNE_STKH-2, @DUNE-DUNE_STKH-79, and @DUNE-DUNE_STKH-110, and should be removed.!decisionneeded
2024/10/01 Kyle Knoepfel	After discussion with DUNE, the proposal is to remove this requirement after specifying the relationships to the stakeholder requirements in the previous comment.

5   DUNE-DUNE_STKH-8 Full utilization of DUNE computing resources

Requirement	The framework shall run on widely-used scientific computing systems in order to fully utilize DUNE computing resources.
Notes	This is ID #05 from the original DUNE document.
Status	Approved
Tags	General, Original, Reproducibility

Upstream Relationships:

Relationships:

Comments:

2024/09/30 Barnali Chowdhury	This is persistence related and we should tag this requirement as "Persistence".Re: @DUNE-DUNE_STKH-8
2024/06/12 Saba Sehrish	!question Devs response: Does the requirement refer to computing sites (like HPCs and grid farms), specific hardware, and/or operating systems? If it is referring to sites, it would be helpful to have an enumeration (albeit incomplete) of what sites you're expecting in the next few years.Re: @DUNE-DUNE_STKH-8
2024/06/12 Saba Sehrish	DUNE response: We should check where else we mention running on hardware to avoid overlap (overlap with req. 19 – do they need to be separate). Certainly this refers to computing sites and different hardware. We want to check the policy on operating systems in DUNE, we're unsure if DUNE is generally planning to support one OS. Note: req. 19 refers to the requirement number 19 in the original google sheet.
2024/06/12 Saba Sehrish	Devs response: We will await your response.

2024/07/31 Kyle Knoepfel	Original requirement 19 corresponds to @DUNE-DUNE_STKH-29.

6   DUNE-DUNE_STKH-9 Algorithm hardware requirements

Requirement	The framework shall provide an API that allows users to express hardware requirements of the algorithms.
Notes	This is ID #06 from the original DUNE document.
Status	Approved
Tags	General, Original, Reproducibility

Upstream Relationships:

Relationships:

Comments:

2024/06/12 Saba Sehrish	!question Devs response: What types of hardware requirements are you referring to (e.g. number of CPUs, GPU resources, memory limitations)? And what should happen when those requirements cannot be met? For example, there could be a mechanism to specify hardware requirements in a way that the framework can tell you what is available, and you make the decision of what to do if the resource is not available. Is that the type of thing you have in mind?Re: @DUNE-DUNE_STKH-9
2024/06/12 Saba Sehrish	DUNE response: The hardware requirements you suggest are certainly what we had in mind, but there may be more that could be needed in future. We think that if the requirements aren't met, the FW should not run, and should report the problem as suggested. The hardware requirements for an algorithm might include OR.
2024/06/12 Saba Sehrish	Devs response: We then propose breaking this requirement into several: "(a) The framework MUST support running algorithms that require multiple CPUs (b) The framework MUST support running algorithms that require a GPU (c) The framework MUST provide an API so that users can express the CPU and GPU requirements of the algorithms (d) The framework MUST shut

	down and emit an appropriate diagnostic if one or more of these requirements is not met by the platform on which it is running."

6.1 DUNE-DUNE_STKH-10 Algorithms can use multiple CPUs

Requirement	The framework shall support running algorithms that require multiple CPUs.
Status	Approved
Tags	General, Reproducibility

Upstream Relationships:

Relationships:

Comments:

6.2 DUNE-DUNE_STKH-11 Algorithms can use a GPU

Requirement	The framework shall support running algorithms that require a GPU.
Status	Approved
Tags	General, Accelerators, Reproducibility

Upstream Relationships:
- DUNE-STY-8 Share GPU between jobs (Related to)

Relationships:

Comments:

2024/07/11 Amit Bashyal	This related to Accelerator Integration EffortRe: @DUNE-DUNE_STKH-11

6.2.1 DUNE-DUNE_STKH-42 Remote GPU algorithm support

Requirement	The framework shall support algorithms that perform calculations using a remote GPU.
Notes	This is ID #70 as proposed to DUNE.
Status	Ready for Approval
Tags	Services

Upstream Relationships:
- DUNE-DUNE_STKH-44 Thread-safety declarations of services (Split Into)

Relationships:

Comments:

2024/07/11 Amit Bashyal	Relates to Accelerator Integration EffortRe: @DUNE-DUNE_STKH-42
2024/07/03 Kyle Knoepfel	!question Devs response: [Original r]equirements 47-49 are very specific wrt. the concept of services. We'd like to understand the needs that motivate these requirements. One example is the ability for algorithms to consume data from sources outside of the framework's I/O system (e.g. external databases). Are there other examples you have in mind?Re: @DUNE-DUNE_STKH-35
2024/07/03 Kyle Knoepfel	DUNE response: This may hinge on the definition of service, so might require a conversation. Examples we were thinking of include: calibration databases etc (as you suggested); random number generators for the use of algorithms; potentially GPU access as a service; or, inference as a service (e.g., a service that runs a CNN and provides a classification for images on request)
2024/07/03 Kyle Knoepfel	Devs response: We would like to discuss with you replacing the concept of a service with simply a data-providing algorithm. Please see requirements 68–70 [@DUNE-DUNE_STKH-40, @DUNE-DUNE_STKH-41, @DUNE-DUNE_STKH-42], which (given your approval) replace requirements 47–49.

6.2.2   DUNE-DUNE_STKH-41 Local GPU algorithm support

Requirement	The framework shall support algorithms that perform calculations using a local GPU.
Notes	This is ID #69 as proposed to DUNE.
Status	Ready for Approval
Tags	Services

Upstream Relationships:
- DUNE-DUNE_STKH-44 Thread-safety declarations of services (Split Into)

Relationships:

Comments:

2024/07/11 Amit Bashyal	Relates to Accelerator Integration EffortRe: @DUNE-DUNE_STKH-41
2024/07/03 Kyle Knoepfel	!question Devs response: [Original r]equirements 47-49 are very specific wrt. the concept of services. We'd like to understand the needs that motivate these requirements. One example is the ability for algorithms to consume data from sources outside of the framework's I/O system (e.g. external databases). Are there other examples you have in mind?Re: @DUNE-DUNE_STKH-35
2024/07/03 Kyle Knoepfel	DUNE response: This may hinge on the definition of service, so might require a conversation. Examples we were thinking of include: calibration databases etc (as you suggested); random number generators for the use of algorithms; potentially GPU access as a service; or, inference as a service (e.g., a service that runs a CNN and provides a classification for images on request)
2024/07/03 Kyle Knoepfel	Devs response: We would like to discuss with you replacing the concept of a service with simply a data-providing algorithm. Please see requirements 68–70 [@DUNE-DUNE_STKH-40, @DUNE-DUNE_STKH-41, @DUNE-DUNE_STKH-42], which (given your approval) replace requirements 47–49.

6.3   DUNE-DUNE_STKH-13 Shut down upon unmet algorithm hardware requirements

Requirement	The framework shall shut down if the platform fails to meet each specified hardware requirement.
Status	Approved
Tags	General

Upstream Relationships:

Relationships:

Comments:

6.4 DUNE-DUNE_STKH-112 Emit diagnostic upon unmet algorithm hardware requirements

Requirement	The framework shall emit a diagnostic message for each hardware requirement the platform fails to meet.
Status	Approved
Tags	General

Upstream Relationships:

Relationships:

Comments:

7 DUNE-DUNE_STKH-14 Support for multiple programming languages

Requirement	The framework shall support the invocation of algorithms written in multiple programming languages.
Notes	This is ID #07 from the original DUNE document. If DUNE decides that additional languages should be supported in the future, a specific requirement can be added for that language as a sub-requirement.
Status	Approved
Tags	Original, General

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-16 Provide instructions for writing algorithms in supported languages (Related to)

Comments:

2024/06/12 Marc Paterno	!question Devs response: Do you mean modules (which is a framework term, related to connecting the framework with user code) or algorithms (which may also need to use framework API), or both? And what languages are you interested in?Re: @DUNE-DUNE_STKH-14

2024/06/12 Marc Paterno	DUNE response: We used modules to mean a FW compatible container that holds user provided code (algorithms). So we intended for it to be possible to write algorithms in a variety of languages, but are agnostic about the wrapper that forms the module. We certainly expect C++ and python algorithms, but there may be more in future. This implies that build instructions need to be distributed along with the FW, and it must be clear to the users how to build plug-ins – maybe related to a service agreement?
2024/06/12 Marc Paterno	Devs response: We then propose rewording this into two requirements: "The framework MUST support executing algorithms written in C++ and Python."

7.1   DUNE-DUNE_STKH-81 Support C++ algorithms

Requirement	The framework shall support the invocation of algorithms written in C++.
Status	Approved
Tags	General

Upstream Relationships:

Relationships:

Comments:

7.2   DUNE-DUNE_STKH-82 Support Python algorithms

Requirement	The framework shall support the invocation of algorithms written in Python.
Status	Approved
Tags	General

Upstream Relationships:

Relationships:

Comments:

8   DUNE-DUNE_STKH-16 Provide instructions for writing algorithms in supported languages

Requirement	The framework documentation shall provide instructions for writing framework-executable algorithms in supported languages.
Status	Approved
Tags	Documentation

Upstream Relationships:
- DUNE-DUNE_STKH-14 Support for multiple programming languages (Related to)

Relationships:

Comments:

9   DUNE-DUNE_STKH-17 Persist user-defined metadata

Requirement	The framework shall provide user-accessible persistence of user-defined metadata.
Notes	This is ID #08 from the original DUNE document.
Status	Approved
Tags	General, Original

Upstream Relationships:

Relationships:

Comments:

2024/06/12 Marc Paterno	!question Devs response: What do you mean by metadata? Are you referring to user-defined metadata (e.g. SAM information) that is kept in the same file as physics data? And what do you mean by store?Re: @DUNE-DUNE_STKH-17
2024/06/12 Marc Paterno	DUNE response: We give details on what metadata we think should be stored in the configuration/reproducibility sections (maybe we have overlap?). By store, we expect FW output files to contain the metadata that describes itself and the processes by which it was produced. Note the sections name correspond to the original spreadsheet.
2024/06/12 Marc Paterno	Devs response: We propose the following rewording: "The framework MUST provide tools

	(e.g. API or CLI) for storing and retrieving user-defined information that is ancillary to physics data."

10 DUNE-DUNE_STKH-19 Framework shall read its own output files

Requirement	The framework shall provide the ability to read a framework-produced file as input to a subsequent framework job so that the physics data are equivalent to the physics data obtained from a single execution of the combined job.
Notes	This is ID #10 from the original DUNE document.
Status	Approved
Tags	Physics Analysis, Original

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-20 Presenting data to subsequent algorithms (Related to)

Comments:

2024/06/14 Saba Sehrish	!question Devs response: Under what circumstances would you need to modify an existing file? Do you expect to need to read data from an output file configured for the current job? What are the difficulties in doing entire analyses within existing frameworks?Re: @DUNE-DUNE_STKH-19
2024/06/14 Saba Sehrish	DUNE response: [In progress] The intent here is that framework produced files could always be read into a subsequent job that would allow modifications to the event record and production of a new framework file. We anticipate that there will be algorithms that will need access to data written into output during a given job. For example, a clustering algorithm that uses hits generated in an upstage part of the job. DUNE's physics analysis framework is not currently well established and we would prefer that this framework remains an option for that. We will gather additional input from physics analyzers to help define some examples of where this is difficult with existing frameworks.

	Generally the issue is the frameworks like ART are designed with event by event analysis in mind while final physics analysis often requires summary data from samples like POT, toy MC throws, etc.
2024/06/14 Saba Sehrish	Devs response: Are the following rewordings more along the lines of what you are thinking? "The framework MUST be able to read a framework-produced file as input to a subsequent framework job so that the physics data are "indistinguishable" from the physics data that would be obtained from a single execution of both jobs." We specifically mention physics data to distinguish it from (e.g.) processing history, which would be able to detect multiple program executions. Also: "The framework MUST be capable of re-presenting (possibly re-reading from disk) data to the user from an already executed algorithm." We will await your further response.

11 DUNE-DUNE_STKH-20 Presenting data to subsequent algorithms

Requirement	The framework shall present data produced by an already executed algorithm to each subsequent, requesting algorithm.
Status	Approved
Tags	Physics Analysis

Upstream Relationships:
- DUNE-DUNE_STKH-19 Framework shall read its own output files (Related to)

Relationships:

Comments:

12 DUNE-DUNE_STKH-21 Mix input streams

Requirement	The framework shall support the creation of data sets composed of data products derived from data originating from disparate input sources.
Notes	This is ID #11 from the original DUNE document.
Status	Approved

Tags	Physics Analysis, Original

Upstream Relationships:
- DUNE-STY-2 Overlaying simulation and data (Related to)

Relationships:
- DUNE-DUNE_STKH-113 Algorithm invocation with data products from multiple data families (Related to)
- DUNE-DUNE_STKH-89 Algorithm invocation with data products from multiple data sets (Related to)

Comments:

2024/06/14 Saba Sehrish	!question Devs response: Are there other examples of mixing data from multiple input streams other than just overlaying simulation and real data? What types of synchronization requirements are there across input streams?Re: @DUNE-DUNE_STKH-21
2024/06/14 Saba Sehrish	DUNE response: Other examples include sim to sim overlay/mixing (rock+detector interactions, special rare process generations plus main sample, ND Genie with FD etc.). Other examples include processing of streams from multiple ND detectors or FD components (modules, APAs, etc). For synchronization, we imagine that the samples would have timing information that would need to be properly aligned (ex radiological sample with beam interactions). It not clear to us if this would be the purview of the framework or the algorithms.
2024/06/14 Saba Sehrish	!question Devs response: Thank you for the clear use cases. Are there things that art cannot do regarding this that you require? Or are there things that art can do that you think are unnecessary?

13 DUNE-DUNE_STKH-22 Flexible data units

Requirement	The framework shall support flexibly defined, context-aware processing units to address the varying granularity necessary for processing different kinds of data.
Notes	This is ID #12 from the original DUNE document.

Status	Approved
Tags	Original, Flexible Processing Unit (FPU)

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-86 Creation of data sets (Split Into)
- DUNE-DUNE_STKH-87 Definition of data families (Split Into)
- DUNE-DUNE_STKH-88 Definition of data family hierarchies (Split Into)
- DUNE-DUNE_STKH-85 Definition of data products (Split Into)

Comments:

2024/09/30 Barnali Chowdhury	This is equally applicable for persistence. The requirement should be tagged as "Persistence".Re: @DUNE-DUNE_STKH-22
2024/09/04 Kyle Knoepfel	Related to @DUNE-DUNE_STKH-22 Re: @DUNE-STY-5
2024/09/04 Kyle Knoepfel	Related to @DUNE-DUNE_STKH-22 Re: @DUNE-STY-4
2024/09/04 Kyle Knoepfel	Related to @DUNE-DUNE_STKH-22.Re: @DUNE-STY-3
2024/07/11 Amit Bashyal	Data models that touch the GPU related modules of the framework will also relate to the Accelerator Integration Effort. Re: @DUNE-DUNE_STKH-22
2024/08/07 Kyle Knoepfel	"Data model" for this requirement refers to the overall relationships among data products and not the in-memory or on-disk representations of those data products.
2024/07/03 Chris Green	Devs response: We suggest breaking this into four requirements. The framework MUST provide the ability for experiment code to define: (a) Data products (b) Data sets to which each data product belongs (c) Data families to which each data set belongs (d) Data family hierarchies Re: @DUNE-DUNE_STKH-22
2024/07/03 Chris Green	DUNE response: The restated requirements present seem to point to a generalized scheme that may meet the DUNE requirement. DUNE should develop concrete examples of different

	ways the DUNE data will be structured to hold up against this model, but it seems like a step beyond requirements.
2024/07/03 Chris Green	Do you agree enough with the proposed requirements that they can be adopted?!question As a reminder, all requirements are subject to future revision as necessary.

13.1 DUNE-DUNE_STKH-85 Definition of data products

Requirement	The framework shall provide the ability for user-level code to define data products.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-DUNE_STKH-22 Flexible data units (Split Into)

Relationships:

Comments:

13.2 DUNE-DUNE_STKH-86 Creation of data sets

Requirement	The framework shall provide the ability for user-level code to create new data sets.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-DUNE_STKH-22 Flexible data units (Split Into)

Relationships:

Comments:

13.3 DUNE-DUNE_STKH-87 Definition of data families

Requirement	The framework shall provide the ability for user-level code to define data families.
Status	Approved

Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-DUNE_STKH-22 Flexible data units (Split Into)

Relationships:

Comments:

13.4 DUNE-DUNE_STKH-88 Definition of data family hierarchies

Requirement	The framework shall provide the ability for user-level code to define hierarchies of data families.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-STY-4 Hit-level data for near detector (Related to)
- DUNE-STY-3 Spill-level data for near detector (Related to)
- DUNE-DUNE_STKH-22 Flexible data units (Split Into)

Relationships:

Comments:

14 DUNE-DUNE_STKH-23 FPU accessibility within a job

Requirement	The FPU of importance MUST be able to change within a single job, with a different granularity of information or different groupings of sub-units by different modules. For example, it must be possible to process entire ND spills, then call a process on a subset of the hits in that spill, then return to processing the full spill.
Notes	This is ID #13 from the original DUNE document.
Status	Removed
Tags	Original, Flexible Processing Unit (FPU)

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-113 Algorithm invocation with data products from multiple data families (Split Into)

- DUNE-DUNE_STKH-89 Algorithm invocation with data products from multiple data sets (Split Into)
- DUNE-DUNE_STKH-90 Specification of algorithm output FPUs (Split Into)

Comments:

2024/10/01 Barnali Chowdhury	Persistence should be able to write/read the entire TR or a subset of TR. Let's tag this as "Persistence".Re: @DUNE-DUNE_STKH-23
2024/07/03 Chris Green	Devs response: We suggest splitting this requirement into two: "The invocation of an algorithm MUST NOT be limited to accessing data products from a single data set or family," and "A user MUST be able to specify the data family to which an algorithm's created data products belong." Re: @DUNE-DUNE_STKH-23
2024/07/03 Chris Green	DUNE response: We discussed whether the specification of families would need to be set at run time or if establishing them via configuration is sufficient.
2024/07/03 Chris Green	Devs response: Let's discuss this in a meeting.

15 DUNE-DUNE_STKH-24 Data product I/O independence

Requirement	The framework shall support reading from disk only the data products required by a given algorithm.
Notes	This is ID #14 from the original DUNE document.
Status	Approved
Tags	Original, Data I/O layer

Upstream Relationships:

Relationships:

Comments:

2024/10/01 Barnali Chowdhury	This requirement should be tagged as "Persistence".Re: @DUNE-DUNE_STKH-24
2024/07/03 Chris Green	Devs response: We suggest rewording this requirement as: "The framework MUST support reading from disk only the data products of a dataset that are required for a given algorithm."

	If you also require the ability to read from disk a part of a persistable data product (to satisfy a different presentation of the data to algorithms than what was written), that should be a separate and explicit requirement. Re: @DUNE-DUNE_STKH-24
2024/07/03 Chris Green	DUNE response: We would like clarity on the definition of dataset in this statement.
2024/07/03 Chris Green	Devs response: We are using the term "data set" (we neglected to include the space between "data" and "set" earlier) in the same way we define it on the all-responses page: "A mathematical set of data products that is identifiable by the framework and used to determine which data products serve as inputs to an algorithm." Note that a data set is a member of a data family. An example of a data set in art would be a specific event. That said, we believe the concept of data set does not need to appear in this requirement, and we suggest a slightly different rewording: "The framework MUST support reading from disk only the data products required by a given algorithm." Does that help?!question

16 DUNE-DUNE_STKH-25 Process collections of unconstrained size

Requirement	The framework shall support processing of collections that are too large to fit into memory at one time.
Notes	This originates from ID #16 from the original DUNE document.
Status	Ready for Approval
Tags	Original, Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-STY-5 APA/CRP data for far detector (Related to)
- DUNE-STY-1 Time-ordered processing of DAQ data (Related to)

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: Are you requiring that the user have the ability to specify time-ordering of data for any number of user-specified algorithms, for all algorithms in the program, for the order in which data are written, or something else? !question If there is a requirement that data must be presented to an algorithm in a time-ordered manner, we believe that this may be ill-defined if the algorithm is concurrently invoked by multiple threads. Re: @DUNE-DUNE_STKH-25
2024/07/03 Chris Green	DUNE response: Example: need to process subsequent 3ms of raw data during signal processing to avoid edge artifacts. But, in other cases, framework should be allowed to process chunks of data in parallel
2024/07/03 Chris Green	Devs response: Are you asking that the framework present to the same invocation of the algorithm the "current" data as well as the subsequent 3ms of raw data? If not, what are you imagining the interface might look like for your case?!question
2024/08/07 Kyle Knoepfel	The developers need further information on the use cases motivating this requirement in order to refine it sufficiently to inform a design.
2024/10/28 Chris Green	Based on discussions at the workshop, and the resulting user stories, the developers came to the conclusion that the reference to time ordering came about as a result of thinking about how to handle the need to process data too large for memory constraints.

16.1 DUNE-DUNE_STKH-32 Read collections of unconstrained size

Requirement	The framework shall support the reading of collections too large to hold in memory.
Notes	This is ID #57 from the original DUNE document. Very large data collections could be 10s of GBs in size, and the memory budget may be a maximum of a few GBs.
Status	Ready for Approval
Tags	Original, Memory management

Upstream Relationships:
- DUNE-STY-5 APA/CRP data for far detector (Related to)
- DUNE-STY-1 Time-ordered processing of DAQ data (Related to)

Relationships:

Comments:

2024/10/01 Barnali Chowdhury	This also affects persistence and should be tagged as "Persistence". Re: @DUNE-DUNE_STKH-32

16.2 DUNE-DUNE_STKH-120 Write collections of unconstrained size

Requirement	The framework shall support the writing of collections too large to hold in memory.
Status	Ready for Approval
Tags	Original, Memory management

Upstream Relationships:

Relationships:

Comments:

17 DUNE-DUNE_STKH-26 Process neighboring FPUs

Requirement	The framework MUST be able to bring neighboring FPUs (where neighboring is context-dependent and may mean related in time or space) into the scope of a task. For example, one FPU may be 3 ms of a 6 ms readout window of one CRP, and would need to know that there is a relationship with a neighboring FPU, and there could be a hit that straddles the edge.
Notes	This is ID #15 from the original DUNE document.
Status	Removed
Tags	Original, Flexible Processing Unit (FPU)

Upstream Relationships:

Relationships:

- DUNE-DUNE_STKH-91 Algorithm invocation with data products from adjacent data sets (Split Into)
- DUNE-DUNE_STKH-92 User-defined adjacency (Split Into)

Comments:

2024/07/03 Chris Green	Devs response: We suggest two requirements: "The framework MUST be able to present to an algorithm data belonging to adjacent data sets, where adjacency is user-defined," and "The framework MUST support algorithms that define adjacency of data sets within a family." Re: @DUNE-DUNE_STKH-26

18 DUNE-DUNE_STKH-98 Configuration comparison

Requirement	The framework shall provide the ability to compare two configurations.
Status	Approved
Tags	Configuration

Upstream Relationships:

Relationships:

Comments:

19 DUNE-DUNE_STKH-29 Reproducible and deterministic results

Requirement	The framework MUST be capable of running across a wide range of present hardware (e.g. CPU architectures (ARM, ...), with GPU, TPU, co-processors, ...) in a reproducible and deterministic fashion. (a) This is not a requirement for identical bitwise output as that depends on the behavior of modules and services.
Notes	This is ID #19 from the original DUNE document. **DUNE is requested to approval the removal of this requirement.**
Status	Ready for Approval
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-99 Framework hardware independence (Split Into)

Comments:

2024/07/03 Chris Green	Devs response: For what aspects of the framework (as opposed to user code) do you require determinism and reproducibility? Part (b) is not phrased as a framework requirement. How about the following rewording: "The framework MUST NOT rely on unique characteristics of existing hardware." Re: @DUNE-DUNE_STKH-29
2024/07/03 Chris Green	DUNE response: [For the first question] NOTE - Review the note on this one. Perhaps there could be a unit test that the framework must pass on a specific architecture?? [For the reworded requirement] Agree.
2024/07/03 Chris Green	Devs response: Wikipedia defines a determistic algorithm as "an algorithm that, given a particular input, will always produce the same output, with the underlying machine always passing through the same sequence of states". Do you intend to distinguish between deterministic and reproducible? Are you asking that the control flow of the framework (i.e. the execution order of user-supplied algorithms) be implemented using only deterministic algorithms? Note that such a requirement will substantially hamper the ability to optimally execute user-provided algorithms. Re. part (b), we will list the reworded requirement separately from the issue of determinism and reproducibility.
2024/10/23 Chris Green	Upon discussion, the developers believe that this requirement as stated does not articulate something that *the framework* can ensure without consideration for the content of

	algorithms. We therefore request that DUNE approve its removal.!decisionneeded
2024/06/12 Saba Sehrish	!question Devs response: Does the requirement refer to computing sites (like HPCs and grid farms), specific hardware, and/or operating systems? If it is referring to sites, it would be helpful to have an enumeration (albeit incomplete) of what sites you're expecting in the next few years.Re: @DUNE-DUNE_STKH-8
2024/06/12 Saba Sehrish	DUNE response: We should check where else we mention running on hardware to avoid overlap (overlap with req. 19 – do they need to be separate). Certainly this refers to computing sites and different hardware. We want to check the policy on operating systems in DUNE, we're unsure if DUNE is generally planning to support one OS. Note: req. 19 refers to the requirement number 19 in the original google sheet.
2024/06/12 Saba Sehrish	Devs response: We will await your response.
2024/07/31 Kyle Knoepfel	Original requirement 19 corresponds to @DUNE-DUNE_STKH-29.

20 DUNE-DUNE_STKH-30 Record execution environment

Requirement	The framework shall record the job's execution environment.
Notes	This is ID #20 from the original DUNE document.
Status	Ready for Approval
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:
- DUNE-DUNE_STKH-101 Save user-selected execution environment components (Split Into)
- DUNE-DUNE_STKH-124 User-provided execution environment information (Split Into)
- DUNE-DUNE_STKH-100 User-selectable list of recordable execution environment components (Split Into)

Comments:

2024/07/03 Chris Green	Devs response: We propose two separate but related requirements:

	"The framework MUST provide a facility for the storage of execution-environment information provided by user code (as distinct from the evaluated configuration)," and "The framework MUST explicitly record the following execution environment information: <list>." Re: @DUNE-DUNE_STKH-30
2024/07/03 Chris Green	DUNE response: [For the first proposed requirement] If this means the FW will capture the environment that user code was run in.. then yes. [For the second proposed requirement] "The FW should allow the user to request that specified environmental info is saved."
2024/07/03 Chris Green	Devs response: We then suggest the following requirements instead: "The framework MUST specify a set of descriptions of the execution environment that may be saved." "The framework MUST save any of the execution-environment descriptions specified by the user."

20.1 DUNE-DUNE_STKH-100 User-selectable list of recordable execution environment components

Requirement	The framework shall provide the list of recordable components of the execution environment.
Status	Ready for Approval
Tags	Reproducibility, Provenance

Upstream Relationships:
- DUNE-DUNE_STKH-30 Record execution environment (Split Into)

Relationships:

Comments:

20.2 DUNE-DUNE_STKH-101 Save user-selected execution environment components

Requirement	The framework shall save each execution-environment description selected by the user from the framework-provided-list.
Status	Ready for Approval
Tags	Reproducibility, Provenance

Upstream Relationships:
- DUNE-DUNE_STKH-30 Record execution environment (Split Into)

Relationships:
- DUNE-DUNE_STKH-123 Record user-selected items from the shell environment (Split Into)

Comments:

20.2.1 DUNE-DUNE_STKH-123 Record user-selected items from the shell environment

Requirement	The framework shall record user-selected items from the shell environment.
Status	Ready for Approval
Tags	Provenance

Upstream Relationships:
- DUNE-DUNE_STKH-101 Save user-selected execution environment components (Split Into)

Relationships:

Comments:

20.3 DUNE-DUNE_STKH-124 User-provided execution environment information

Requirement	The framework shall record labelled execution environment information provided by the user.
Status	Ready for Approval
Tags	Provenance

Upstream Relationships:
- DUNE-DUNE_STKH-30 Record execution environment (Split Into)

Relationships:

Comments:

21 DUNE-DUNE_STKH-33 Unfolding data products

Requirement	The framework shall allow the unfolding of data products into a sequence of finer-grained data products.
Notes	This is ID #58 from the original DUNE document.
Status	Ready for Approval
Tags	Memory management, Original, Flexible Processing Unit (FPU)

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Kyle Knoepfel	!question Devs respose: Would satisfying this requirement meet the needs implied by [original] requirement 57?Re: @DUNE-DUNE_STKH-33
2024/07/03 Kyle Knoepfel	DUNE response: no, we need to be able to keep down the memory footprint regardless of the size of the FPU at whatever the relevant granularity in the job is. So satisfying this requirement may not be sufficient to satisfy 57.
2024/07/03 Kyle Knoepfel	Devs response: Let's discuss requirements 57 and 58 in person.
2024/10/16 Kyle Knoepfel	We should create a related system requirement that the IO API provide information to the IO backend enabling it to account for the presence/absence of the original data products (before unfolding) in the output.

22 DUNE-DUNE_STKH-34 Storage of configuration supporting reproducibility

Requirement	The framework shall support the storage of runtime configuration to enable the reproduction of physics output.
Notes	This is ID #21 from the original DUNE document. **We suggest the removal of this requirement (see comments).**
Status	Ready for Approval
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	The example provided does not seem related to the first sentence—the ability of the framework to "roll back" to a particular version of an algorithm is a separate issue from returning configuration and execution-environment information of that algorithm. At minimum, this requirement should be separated into two. If one accepts the proposition that, "conditions data are data in the same way that physics data are data," then it would follow that a "rollback date" represents a request (via user configuration) of an external system (e.g. a conditions database) for temporally-specified data. Re: @DUNE-DUNE_STKH-34
2024/07/03 Chris Green	DUNE response: This should be broken into 2 requirements 1) the FW can request the config/environment of modules and services and store them. 2) The information can then be handed back to the module/service later to attempt to setup things up in a reproducible manner.
2024/07/03 Chris Green	Devs response: Let's discuss this in a meeting.
2024/10/25 Chris Green	The example as stated in the original requirement, and the specific references to art-style "modules and services" were considered to be inapplicable to the DUNE framework as currently conceived. The clause referencing "execution environment" is addressed by https://fnal-prod.jamacloud.com/perspective.req?docId=14590&projectId=63 and related requirements.
2024/10/31 Kyle Knoepfel	This requirement duplicates @DUNE-DUNE_STKH-27 and we propose its removal.
2024/09/06 Kyle Knoepfel	This question arose when analyzing @DUNE-DUNE_STKH-34 Re: @DUNE-TXT-24

23 DUNE-DUNE_STKH-35 Access to external data sources

(Requirement statement on next page)

Requirement	The framework shall support access to external data sources.
Notes	This is ID #47 from the original DUNE document. By "external data sources," we mean "data sources **other than** framework-readable data files containing detector readout or simulated physics data."
Status	Ready for Approval
Tags	Original, Services

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Kyle Knoepfel	Refer to comments @DUNE-DUNE_STKH-35 Re: @DUNE-DUNE_STKH-44
2024/07/03 Kyle Knoepfel	Refer to comments @DUNE-DUNE_STKH-35 Re: @DUNE-DUNE_STKH-43
2024/07/03 Kyle Knoepfel	!question Devs response: [Original r]equirements 47-49 are very specific wrt. the concept of services. We'd like to understand the needs that motivate these requirements. One example is the ability for algorithms to consume data from sources outside of the framework's I/O system (e.g. external databases). Are there other examples you have in mind?Re: @DUNE-DUNE_STKH-35
2024/07/03 Kyle Knoepfel	DUNE response: This may hinge on the definition of service, so might require a conversation. Examples we were thinking of include: calibration databases etc (as you suggested); random number generators for the use of algorithms; potentially GPU access as a service; or, inference as a service (e.g., a service that runs a CNN and provides a classification for images on request)
2024/07/03 Kyle Knoepfel	Devs response: We would like to discuss with you replacing the concept of a service with simply a data-providing algorithm. Please see requirements 68–70 [@DUNE-DUNE_STKH-40, @DUNE-DUNE_STKH-41,

	@DUNE-DUNE_STKH-42], which (given your approval) replace requirements 47–49.

23.1 DUNE-DUNE_STKH-40 Calibration database algorithms

Requirement	The framework shall support algorithms that provide data from calibration databases.
Notes	This is ID #68 as proposed to DUNE.
Status	Ready for Approval
Tags	Services

Upstream Relationships:
- DUNE-DUNE_STKH-44 Thread-safety declarations of services (Split Into)

Relationships:

Comments:

2024/07/03 Kyle Knoepfel	!question Devs response: [Original r]equirements 47-49 are very specific wrt. the concept of services. We'd like to understand the needs that motivate these requirements. One example is the ability for algorithms to consume data from sources outside of the framework's I/O system (e.g. external databases). Are there other examples you have in mind?Re: @DUNE-DUNE_STKH-35
2024/07/03 Kyle Knoepfel	DUNE response: This may hinge on the definition of service, so might require a conversation. Examples we were thinking of include: calibration databases etc (as you suggested); random number generators for the use of algorithms; potentially GPU access as a service; or, inference as a service (e.g., a service that runs a CNN and provides a classification for images on request)
2024/07/03 Kyle Knoepfel	Devs response: We would like to discuss with you replacing the concept of a service with simply a data-providing algorithm.

Please see requirements 68–70 [@DUNE-DUNE_STKH-40, @DUNE-DUNE_STKH-41, @DUNE-DUNE_STKH-42], which (given your approval) replace requirements 47–49. |

24 DUNE-DUNE_STKH-38 Record workflow configuration and conditions

Requirement	The framework MUST support inclusion of external processing information.
Notes	This is ID #24 from the original DUNE document. **We suggest the removal of this requirement (see comments).**
Status	Ready for Approval
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: "The framework MUST be able to record external processing information." Note that requirements 17 and 18 already imply propagating the recorded parentage of the input files for the current framework job. Is this requirement intended to make that propagation explicit, or did you have something else in mind? Re: @DUNE-DUNE_STKH-38
2024/07/03 Chris Green	DUNE response: Yes, this was simply making that propagation explicit. We are happy with the suggested rewording.
2024/07/03 Chris Green	Devs response: We have added req. 61 that explicitly refers to propagating the provenance of input files. Do you think that the rephrased requirement here is then different from req. 8? If not, we could consider removing req. 24.
2024/07/03 Kyle Knoepfel	Req. 61 is included @DUNE-DUNE_STKH-62.

25 DUNE-DUNE_STKH-39 Algorithm code versioning and build information

Requirement	The framework shall have an option to record build information, including the source code version, associated with each algorithm.
Notes	This is ID #25 from the original DUNE document.
Status	Ready for Approval

Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:
- DUNE-DUNE_SYS-7 Recording user plugin source code information (Split Into)

Comments:

2024/07/15 Chris Green	It's not obvious to me that the downstream requirements properly capture the full intent of the original: we should consider during review.Re: @DUNE-DUNE_STKH-39
2024/07/03 Chris Green	Devs response: What does "itself" refer to? Does "code versioning" refer to including the versions of external software libraries used when processing the physics data? Regarding an interface that can be used by an algorithm—do you require just the ability to access the parentage within an algorithm (see our response to req. 18)? Note: it is technically feasible for the framework to automatically retrieve (e.g.) code versioning and compiler information for any plugin of which it is aware that has been built according to a specific prescription. Is this sufficient, or does the framework need to be able to receive and record arbitrary user-provided information of this type? Re: @DUNE-DUNE_STKH-39
2024/07/03 Chris Green	DUNE response: "itself" just means the core FW which is probably degenerate with previous requirements and could be reworded. We believe this information for modules needs to be available both during module execution, and stored in the FW output file. [Re. our note] The FW needs to be able to receive and record information from the modules, although we could discuss limitations to what the modules/plugins will provide.
2024/07/03 Chris Green	Devs response: Are you thinking along these lines?

	(a) "The framework MUST record an identifier encoding the source code version and build information for every algorithm executed in a framework job." (b) "The framework MUST provide a mechanism by which the source code version and build information identifier is associated with each algorithm."

26 DUNE-DUNE_STKH-43 Algorithms independent of framework interface

Requirement	The framework shall support the registration of algorithms that are independent of framework interface.
Notes	This is ID #48 from the original DUNE document.
Status	Ready for Approval
Tags	Services, Original

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Kyle Knoepfel	Refer to comments @DUNE-DUNE_STKH-35 Re: @DUNE-DUNE_STKH-43

27 DUNE-DUNE_STKH-59 Monitoring global memory use

Requirement	The framework shall be able to report the global memory use of the framework program at user-specified points in time.
Notes	This is ID #67 as proposed to DUNE.
Status	Approved
Tags	General

Upstream Relationships:
* DUNE-DUNE_STKH-3 Modularity of Services (Split Into)

Relationships:

Comments:

2024/06/12 Saba Sehrish	!question Devs response: We do not assume that a "service" will be an essential ingredient of the framework. Rather the specific request seems to be that the framework should provide logging facilities that can be used by algorithms, and that the framework should provide time- and memory-profiling information for those algorithms. Did you have something more general in mind? Can you reword this accordingly?Re: @DUNE-DUNE_STKH-3
2024/06/12 Saba Sehrish	DUNE response: Perhaps "facilities" would be a better term than "services" here. We want to ensure that useful features and facilities can be utilized by code that lives outside the FW. E.g. if an algorithm is typically run by the FW and utilizes various features (e.g., logging), it would be desirable for it to still be executable outside the framework.
2024/06/12 Saba Sehrish	Devs response: Even the term "facilities" is too general to form a meaningful requirement, as facilities vary widely in scope and purpose. Requirements must be placed on explicitly enumerated facilities. For logging, we propose the following: "The framework MUST adopt a logging solution that can be used in an algorithm without the algorithm explicitly relying on the framework." Please see requirements 65–67, which (given your approval) will replace this requirement 3. Note: requirement numbers 65-67 correspond to the original spreadsheet. Please see the associated requirements.
2024/07/03 Kyle Knoepfel	Original requirements 65-67 correspond to @DUNE-DUNE_STKH-59, @DUNE-DUNE_STKH-60, and @DUNE-DUNE_STKH-61.

28 DUNE-DUNE_STKH-60 Elapsed time information

Requirement	The framework shall have an option to provide elapsed time information for each algorithm executed in a framework program.
Notes	This is ID #66 as proposed to DUNE.

	This option is intended to capture wall-clock time and not CPU time. If more granular reporting of CPU vs. IO time is required, dedicated profiling tools like VTune or Linaro Forge should be used.
Status	Approved
Tags	General

Upstream Relationships:

- DUNE-DUNE_STKH-3 Modularity of Services (Split Into)

Relationships:

Comments:

2024/08/05 Kyle Knoepfel	This question arose when analyzing @DUNE-DUNE_STKH-60.Re: @DUNE-TXT-19
2024/06/12 Saba Sehrish	!question Devs response: We do not assume that a "service" will be an essential ingredient of the framework. Rather the specific request seems to be that the framework should provide logging facilities that can be used by algorithms, and that the framework should provide time- and memory-profiling information for those algorithms. Did you have something more general in mind? Can you reword this accordingly?Re: @DUNE-DUNE_STKH-3
2024/06/12 Saba Sehrish	DUNE response: Perhaps "facilities" would be a better term than "services" here. We want to ensure that useful features and facilities can be utilized by code that lives outside the FW. E.g. if an algorithm is typically run by the FW and utilizes various features (e.g., logging), it would be desirable for it to still be executable outside the framework.
2024/06/12 Saba Sehrish	Devs response: Even the term "facilities" is too general to form a meaningful requirement, as facilities vary widely in scope and purpose. Requirements must be placed on explicitly enumerated facilities. For logging, we propose the following: "The framework MUST adopt a logging solution that can be used in an algorithm without the algorithm explicitly relying on the framework." Please see requirements 65–67, which (given your approval) will replace this requirement 3.

	Note: requirement numbers 65-67 correspond to the original spreadsheet. Please see the associated requirements.
2024/07/03 Kyle Knoepfel	Original requirements 65-67 correspond to @DUNE-DUNE_STKH-59, @DUNE-DUNE_STKH-60, and @DUNE-DUNE_STKH-61.

29 DUNE-DUNE_STKH-61 Framework-independent message logging

Requirement	The framework shall support a logging solution that is usable in an algorithm without that algorithm explicitly relying on the framework.
Notes	This is ID #65 as proposed to DUNE.
Status	Approved
Tags	General

Upstream Relationships:
- DUNE-DUNE_STKH-3 Modularity of Services (Split Into)

Relationships:

Comments:

2024/06/12 Saba Sehrish	!question Devs response: We do not assume that a "service" will be an essential ingredient of the framework. Rather the specific request seems to be that the framework should provide logging facilities that can be used by algorithms, and that the framework should provide time- and memory-profiling information for those algorithms. Did you have something more general in mind? Can you reword this accordingly?Re: @DUNE-DUNE_STKH-3
2024/06/12 Saba Sehrish	DUNE response: Perhaps "facilities" would be a better term than "services" here. We want to ensure that useful features and facilities can be utilized by code that lives outside the FW. E.g. if an algorithm is typically run by the FW and utilizes various features (e.g., logging), it would be desirable for it to still be executable outside the framework.
2024/06/12 Saba Sehrish	Devs response: Even the term "facilities" is too general to form a meaningful requirement, as facilities vary widely in scope and purpose.

	Requirements must be placed on explicitly enumerated facilities. For logging, we propose the following: "The framework MUST adopt a logging solution that can be used in an algorithm without the algorithm explicitly relying on the framework." Please see requirements 65–67, which (given your approval) will replace this requirement 3. Note: requirement numbers 65-67 correspond to the original spreadsheet. Please see the associated requirements.
2024/07/03 Kyle Knoepfel	Original requirements 65-67 correspond to @DUNE-DUNE_STKH-59, @DUNE-DUNE_STKH-60, and @DUNE-DUNE_STKH-61.

30 DUNE-DUNE_STKH-64 Configuration validation

Requirement	The framework shall validate an algorithm's configuration against specifications provided at registration time.
Notes	This is ID #42 from the original DUNE document.
Status	Ready for Approval
Tags	Original, Configuration

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: By "method" are you just referring to a mechanism? What do you mean by additional imported configuration files? Is the requested behavior intended to extend to validating the configuration of third-party libraries (e.g. Geant4 configuration with GDML, WireCell configuration with Jsonnet)?!question Re: @DUNE-DUNE_STKH-64
2024/07/03 Chris Green	DUNE response: We are referring to a mechanism that would validate the configure (including all inherited configuration) at runtime that would then alert users to errors upfront rather than mid job. This would also be available as a standalone debug feature.

2024/07/03 Chris Green	Devs response: The art framework splits job initialization into 3 phases:
	(1) configuration syntax validation
	(2a) framework configuration schema validation
	(2b) configuration-specified libraries match the schema
	(3) All plugins can be initialized (e.g. all constructors succeed)
	The framework provides separate debugging facilities to check phase 1, phases 1+2, and phases 1+2+3. It sounds like you are asking for 1 and 2a. Is that correct?

31 DUNE-DUNE_STKH-66 Serialized algorithm configuration

Requirement	The framework MUST NOT allow a module to concurrently execute on more than one thread when the module configuration method is called.
Notes	This is ID #41 from the original DUNE document.
	We recommend the removal of this requirement as an algorithm is configured on only one thread.
Status	Ready for Approval
Tags	Original, Configuration

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: Is this an expression of concern for multi-thread safety during module construction, the idea of module reconfiguration, or something else? !question Re: @DUNE-DUNE_STKH-66
2024/07/03 Chris Green	DUNE response: This is expressing a need for thread safety during module construction.
2024/07/03 Chris Green	Devs response: During C++ object construction, only the thread on which construction is taking place has access to the object. If that was your concern, then we propose to strike this requirement.

32 DUNE-DUNE_STKH-67 Algorithm configuration schema availability

Requirement	The framework shall have an option to emit an algorithm's configuration schema in human-readable form.
Notes	This is ID #43 from the original DUNE document.
Status	Ready for Approval
Tags	Original, Configuration

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: Are you asking for the equivalent of the fhicl-cpp description API that art users place in their module code?Re: @DUNE-DUNE_STKH-67

33 DUNE-DUNE_STKH-69 One configuration per framework execution

Requirement	The framework shall accept exactly one configuration per program execution.
Notes	This is ID #44 from the original DUNE document.
Status	Ready for Approval
Tags	Original, Configuration

Upstream Relationships:

Relationships:

Comments:

2024/07/16 Marc Paterno	Devs response: Is this saying that an individual module needs to be able to specify that it is not reconfigurable (i.e. the same module object cannot be configured more than once in the same framework job)? !questionRe: @DUNE-DUNE_STKH-69
2024/07/16 Marc Paterno	DUNE response: yes, your reading is correct.
2024/07/16 Marc Paterno	Devs response: Suggested rewording of this requirement: "The framework MUST provide a mechanism to declare that an algorithm may be

	provided with an evaluated configuration only once." Note: This implies that the default behavior is that the framework should assume each algorithm is reconfigurable. Up until now, there has been no request for reconfiguration. If reconfiguration is desired, what are the use cases that motivate it? If reconfiguration is not required, requirement 44 is moot.

34   DUNE-DUNE_STKH-70 Eager validation of algorithm configuration

Requirement	The framework shall validate the configuration of each algorithm before that algorithm processes data.
Notes	This is ID #45 from the original DUNE document. Validation includes any reading, parsing, canonicalizing, and checking against applicable schemata.
Status	Ready for Approval
Tags	Original, Configuration

Upstream Relationships:

Relationships:

Comments:

2024/11/18 Kyle Knoepfel	Devs: Does this mean that you want to be able to specify a single configuration on the command line to start a program? (single point in user interface) Or does this mean that all configuration must be completed before data processing begins? (single point in time) Or is this getting at something else?!question Re: @DUNE-DUNE_STKH-70
2024/11/20 Kyle Knoepfel	From Ken Herner: "I think the second. You don't want to start processing something without parsing the entire config structure only to have it fail later because of some problem that you could have foreseen had you processed all of the configuration up front. That's my take on it anyway."

2024/11/20 Kyle Knoepfel	From Callum Wilkinson: "I agree with [Ken]. I think there was an additional concern that people wanted to capture with that requirement, that the configuration should be minimally magic, e.g., the user knows where all of their configuration is coming from. But that maybe is poorly framed as a requirement"

35 DUNE-DUNE_STKH-71 Discard unneeded configuration

Requirement	The framework SHOULD only initialize those modules specified to be run. Any additional configuration data for other modules not in the list to be run (e.g. something pulled in from including/importing various common config files) should be ignored.
Notes	This is ID #46 from the original DUNE document. **We suggest the removal of this requirement (see comment).**
Status	Ready for Approval
Tags	Original, Configuration

Upstream Relationships:

Relationships:

Comments:

2024/11/18 Kyle Knoepfel	See @DUNE-DUNE_STKH-71 comment.Re: @DUNE-TXT-30
2024/11/18 Kyle Knoepfel	We're unable to rephrase this as a functional requirement as the current wording suggests an implementation solving an unstated problem. We have tentatively reclassified this requirement as a design idea (https://fnal-prod.jamacloud.com/perspective.req?docId=15503&projectId=63).Re: @DUNE-DUNE_STKH-71

36 DUNE-DUNE_STKH-72 Framework configuration language

Requirement	The framework shall provide the ability to configure the execution of a framework program at runtime using a human-readable language.
Notes	This is ID #60 as proposed to DUNE.
Status	Approved
Tags	Configuration

Upstream Relationships:

Relationships:

Comments:

2024/08/07 Kyle Knoepfel	We believe this requirement is now redundant with @DUNE-DUNE_STKH-72.Re: @DUNE-DUNE_STKH-93

37 DUNE-DUNE_STKH-89 Algorithm invocation with data products from multiple data sets

Requirement	The framework shall allow a single invocation of an algorithm with data products from multiple data sets.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-STY-6 Different ND data groupings in the same job (Related to)
- DUNE-STY-4 Hit-level data for near detector (Related to)
- DUNE-STY-3 Spill-level data for near detector (Related to)
- DUNE-DUNE_STKH-23 FPU accessibility within a job (Split Into)
- DUNE-DUNE_STKH-21 Mix input streams (Related to)

Relationships:

Comments:

2024/09/09 Kyle Knoepfel	This user story is related to @DUNE-DUNE_STKH-89, @DUNE-DUNE_STKH-90, and @DUNE-DUNE_STKH-113, which are specific to algorithms. This user story, however, refers to features of the entire job, not just a given algorithm.Re: @DUNE-STY-6

37.1 DUNE-DUNE_STKH-91 Algorithm invocation with data products from adjacent data sets

Requirement	The framework shall support the invocation of an algorithm with data products belonging to adjacent data sets.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:

- DUNE-STY-5 APA/CRP data for far detector (Related to)
- DUNE-STY-7 Neighboring data groupings (Related to)
- DUNE-STY-1 Time-ordered processing of DAQ data (Related to)
- DUNE-DUNE_STKH-26 Process neighboring FPUs (Split Into)

Relationships:

Comments:

37.1.1 DUNE-DUNE_STKH-92 User-defined adjacency

Requirement	The framework shall support user code that defines adjacency of data sets within a data family.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-DUNE_STKH-26 Process neighboring FPUs (Split Into)

Relationships:

Comments:

38 DUNE-DUNE_STKH-113 Algorithm invocation with data products from multiple data families

Requirement	The framework shall allow a single invocation of an algorithm with data products from multiple data families.
Status	Approved
Tags	Flexible Processing Unit (FPU)

Upstream Relationships:
- DUNE-STY-6 Different ND data groupings in the same job (Related to)
- DUNE-STY-4 Hit-level data for near detector (Related to)
- DUNE-DUNE_STKH-23 FPU accessibility within a job (Split Into)
- DUNE-DUNE_STKH-21 Mix input streams (Related to)

Relationships:

Comments:

2024/09/09 Kyle Knoepfel	This user story is related to @DUNE-DUNE_STKH-89, @DUNE-DUNE_STKH-90, and @DUNE-

	DUNE_STKH-113, which are specific to algorithms. This user story, however, refers to features of the entire job, not just a given algorithm.Re: @DUNE-STY-6

39 DUNE-DUNE_STKH-99 Framework hardware independence

Requirement	The framework shall operate independently of unique characteristics of existing hardware.
Status	Approved
Tags	General

Upstream Relationships:
- DUNE-DUNE_STKH-29 Reproducible and deterministic results (Split Into)

Relationships:

Comments:

40 DUNE-DUNE_STKH-110 Algorithm-Data Separability

Requirement	The data objects exchanged among algorithms shall be separable from those algorithms.
Status	Approved
Tags	General

Upstream Relationships:
- DUNE-DUNE_STKH-1 Algorithm Decomposability (Split Into)
- DUNE-DUNE_STKH-7 Multiple persistent data formats (Related to)

Relationships:

Comments:

2024/06/12 Saba Sehrish	Devs response: The framework MUST not prevent reading and writing multiple persistent data formats.Re: @DUNE-DUNE_STKH-7
2024/06/12 Saba Sehrish	DUNE response: We should re-open a discussion internally about whether we need the FW to be able to convert between different persistent formats or not.

2024/06/12 Saba Sehrish	Devs response: Note: if the framework can read and write data using format A, and it can read and write data using format B, it can then read data using format A and write it using format B (and vice versa).
2024/07/26 Chris Green	Saba and Chris believe that this requirement is redundant when considered with @DUNE-DUNE_STKH-2, @DUNE-DUNE_STKH-79, and @DUNE-DUNE_STKH-110, and should be removed.!decisionneeded
2024/10/01 Kyle Knoepfel	After discussion with DUNE, the proposal is to remove this requirement after specifying the relationships to the stakeholder requirements in the previous comment.

41 DUNE-DUNE_STKH-116 ProtoDUNE single-phase raw data

Requirement	The framework ecosystem shall support processing ProtoDUNE single-phase raw data.
Notes	ProtoDUNE single-phase was used in run 1.
Status	Ready for Approval
Tags	Backwards compatibility, Data I/O layer

Upstream Relationships:

Relationships:

Comments:

42 DUNE-DUNE_STKH-117 ProtoDUNE dual-phase raw data

Requirement	The framework ecosystem shall support processing ProtoDUNE dual-phase raw data.
Notes	ProtoDUNE dual-phase was used in run 1.
Status	Ready for Approval
Tags	Backwards compatibility, Data I/O layer

Upstream Relationships:

Relationships:

Comments:

43 DUNE-DUNE_STKH-118 ProtoDUNE II horizontal-drift raw data

Requirement	The framework ecosystem shall support processing ProtoDUNE II horizontal-drift raw data.
Status	Approved
Tags	Backwards compatibility, Data I/O layer

Upstream Relationships:

Relationships:

Comments:

44 DUNE-DUNE_STKH-119 ProtoDUNE II vertical-drift raw data

Requirement	The framework ecosystem shall support processing ProtoDUNE II vertical-drift raw data.
Status	Approved
Tags	Backwards compatibility, Data I/O layer

Upstream Relationships:

Relationships:

Comments:

45 DUNE-DUNE_STKH-121 Provenance discovery

Requirement	The framework shall enable users to discover the provenance of data products.
Status	Ready for Approval
Tags	Provenance

Upstream Relationships:

Relationships:

Comments:

45.1 DUNE-DUNE_STKH-27 Framework configuration persistency

Requirement	The framework shall provide an option to persist the configuration of each framework execution to the output of that execution.
Notes	This is ID #17 from the original DUNE document. This requirement is in support of documenting and reproducing previous results.
Status	Approved
Tags	Original, Configuration, Reproducibility, Provenance

Upstream Relationships:

Relationships:
- DUNE-DUNE_SYS-1 Canonical form of configuration (Related to)

Comments:

2024/07/03 Chris Green	The example provided does not seem related to the first sentence—the ability of the framework to "roll back" to a particular version of an algorithm is a separate issue from returning configuration and execution-environment information of that algorithm. At minimum, this requirement should be separated into two. If one accepts the proposition that, "conditions data are data in the same way that physics data are data," then it would follow that a "rollback date" represents a request (via user configuration) of an external system (e.g. a conditions database) for temporally-specified data. Re: @DUNE-DUNE_STKH-34
2024/07/03 Chris Green	DUNE response: This should be broken into 2 requirements 1) the FW can request the config/environment of modules and services and store them. 2) The information can then be handed back to the module/service later to attempt to setup things up in a reproducible manner.
2024/07/03 Chris Green	Devs response: Let's discuss this in a meeting.
2024/10/25 Chris Green	The example as stated in the original requirement, and the specific references to art-style "modules and services" were

	considered to be inapplicable to the DUNE framework as currently conceived. The clause referencing "execution environment" is addressed by https://fnal-prod.jamacloud.com/perspective.req?docId=14590&projectId=63 and related requirements.
2024/10/31 Kyle Knoepfel	This requirement duplicates @DUNE-DUNE_STKH-27 and we propose its removal.
2024/08/07 Kyle Knoepfel	Chris and Kyle believe that since (a) the canonical configuration is persisted to an output file, and (b) there is only one canonical configuration associated with a given framework execution, the concept of configuration of versioning has no meaning here. A changed configuration will be associated with a different framework execution and the consequent storage of that canonical configuration. The framework would not know anything of any external management that provides the configuration files that serve as input to a given framework execution.Re: @DUNE-DUNE_STKH-27
2024/07/03 Chris Green	Devs response: Is this an example of the type of "metadata" that you would like to store in a framework output file (req. 8)? Are there additional things to what art already stores that you think you need? Does art store things that are not helpful? We propose breaking part (a) into separate requirements: (a) The framework MUST create a configuration. (What does this mean?) (b) The framework MUST tag a configuration. (What does this mean?) (c) The framework MUST check-sum a configuration. (d) The framework MUST store a configuration. (e) The framework MUST provide facilities to compare fully evaluated configurations. Re: @DUNE-DUNE_STKH-27
2024/07/03 Chris Green	DUNE response: [Re. (a)] The FW should be able to interpret configuration in some language and produce an internal memory representation??? [Re. (b)] This seems like it is the job of release managers to tag the configuration, not the framework?? [Re. (c)] The framework must produce a checksum from configuration that has been transformed into a canonical form. [Whitespace etc should not affect this…]
2024/07/03 Chris Green	Devs response: [Awaiting response to other questions.] Based on your responses we suggest the following: (a) Replaced by (f)–(i).

	(b) No longer relevant. (c) Stricken in favor of (j): (d) The framework MUST store a configuration. Replaced by (h). (e) The framework MUST provide facilities to compare fully evaluated configurations. Replaced by (k). (f) The framework MUST accept a configuration specification conforming to requirement 60. (g) The framework MUST provide an API allowing access to the evaluated configuration from user code. (h) The framework MUST enable the persistent storage of an evaluated configuration in canonical form. (i) The framework MUST provide the facility to convert a stored configuration to a representation conforming to requirement 60. (j) The framework MUST check-sum a configuration from its canonical form. (k) The framework MUST provide the facility to compare the canonical forms of two stored and/or req.60-conforming configurations.

45.2 DUNE-DUNE_STKH-28 Framework recording of metadata for reproduction of output data

Requirement	The framework shall record metadata to output enabling the reproduction of the processing steps used to produce the data recorded in that output.
Notes	This is ID #18 from the original DUNE document.
Status	Approved
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:
- DUNE-DUNE_SYS-10 Copy provenance information from input to output (Verified by)
- DUNE-DUNE_SYS-8 Framework recording of algorithm input metadata (Verified by)
- DUNE-DUNE_SYS-9 Framework recording of algorithmic origin of output data (Verified by)

Comments:

2024/07/03 Chris Green	Devs response: Does entry mean a data product or the entity that contains the data product (e.g. an event), or something else? Can you more specifically define what parentage is (req. 25 refers only to code versioning and compiler

	information)? Do you need to be able to retrieve and interrogate parentage within user code executed by the framework, and—if so—in what context (e.g. that in which the data are produced, later on, or if some other context, please specify)? We propose rewording part (a) as: "We MUST be able to redo the processing of the same input and/or intermediate data." Re: @DUNE-DUNE_STKH-28
2024/07/03 Chris Green	DUNE response: Re. definition of entry] This refers to the based processing unit. [Re. definition of parentage] Parentage should include information about software, configuration, version and compilers used to create output. [Regarding the rewording] This sounds like a reasonable suggestion.
2024/07/03 Chris Green	Devs response: We have added some definitions to the top-level responses page. Please see if these definitions are clear to you. We would classify as provenance what you list as examples of parentage. We will await your response on the question of when/where you need to retrieve and interrogate the provenance.

45.3 DUNE-DUNE_STKH-37 Framework data input and algorithmic parentage

Requirement	The framework MUST record the input and algorithmic parentage of each data product.
Notes	This is ID #23 from the original DUNE document. DUNE is asked to approve the removal of this requirement due to its replacement by Framework recording of algorithm input metadata and Framework recording of algorithmic parentage.
Status	Ready for Approval
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: "The framework MUST record the input and algorithmic parentage of each data product."Re: @DUNE-DUNE_STKH-37
2024/07/03 Chris Green	DUNE response: OK

46   DUNE-DUNE_STKH-122 Reproducibility of data products

Requirement	The framework shall support the reproduction of data products from the provenance stored in the output.
Status	Ready for Approval
Tags	Reproducibility, Provenance

Upstream Relationships:

Relationships:

Comments:

46.1   DUNE-DUNE_STKH-36 Reproducibility with pseudo-random numbers

Requirement	The framework shall provide a facility to produce random numbers enabling algorithms to create reproducible data in concurrent contexts.
Notes	This is ID #22 from the original DUNE document.
Status	Ready for Approval
Tags	Original, Reproducibility, Provenance

Upstream Relationships:

Relationships:

Comments:

2024/07/03 Chris Green	Devs response: There are counter-based random number generators (CBRNGs) that aim to provide deterministic outputs (e.g. http://www.thesalmons.org/john/random123/releases/latest/docs/index.html). Is this the type of facility that you would like the framework to provide access to? Or are you thinking of something else? Note that reproducibility cannot generally be provided for stateful random number generators (such as CLHEP's) without significant consequences on efficiency (often requiring serialization).Re: @DUNE-DUNE_STKH-36

2024/07/03 Chris Green	DUNE response: Yes. We would like to better understand the proposed implementation here…
2024/07/03 Chris Green	Devs response: We then suggest the following rewording: "The framework MUST provide a facility to produce random numbers so that the algorithms using them create reproducible data in concurrent contexts."

46.2 DUNE-DUNE_STKH-63 Independence from unique hardware characteristics

Requirement	The framework shall operate independently of unique characteristics of existing hardware.
Notes	This is ID #62 as proposed to DUNE.
Status	Ready for Approval
Tags	Reproducibility, Provenance

Upstream Relationships:

Relationships:

Comments: