

ID : 1

## **1 Introduction**

*Is Requirement* : No

ID : 2

### **1.1 Purpose of the Document**

*Is Requirement* : No

ID : 110

The purpose of this document is to specify stakeholder requirements for a demonstration science data processing and quality monitoring system for the Joint Dark Energy Mission (JDEM). JDEM is a collaborative project to build and operate a space telescope for the study of dark energy, and will be jointly funded and developed by NASA and the Office of High Energy Physics at the U.S. Department of Energy (DOE). The demonstration system will be developed at Fermilab for the JDEM Science Operations Center (SOC). The SOC is part of the JDEM Ground Data System. The demonstration system is referred to as the JDEM Demonstration Data Processing System (JDDPS).

*Is Requirement* : No

### **1.2 Scope of the Software**

*Is Requirement* : No

ID : 32

*Is Requirement* : No

ID : 111

*Is Requirement* : No

ID : 113

The scope of work includes the development of software for simulations, science data processing, and quality monitoring for slitless spectroscopy and NIR image processing. Simulations and science data processing will be implemented as prototype workflows that include rudimentary workflow management and provenance tracking capabilities. Workflow capabilities will be implemented by using the Kepler workflow system. This amounts to an evaluation of Kepler, which is therefore included in the scope of work. Also included in the scope is the evaluation of OpenSplice DDS message passing software for quality monitoring, and the use of databases for processing simulated slitless spectroscopy data. Each of these evaluations will be performed within the context of JDDPS.

*Is Requirement* : No

ID : 163

### **1.3 Stakeholders**

*Is Requirement* : No

ID : 164

This stakeholder requirements document for JDDPS is intended for the following stakeholders:

- DOE's JDEM Project Office,

- DOE's JDEM GDS team,
- DOE's JDEM Scientists,
- Fermilab Computer Security Team, and
- Fermilab Management.

It is worth noting that the JDEM Interim Science Working Group (ISWG) is not listed as a stakeholder, but members of the ISWG are included as stakeholders through the JDEM Project Office and as JDEM scientists and are therefore able to influence the development of JDDPS.

*Is Requirement* : No

*ID* : 170

DOE's JDEM GDS team and JDEM scientists include both software developers and users of JDDPS. The developers and users are organized according to roles that identify different types of functional requirements. We define four roles:

- Scientist algorithm developer - A scientist with specific expertise in developing and implementing algorithms for processing JDEM data.
- Scientist pipeline developer - A scientist with expertise in developing and implementing data processing pipelines to produce data products.
- Scientist data analyst - A scientist who uses data products to perform scientific analyses of data.
- Operator - A person who operates data processing pipelines.

*ID* : 171

Three of the roles are usually (but not exclusively) associated with work done by scientists, so we use the term "scientist" to define these roles. The "operator" role is often performed by a scientist, but we do not view this role as requiring much of a background in science. Other roles associated with software development, such as developing software infrastructure, are not included in this stakeholder requirements document, but are important for system requirements.

*Is Requirement* : No

*ID* : 4

## **1.4 Definitions, Acronyms, and Abbreviations**

*ID* : 160

The following definitions, acronyms, and abbreviations provide the meanings for terms used in this document.

*Is Requirement* : No

*ID* : 130

Actor - A workflow component that performs a task, typically by reading input and producing output.

*Is Requirement* : No

*ID* : 118

API - Application Programming Interface

*Is Requirement* : No

*ID* : 176

Application - Self-contained software that consists of one or more programs that perform a task. For example, the software that processes simulated slitless spectroscopy data

constitutes an application. Viewed in the context of JDDPS it is a program that is run by JDDPS.

*Is Requirement* : No

*ID* : 122

BAO - Baryon Acoustic Oscillations

*Is Requirement* : No

*ID* : 131

Campaign - A workflow initiated by a human.

*Is Requirement* : No

*ID* : 119

DDS - Data Distribution Service, a customizable quality of service publish/subscribe standard from the Object Management Group.

*Is Requirement* : No

*ID* : 124

DOE - Department of Energy

*Is Requirement* : No

*ID* : 120

GDS - Ground Data System

*ID* : 123

JDEM - Joint Dark Energy Mission

*Is Requirement* : No

*ID* : 177

JDDPS - JDEM Demonstration Data Processing System is an execution framework for running JDEM applications.

*Is Requirement* : No

*ID* : 132

Instantiated workflow - A workflow in which all data sources and participants, as well as their configurations, are specified.

*Is Requirement* : No

*ID* : 133

Job - A submission to a batch processing queue.

*Is Requirement* : No

*ID* : 125

NASA - National Aeronautics and Space Administration

*Is Requirement* : No

*ID* : 121

NIR - Near Infrared

*Is Requirement* : No

*ID* : 126

OMG - Object Management Group, a group that maintains standards for distributed, object-oriented software systems.

*Is Requirement* : No

Participant - An actor whose action is triggered by a workflow engine. From the point of view of the workflow engine, the task carried out by a participant is atomic in that the task completes successfully or fails to complete. Moreover, the workflow engine does not

manipulate the internal state of a participant. A participant might, for example, be a shell script that runs a data processing application to perform a specific task.

*Is Requirement* : No

*ID* : 178

Participant configuration - The configuration of a participant includes the components that are needed to run the participant. The configuration may include one or more of the following components: the program that is executed, settings for tunable parameters, and specification of the data to be processed.

*Is Requirement* : No

*ID* : 135

*Is Requirement* : No

*ID* : 158

Provenance - A record that identifies the configured participant and the unit of work that was responsible for creating a data product.

*Is Requirement* : No

*ID* : 136

Publisher - A software entity that prepares data for transmission based on one or more writers.

*Is Requirement* : No

*ID* : 127

QC - Quality Control

*Is Requirement* : No

*ID* : 179

Quality control products - Data products that are created for quality control purposes by simulation and data processing applications, and by processes that monitor the distributed computing environment. The purpose is for testing and determining the quality of science data products.

*Is Requirement* : No

*ID* : 128

QoS - Quality of Service, is the ability to provide different priorities to data flow in a network. QoS refers to control mechanisms that are used to reserve resources for different applications, users, or data flows, or to guarantee a certain level of performance.

*Is Requirement* : No

*ID* : 137

Reader - A software entity that reconstitutes an object from the data that has been received from a writer.

*Is Requirement* : No

*ID* : 129

SOC - Science Operations Center

*Is Requirement* : No

*ID* : 180

Software artifacts - Software that is entered in a repository so that different versions of the software can be managed over time. Examples of software artifacts include source code, configuration files, workflow specifications, documentation, test data and test scripts.

*Is Requirement* : No

*ID* : 138

Stream of data - A sequence of units of work of the same type.

*Is Requirement* : No

*ID* : 139

Submit node - A node that manages a job and can monitor other nodes involved in executing a job.

*Is Requirement* : No

*ID* : 140

Subscriber - A software entity that receives data using one or more readers.

*Is Requirement* : No

*ID* : 181

Test framework - Software that is part of the development environment that allows users to plug in unit-test and integration-test applications or scripts for automated testing.

*Is Requirement* : No

*ID* : 182

Test suite - The collection of unit tests and integration tests that are built using the test framework to validate algorithms and applications.

*Is Requirement* : No

*ID* : 141

Unit of work - The smallest data element that is processed in its entirety by an actor. Usually actors operate on a sequence of data elements.

*Is Requirement* : No

*Is Requirement* : No

*ID* : 142

Workflow - A collection of participants and a defined set of rules specifying when (under what conditions) and how (with what parameters, configuration, and input data) the actions performed by participants should be triggered.

*Is Requirement* : No

*ID* : 143

Workflow engine - A software application that manages and executes workflows.

*Is Requirement* : No

*ID* : 144

Workflow management system - A software application that triggers the actions performed by participants according to the rules that define the workflow. A workflow management system may additionally record provenance and other metadata about the execution of the workflow, and may provide tools for users to specify workflows.

*Is Requirement* : No

*ID* : 145

Workflow template - A workflow in which a subset of the data and/or participants are specified abstractly. At a minimum, a workflow template contains only the workflow rules, which refer to data and actors using undefined symbols.

*Is Requirement* : No

*ID* : 146

*Is Requirement* : No

*ID* : 5

## **1.5 References**

*Is Requirement* : No

*ID* : 94

This requirements document shall be used in conjunction with the documents listed in this section. URLs are provided for documents that are available. Documents that are not yet available are under review and will be released as soon as they become available. When documents are superseded by an approved revision, the revision shall apply.

*Is Requirement* : No

*ID* : 147

1) JDDPS Concept of Operations - currently under review

*Is Requirement* : No

*ID* : 148

2) QuIDS Concept of Operations - <http://cdcv.s.fnal.gov/redmine/documents/show/61>

*Is Requirement* : No

*ID* : 6

## **1.6 Overview of the Document**

*Is Requirement* : No

*ID* : 95

This document specifies stakeholder requirements for JDDPS. Section 2 provides a general description of JDDPS. It describes the relationship of JDDPS to other projects, functions and purpose, operating environment, and general constraints. Section 3 stipulates the requirements, and is organized into several subsections. The subsection titled "Functional Requirements" is organized according to the four roles defined in Section 1.3 for different types of software developers and users. Each role corresponds to a set of functions that will be provided by JDDPS.

*Is Requirement* : No

*ID* : 7

## **2 General Description**

*Is Requirement* : No

*ID* : 8

### **2.1 Relation to Current Projects**

*Is Requirement* : No

*ID* : 96

JDDPS benefits from its relationship to three projects at Fermilab and one project in which Fermilab is collaborating with Tech-X Corporation (<http://www.txcorp.com/>). The projects are LQCD, NOvA, QuIDS, and the FermiCloud project. The relationship of these projects to JDDPS is described in the following paragraphs.

LQCD is a project that uses the Lattice QCD (quantum chromodynamics) approach to solving theoretical problems in particle physics. LQCD faces significant computational challenges and has often been at the forefront of developing high-performance computing capabilities. At Fermilab the work has included a recent evaluation of workflow

management systems (see JDDPS ConOps document referenced in Section 1.4), and has influenced our approach to workflow management for JDDPS.

The NOvA project is a particle physics experiment that is being constructed to detect and analyze neutrino particles and their interactions. Members of the NOvA team have worked with the JDEM GDS team to evaluate specific implementations of the Data Distribution Service (DDS) message passing system. The NOvA team concluded that OpenSplice DDS satisfies their requirements for the NOvA data acquisition system, and has proceeded to implement functions using OpenSplice DDS software. The JDEM GDS team has not completed its evaluation, but will do so as part of the development of quality monitoring capabilities for JDDPS. If DDS satisfies JDEM requirements, then JDEM may benefit by using software developed for NOvA. For example, a message logger that is being developed for NOvA may be applicable to JDEM.

Members of the GDS team have been involved in the evaluation of a message passing system for quality control (QC) data. The name of the system is "Quality Information Distribution Service" (QuIDS), and it is based on DDS. QuIDS is being developed together with Tech-X, which received a Phase-1 SBIR grant to evaluate different DDS implementations and to develop a test system to measure performance limits of the messaging software. Having successfully completed Phase 1 of the SBIR, Tech-X has submitted a proposal for a Phase-2 grant to implement QC capabilities for JDEM. JDEM has contributed computing hardware to the FermiCloud project, which aims to investigate and implement "cloud computing" capabilities at Fermilab. The computing hardware will be used to run tests and demonstrations of JDDPS. The GDS team benefits from this collaboration since hardware installation and maintenance is managed by the FermiCloud project.

*Is Requirement* : No

*ID* : 9

## **2.2 Relationship to Predecessor and Successor Projects**

*Is Requirement* : No

*ID* : 97

JDDPS benefits from expertise of JDEM GDS team members with previous experience from the Sloan Digital Sky Survey (SDSS), a survey that shares many of the features of JDEM. Members of Fermilab's Experimental Astrophysics Group (EAG) have extensive experience with data collection, processing, distribution and analysis of SDSS data. The EAG hosted and operated those portions of SDSS science data processing infrastructure that corresponds to infrastructure that needs to be developed for JDEM. Furthermore, much of the SDSS infrastructure was developed by the EAG.

*ID* : 161

One of the important contributions that the EAG made to SDSS was the workflow system. The SDSS workflow system was adequate for its time and satisfied the requirements for SDSS science data processing. However, the system is outdated and specific to SDSS infrastructure. Therefore, it is not suitable for JDDPS or for JDEM.

*Is Requirement* : No

*ID* : 10

## **2.3 Functions and Purpose**

*Is Requirement* : No

*ID* : 100

The purpose of JDDPS is to provide an execution framework for simulations and science data processing and demonstrate science data processing capabilities for slitless spectroscopy and NIR imaging for JDEM. JDDPS functions will be developed to satisfy the goals identified in the JDDPS ConOps document referenced in Section 1.4. The goals are reiterated here:

*Is Requirement* : No

*ID* : 112

- Develop slitless spectroscopy simulations and data analysis capabilities for JDEM to support mission concept studies and assist in the development of science data processing algorithms, applications and software infrastructure.

*Is Requirement* : No

*ID* : 114

- Develop a prototype workflow for slitless spectroscopy and NIR science data processing.

*Is Requirement* : No

*ID* : 115

- Develop rudimentary workflow management and provenance tracking capabilities for the prototype workflow system.

*Is Requirement* : No

*ID* : 116

- Evaluate the OpenSplice implementation of the Data Distribution Service standard for data processing quality control.

*Is Requirement* : No

*ID* : 117

- Evaluate the use of databases for slitless spectroscopy and NIR science data processing.

*Is Requirement* : No

*ID* : 11

*2.4 Is Requirement* : No

The Fermilab Computing Division supplies and/or manages computing systems for user access, application development, production processing, and analysis of experimental data. The production processing infrastructure consists of large clusters (many thousands of computing cores) of either loosely or tightly interconnected worker nodes, shared among various experimental groups but allocated according to needs and dedicated investments. The FermiCloud infrastructure is currently being deployed to provide resources for virtualized application services in support of experimental data processing. JDEM has contributed computing hardware to the FermiCloud project, and JDDPS is being designed to operate on FermiCloud nodes.

*Is Requirement* : No

*ID* : 166

The Computing Division also supplies and/or operates data storage facilities and databases for experimental data. JDEM has purchased two database servers for JDDPS, and these will be managed by members of the Computing Division.

*Is Requirement* : No

*ID : 13*

*Is Requirement : No*

JDDPS does not involve any aspects beyond the usual and customary issues of running data processing systems at Fermilab. JDDPS will comply with Fermilab's Computer Security Policy. There are general constraints in that JDDPS must use existing hardware and services that exist at Fermilab. There are resource constraints in that users of JDDPS will share resources with other users at Fermilab when running on FermiCloud nodes.

*Is Requirement : No*

*ID : 15*

### **3 Specific Requirements**

*Is Requirement : No*

*ID : 152*

Requirements for JDDPS are presented in this section. We use the term "system" to refer to JDDPS. JDDPS provides the tools and the environment for developing and running JDEM applications. The term "production system" refers to an instance of a JDEM application running at Fermilab with all JDDPS functions implemented, such as quality control and provenance tracking. We use the term "development environment" to refer to software development tools that are used to develop, build and test software. We identify two development environments, the "pipeline development environment" and the "system development environment." The pipeline development environment is used by "scientist algorithm developers" and "scientist pipeline developers" (see definition of roles in Section 1.3) to develop and implement JDEM algorithms and applications, respectively. Scientists or software developers who need to use the tools at their home institution must satisfy requirements for use of the development environment, which are not included in this document. The second development environment is the system development environment, which is used by system engineers to develop software infrastructure for JDDPS.

*Is Requirement : No*

*ID : 16*

#### **3.1 Functional Requirements**

*Is Requirement : No*

*ID : 45*

##### **3.1.1 Scientist Algorithm Developer**

*Is Requirement : No*

*ID : 184*

The system shall have conformance to the currently approved JDEM application protocol.

*Is Requirement : Yes*

*ID : 153*

The system shall have a test framework to enable validation of participants.

*Is Requirement : Yes*

*ID : 173*

The system shall have a test suite that validates participants.

*Is Requirement* : Yes

*ID* : 154

Participants that satisfy the test suite shall be executable in a production system.

*Is Requirement* : Yes

*ID* : 77

### **3.1.2 Scientist Pipeline Developer**

*Is Requirement* : No

*ID* : 185

The system shall have conformance to the currently approved JDEM pipeline protocol.

*Is Requirement* : Yes

*ID* : 155

The system shall have a test framework to enable validation of pipelines.

*Is Requirement* : Yes

*ID* : 183

The system shall have a test suite that validates pipelines.

*Is Requirement* : Yes

*ID* : 156

Pipelines that satisfy the test suite shall be executable in a production system.

*Is Requirement* : Yes

*ID* : 186

Pipeline specifications shall describe abstract workflows independent of resource constraints.

*Is Requirement* : Yes

*ID* : 82

The system shall have a method for establishing valid dependencies between participants.

*Is Requirement* : Yes

*ID* : 83

The system shall provide the means to combine participants into applications (**should this be "pipelines or pipeline applications"?**).

*Is Requirement* : Yes

*ID* : 84

The system shall have a method for specifying the configuration of each participant.

*Is Requirement* : Yes

*ID* : 64

### **3.1.3 Scientist Data Analyst**

*Is Requirement* : No

*ID* : 76

A production system shall provide access to provenance for each data product.

*Is Requirement* : Yes

*ID* : 66

For each data product, the system shall provide a method for uniquely identifying all derived data products.

*Is Requirement* : Yes

*ID* : 44

### 3.1.4 Operator

*Is Requirement* : No

*ID* : 31

The system shall execute campaigns.

*Is Requirement* : Yes

*ID* : 87

The system shall provide the ability to execute multiple campaigns simultaneously.

*Is Requirement* : Yes

*ID* : 75

The system shall provide a method for configuring campaigns.

*Is Requirement* : Yes

*ID* : 35

The system shall provide current status information for campaigns.

*Is Requirement* : Yes

*ID* : 36

The system shall provide a list of all active campaigns.

*Is Requirement* : Yes

*ID* : 37

The system shall provide access to quality control products for an active campaign.

*Is Requirement* : Yes

*ID* : 38

The system shall provide access to scientific data products for an active campaign.

*Is Requirement* : Yes

*ID* : 39

The system shall provide the means for immediate termination of an active campaign.

*Is Requirement* : Yes

*ID* : 91

The system shall provide the means for orderly premature termination of an active campaign.

*ID* : 92

The system shall provide the means to restart a prematurely terminated campaign.

*Is Requirement* : Yes

*ID* : 93

The system shall retain data products created by a production system.

*Is Requirement* : Yes

*ID* : 40

The system shall provide remote active display of quality control products.

*Is Requirement* : Yes

*ID* : 61

The system shall provide a list of quality control products that are available for display.

*Is Requirement* : Yes

*ID* : 41

The system shall provide a method to change the selection of quality control products for display during an active campaign.

*Is Requirement* : Yes

*ID* : 67

The system shall provide a method to change the configuration of how quality control products are displayed during an active campaign.

*Is Requirement* : Yes

*ID* : 42

The system shall provide access to quality control products created by all previous campaigns executed by a production system.

*Is Requirement* : Yes

*ID* : 54

The system shall provide a method to request notifications of job failures.

*Is Requirement* : Yes

*ID* : 17

### **3.2 Performance Requirements**

*Is Requirement* : No

*ID* : 101

A production system shall process JDEM data that constitute a 24-hour period of data acquisition in less than 24 hours.

*Is Requirement* : Yes

*ID* : 18

### **3.3 Interface Requirements**

*Is Requirement* : No

*ID* : 175

This section specifies interface requirements for exchanging data with external systems. These requirements do not apply to internal interfaces in JDDPS, nor do they apply to exchanging data with database systems that are used by JDDPS.

*Is Requirement* : No

*ID* : 103

A production system shall be able to accept data as FITS files.

*Is Requirement* : Yes

*ID* : 172

A production system shall be able to write data products as FITS files.

*Is Requirement* : Yes

*ID* : 19

### **3.4 Operational Requirements**

*Is Requirement* : No

*ID* : 69

The system shall provide repositories for all software artifacts.

*Is Requirement* : Yes

*ID* : 70

The system shall provide access to repositories to all authorized collaborators.

*Is Requirement* : Yes

*ID* : 73

The system shall use validated releases for running a production system.

*ID* : 62

The system shall comply with policies established for Fermilab's operational environment.

*Is Requirement* : Yes

*ID* : 21

### **3.5 Verification Requirements**

*Is Requirement* : No

*ID* : 106

JDDPS is being developed using best practices encouraged by the Project Management Institute (<http://www.pmi.org/>). Requirements are being developed and managed using IBM Rational DOORS. DOORS is also being used for validation.

*Is Requirement* : No

*ID* : 105

The system shall have a test suite to validate requirements.

*Is Requirement* : Yes

*ID* : 22

### **3.6 Acceptance Testing Requirements**

*Is Requirement* : No

*ID* : 71

The system shall have a validation procedure for stakeholder requirements acceptance.

*Is Requirement* : Yes

*ID* : 24

### **3.7 Security Requirements**

*Is Requirement* : No

*ID* : 63

The system shall comply with policies established for Fermilab computing security.

*Is Requirement* : Yes

*ID* : 29

### **3.8 Safety Requirements**

*Is Requirement* : No

*ID* : 109

The system shall comply with Fermilab safety policies.

*Is Requirement* : Yes