



NextGRID Data Processing Profile

Editors:	Clive Davenhall	UEDIN

Date	Author	Comments	Version	Status
13 Aug. 2007	Clive Davenhall	Placeholder	0.0	Pre-draft
24 Sept. 2007	Clive Davenhall	Draft for comments	0.1	Draft
10 Oct. 2007	Clive Davenhall	Modified following review	0.2	Draft
14 Dec. 2007	Clive Davenhall	Removed references to EMS standards other than BES and JSDL; added mention of UDAP.	0.3	Draft
12 Feb. 2008	Clive Davenhall	Modified following review	0.4	Draft
1 Mar. 2008	Clive Davenhall	Final edits.	0.5	Draft
11 Mar 2008	D. Snelling	Final QA	1.0	Final



1 INTRODUCTION..... 3

1.1 Profile Overview 3

1.2 Relationships to Other Profiles..... 3

1.3 Relationship to Use Cases..... 4

 1.3.1 Terminology..... 4

1.4 Notational Conventions 5

1.5 Profile Identification and Versioning..... 5

2 PROFILE CONFORMANCE 5

2.1 Conformance Targets 5

2.2 Claiming Conformance 6

3 NEXTGRID DATA PROCESSING 6

3.1 Scope and Functionality 6

3.2 Mapping to WSRF 6

4 NEXTGRID BASIC PROFILE 7

5 NEXTGRID UDAP FRAMEWORK..... 7

6 OGSA BES 7

6.1 Naming..... 8

7 JSDL 8

7.1 Data Staging 8

7.2 Transfer of Staged Data 11

8 REFERENCES..... 12



1 Introduction

This document defines the NextGRID Data Processing Profile 1.0 (hereafter, “this Profile”), consisting of a set of de facto, institutional or evolving institutional Web Services specifications, along with clarifications, refinements, interpretations and amplifications of those specifications that promote interoperability among implementations of those specifications.

Section 1 introduces this Profile, and explains its relationships to other profiles.

Section 2, “Profile Conformance,” explains what it means to be conformant to this Profile.

Subsequent sections present the requirements for a NextGRID Data Processing Service and prescribe the Data Processing Profile in terms of extensions to NextGRID and other profiles.

This Profile forms part of the NextGRID Generalized Specifications, which aim at capturing NextGRID architectural concepts in a set of profiles that may be composed together. These profiles are specified in such a way that they could be implemented in terms of other well known specifications. While overall consistency is achieved at the conceptual level, and captured through the motivating Use Cases accompanying each Generalized Specification, system implementations, which may be based on other specifications, may not be consistent with this Profile. Thus, each profile defines a realisation of the underlying concept that can be implemented. However, implementers of the NextGRID architecture may need to support multiple underlying specifications.

1.1 Profile Overview

This Profile is intended for use when implementing data processing services that are in line with the concepts of NextGRID [1]. It mandates adherence to a certain set of specifications and clarifies their use. A service implementation that uses those specifications in a manner conformant with this Profile may be said to be an “implementation of the NextGRID Data Processing Profile 1.0” or, informally, to be a “NextGRID Data Processing Service”.

The primary issues addressed in this Profile are as follows:

- The specification of NextGRID Data Processing Services in terms of the OGSA-BES (Basic Execution Service) specification and the NextGRID UDAP (Universal Dynamic Activity Package) schema.
- The specification of the given task or ‘activity’ which a NextGRID Data Processing Service is to perform in terms of the JSDL (Job Submission Description Language) standard.

1.2 Relationships to Other Profiles

This Profile extends the *NextGRID Basic Profile* 1.0 [2].



NextGRID UDAP Framework

The UDAP (Universal Dynamic Activity Package) [3] framework provides the means to realise NextGRID Data Processing Services as UDAP services. UDAP services interact with components of the NextGRID environment in order to process data.

NextGRID Workflow Language Profile

NextGRID workflows conform to the *NextGRID Workflow Language Profile* [4].

NextGRID Data Transfer Profile

Data transfer occurs according to the *NextGRID Data Transfer Basic Profile* [5].

NextGRID Naming Profile

Service naming conforms to the *NextGRID Naming Profile* [6].

OGSA-BES

Every activity performing a data transformation is supervised by a NextGRID service. This supervisory service conforms to the OGSA-BES (Basic Execution Service) specification [7].

JSDL

The data transformation that an activity is to perform is specified using the JSDL (Job Submission description Language) specification [8].

1.3 Relationship to Use Cases

This Profile represents a mapping from the use cases described in the *NextGRID Data Processing Use Cases* [9] to the WSRF (Web Services Resource Framework) technology [10]. A data processing service is realised as a UDAP service [3] whose execution is superintended by a NextGRID service conforming to the BES (Basic Execution Service) specification [7]. Particular modifications of those specifications and profiles, as well as requirements on implementations, are prescribed in this Profile.

1.3.1 Terminology

This Profile adopts the terminology of the *NextGRID Data Processing Use Cases and Requirements* [9]. See that document, particularly Section 1.3 ‘Terminology’, for descriptions of

the terms pertaining to data processing used in this Profile. Hereafter terms defined in that section of that document are shown in **bold** in this Profile. See also the OGSA *Glossary of Terms* [11].

1.4 Notational Conventions

The keywords "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119 [12].

Normative statements of requirements in this Profile are presented in the manner detailed in the WS-I *Basic Profile* 1.1 [13], Section 2.1, "Conformance Requirements".

Both requirement statements and extensibility statements can be considered namespace-qualified. This specification uses a number of namespace prefixes throughout; their associated URIs are listed below. Note that the choice of any namespace prefix is arbitrary and not semantically significant.

Table 1 Namespaces used by NextGRID Data Processing Profile 1.0.

Prefix	Namespace
wsa	http://www.w3.org/2005/08/addressing
jsdl	http://schemas.ggf.org/jsdl/2005/11/jsdl
ng-dp	http://www.nextgrid.org/dataprocessingprofile/v-1

1.5 Profile Identification and Versioning

Profile identification and versioning uses the style described in the WS-I *Basic Profile* 1.1 [13] and abides by the normative descriptions contained therein. The name of this Profile is "NextGRID Data Processing Profile" and the version number is "1.0".

2 Profile Conformance

Conformance to this Profile is defined normatively in the WS-I *Basic Profile* 1.1 [13], Section 2, "Profile Conformance". This Profile abides by those definitions.

2.1 Conformance Targets

Since this Profile is an extension of the *NextGRID Basic Profile* 1.0 [2], which itself extends the WS-I *Basic Profile* 1.1 [13] and the *OGSA WSRF Basic Profile* 1.0 [14], it may place further restrictions on conformance targets defined therein.

The following conformance target is used in this Profile:

- **INSTANCE** – an instance of an implementation of a NextGRID Data Processing Service; such an instance necessarily implements a wsdl:port (from *WS-I Basic Profile 1.1*, without “bindingTemplate” from the namespace urn:uddi-org:api_v2).

2.2 Claiming Conformance

Claims of conformance to this Profile and the attachments mechanisms are the same as normatively described in the *WS-I Basic Profile 1.1* [13].

The conformance claim URI for this Profile is
“<http://www.nextgrid.org/dataprocessingprofile/v-1>”.

3 NextGRID Data Processing

3.1 Scope and Functionality

This section summarises the scope and functionality of a NextGRID Data Processing Service, as elucidated from the use cases and requirements [9] (see in particular Section 2 of that document, and also, for the terminology used, its Section 1.3). **Data processing** is the transformation of input data (typically from one or more sources) into output data (typically destined for one or more sinks). This transformation can be arbitrarily complex. In NextGRID the services which perform **data processing** are called **activities**.

An **activity** cannot be further decomposed within the context of the NextGRID architecture. However, **activities** can be **orchestrated**. That is, they can be linked so that they act in concert to perform some more complex task. Specifically a type of **orchestration** supported by NextGRID is the **workflow**, in which the constituent **activities** are marshalled according to a pre-defined pattern. NextGRID **workflows** are specified in the *NextGRID Workflow Language Profile* [4] and are not considered further in this Profile.

3.2 Mapping to WSRF

NextGRID services are specified as a set of services mapped to the WSRF (Web Services Resource Framework) set of specifications [10]. NextGRID Data Processing Services are mapped to WSRF using the OGSA (Open Grid Services Architecture) BES (Basic Execution Service) [7] and the NextGRID UDAP schema [3].

OGSA BES forms part of the OGSA EMS (Execution Management Services) [15] group of specifications. EMS is an architecture and not a standard (and thus forms part of the OGSA ‘informational’ rather than ‘recommendation’ track of documents and specifications). Specifically, it is an architecture for initiating, managing and controlling **activities** executing in a Grid

environment (for example, see *The Open Grid Services Architecture* [16], Section 5.4, p20). Thus, in the context of this Profile, EMS is an architecture for executing **activities**.

Within the EMS architecture BES prescribes services for superintending the execution of **activities**. The **activity** to be performed is initially described to the BES service using a **job description** expressed in the JSDL (Job Submission Description Language) format [8]. However, the **activity** which the BES service superintends is instantiated as a NextGRID UDAP service [3]. This UDAP service manages communication with other NextGRID services and performs the **activity**.

Thus, the sequence of events to instantiate an **activity** is as follows.

1. A BES service [7] is invoked with a JSDL [8] **job description** prescribing the **data processing** task to be performed.
2. The BES service instantiates a UDAP service [3] as the data processing service.
3. The UDAP service performs the **activity** and handles any necessary communication with other NextGRID services.

4 NextGRID Basic Profile

This section of the Profile incorporates the following specification by reference:

- *NextGRID Basic Profile* 1.0 [2].

All requirements stated in the *NextGRID Basic Profile* 1.0 must be observed by implementations of this Profile.

5 NextGRID UDAP Framework

This section of the Profile incorporates the following specification by reference:

- *UDAP Schema* 1.0 [3].

A UDAP service provides the functions described in the *UDAP Use Cases and Requirements* [17] and conforms to the schema presented in the *UDAP Schema* 1.0 [3]. An **activity** is instantiated as a UDAP service, which is superintended by a BES [7] service.

6 OGSA BES

This section of the Profile incorporates the following specification by reference:

- GFD.108: *OGSA Basic Execution Service*, 1.0 [7].

The clarification in the following section should be noted.

6.1 Naming

The BES specification states (Section 3, p6, first bullet point):

“All operations that generate EPRs (e.g. the ‘CreateActivity’ operation) can generate any EndpointReferenceType-compliant data type. That is, the data type MUST be compliant with WS-Addressing EndpointReferenceType, but MAY include profiled data in addition to the required elements, as is the case with WS-Names.”

Names used in NextGRID conform to the *NextGRID Naming Profile* [6], which is itself based on WS-Naming and WS-Addressing. Consequently, this Profile reiterates and extends the above criterion:

R0611 Any INSTANCE MAY be further annotated with NextGRID naming elements, in conformance with the NextGRID Naming Profile [6].

7 JSDL

The core JSDL (Job Submission Description Language) specification is defined in OGF (Open Grid Forum) GFD.56 [8]. Extensions for High-Performance Computing (HPC) processing are specified in GFD.111 [18] and for Single Program Multiple Data (SPMD; that is, parallel) processing in GFD.115 [19]. The core JSDL specification is the NextGRID mechanism for specifying **activities**. The JSDL HPC and SPMD extensions may be used within NextGRID if they are relevant for particular **activities**. Consequently, this section of the Profile incorporates the following specifications by reference:

- GFD.56: *Job Submission Description Language (JSDL)*, V1.0 [8],
- GFD.111: *JSDL HPC Profile Application Extension*, V1.0 [18],
- GFD.115: *JSDL SPMD Application Extension*, V1.0 [19].

It is likely that in the future the current extensions to JSDL for HPC and SPMD will be augmented with extensions for other specialised types of processing. It is hoped that this Profile will be compatible with these extensions, but clearly this cannot be guaranteed.

The JSDL HPC and SPMD extensions are incorporated without further extension. However, the following extensions to the JSDL core specification, GFD.56 [8], should be noted.

7.1 Data Staging

In order that an **activity** can perform **data processing** it needs to be able to input and output remote datasets. The core JSDL specification has facilities for ‘staging’ such datasets and making

them available so that processes running locally can access them. In JSDL, files are specified within the `<jSDL:DataStaging>` element using the `<jSDL:Source>` and `<jSDL:Target>` elements for input and output files, respectively:

```
<jSDL:DataStaging>

  <jSDL:Source>
    <jSDL:URI>http://blue.nowhere.com/somewhere/inputFile</jSDL:URI>
  </jSDL:Source>

  ....

  <jSDL:Target>
    <jSDL:URI>http://blue.nowhere.com/somewhere/outputFile</jSDL:URI>
  </jSDL:Target>

</jSDL:DataStaging>
```

In both cases the remote file is specified by the `<jSDL:URI>` element, whose contents are a URI, as defined in RFC 2396 [20]. This mechanism is inconsistent with NextGRID:

- NextGRID datasets are identified using an EPR; that is, strictly a `<wsa:EndpointReference>` element. Ultimately this element contains an IRI (Internationalised Resource Identifier), as defined in RFC 3987 [21]. URIs and IRIs differ (though there is a mapping between them).
- In NextGRID the `<wsa:EndpointReference>` element can be annotated with NextGRID naming elements [6].
- The actual transfer of the remote dataset to or from a local copy should be performed according to the NextGRID data transfer mechanisms [5].

Consequently, to make JSDL compatible with NextGRID, this profile extends the core JSDL specification [8] by defining the new elements `<ng-dp:Source>` and `<ng-dp:Target>` as follows:

- The `<ng-dp:Source>` element is used to specify a remote NextGRID dataset to be input to the **process**.
- The `<ng-dp:Target>` element is used to specify a local NextGRID dataset to be output to a remote location from the **process**.

The following rules apply:

R0711 An INSTANCE MUST support the `<ng-dp:Source>` element.

R0712 An INSTANCE MUST support the `<ng-dp:Target>` element.

The `<ng-dp:Source>` and `<ng-dp:Target>` elements each contain one and only one `<wsa:EndpointReference>` element, which can be annotated with NextGRID naming elements, as specified in the *NextGRID Naming Profile* [6].

```
<jSDL:DataStaging>

  <ng-dp:Source>
    <wsa:EndpointReference >
      ...
    </wsa:EndpointReference >
  </ng-dp:Source>

  ....

  <ng-dp:Target>
    <wsa:EndpointReference >
      ...
    </wsa:EndpointReference >
  </ng-dp:Target>

</ jSDL:DataStaging>
```

Though the `<ng-dp:Source>` and `<ng-dp:Target>` elements are necessary to utilise the NextGRID naming and data transfer functions, there may be occasions when it is more convenient for an **activity** to use the core JSDL `<jSDL:Source>` and `<jSDL:Target>` elements, and this behaviour is not prohibited:

R0713 An INSTANCE MAY use the `<jSDL:Source>` element.

R0714 An INSTANCE MAY use the `<jSDL:Target>` element.

7.2 Transfer of Staged Data

The transfer of remote datasets for input to an **activity** or output from an **activity**, must be performed according to the data transfer mechanisms specified in the *NextGRID Data Transfer Profile* [5]. Consequently, when staging a dataset specified using either of the `<ng-dp:Source>` and `<ng-dp:Target>` elements, the process must act as a sink or a source of data, respectively, in the context of the *NextGRID Data Transfer Profile*. Consequently:

R0721 An *INSTANCE MUST act as a sink of data in the context of the NextGRID Data Transfer Profile [5] when using the `<ng-dp:Source>` element.*

R0722 An *INSTANCE MUST act as a source of data in the context of the NextGRID Data Transfer Profile [5] when using the `<ng-dp:Target>` element.*

8 References

- [1] D. Snelling, M. Fisher, A. Basermann, F. Wray, P. Wieder and M. SurrIDGE, *NextGRID Vision and Architecture White Paper V5*, v18, NextGRID Project, 17 May 2007, http://www.nextgrid.org/download/publications/NextGRID_Architecture_White_Paper.pdf
- [2] V. Li, D. Snelling and A. Anjomshoaa, *NextGRID Basic Profile V1.0*, NextGRID Project, 2 January 2008.
- [3] A. Anjomshoaa and Ph. Wieder, *UDAP Schema*, V1.0, NextGRID Project, February 2008.
- [4] J. Ferris, B. Cantalupo and N. Matskanis, *NextGRID Workflow Language Profile*, V1.0, NextGRID Project, March 2008.
- [5] M. Drescher, S. Davey and A. Anjomshoaa, *NextGRID Data Transfer Profile*, V1.0, NextGRID Project, 19 February 2008.
- [6] S. Davey and D. Snelling, *NextGRID Naming Profile*, V1.0, NextGRID Project, 26 February 2008.
- [7] I. Foster, A. Grimshaw, P. Lane, W. Lee, M. Morgan, S. Newhouse, S. Pickles, D. Pulsipher, C. Smith and M. Theimer, GFD.108: *OGSA Basic Execution Service*, V1.0, Open Grid Forum, 8 August 2007.
- [8] A. Anjomshoaa, F. Brisard, M. Drescher, D. Fellows, A. Ly, S. McGough, D. Pulsipher and A. Savva, GFD-R.56: *Job Submission Description Language (JSDL)*, V1.0, Open Grid Forum, 7 November 2005.
- [9] C. Davenhall, *NextGRID Data Processing Use Cases and Requirements*, V1.0, NextGRID Project, 11 March 2008.
- [10] T. Banks. *Web Services Resource Framework (WSRF) – Primer v1.2*, Committee Draft 02, Organization for the Advancement of Structured Information Standards (OASIS), 23 May 2006. <http://docs.oasis-open.org/wsrf/wsrf-primer-1.2-primer-cd-02.pdf>
- [11] J. Treadwell, GFD-I.120: *Open Grid Services Architecture Glossary of Terms*, V1.6, Open Grid Forum, 12 December 2007.
- [12] S. Bradner, RFC 2119: *Key words for use in RFCs to Indicate Requirement Levels* (the Internet Engineering Task Force Best Current Practice), March 1997. <http://www.ietf.org/rfc/rfc2119>
- [13] K. Ballinger, D. Ehnebuske, C. Ferris, M. Gudgin, C. K. Liu, M. Nottingham, and P. Yendluri (eds.), *Basic Profile Version 1.1*, Web Services Interoperability Organization Final Material, 10 April 2006. <http://www.ws-i.org/Profiles/BasicProfile-1.1.html>
- [14] I. Foster, T. Maguire, D. Snelling, GFD-R-P.72: *OGSA® WSRF Basic Profile 1.0*, Open Grid Forum, 1 May 2006, <http://www.ogf.org/documents/GFD.72.pdf>
- [15] A. Savva, H. Kishimoto, S. Newhouse and D. Pulsipher, GFD-I.106: *OGSA® EMS Architecture Scenarios Version 1.0*, Open Grid Forum, 18 April 2007. <http://www.ogf.org/documents/GFD.106.pdf>

- [16] I. Foster, H. Kishimoto, A. Savva, D. Berry, A. Djaoui, A. Grimshaw, B. Horn, F. Maciel, F. Siebenlist, R. Subramaniam, J. Treadwell, J. Von Reich, GFD-I.80: *The Open Grid Services Architecture Version 1.5*, Open Grid Forum, 24 July 2006. See in particular Section 3.4, pp20-27. <http://www.ogf.org/documents/GFD.80.pdf>
- [17] A. Anjomshoaa, H. Mersch and Ph. Wieder, *UDAP Use Cases and Requirements*, V1.0, NextGRID Project, February 2008.
- [18] M. Humphrey, C. Smith, M. Theimer and G. Wasson, GFD.111: *JSDL HPC Profile Application Extension*, V1.0, Open Grid Forum, 28 August 2007. <http://www.ogf.org/documents/GFD.111.pdf>
- [19] A. Savva, GFD.115: *JSDL SPMD Application Extension*, V1.0, Open Grid Forum, 28 August 2008. . <http://www.ogf.org/documents/GFD.115.pdf>
- [20] T. Berners-Lee, R. Fielding and L. Masinter, RFC 2396: *Uniform Resource Identifiers (URI): Generic Syntax*, Internet Engineering Task Force, August 1998. <http://www.ietf.org/rfc/rfc2396>
- [21] M. Duerst and M. Suignard, RFC 3987: *Internationalized Resource Identifiers (IRIs)*, Internet Engineering Task Force, January 2005. <http://www.ietf.org/rfc/rfc3987>