# GRAVITE Transfer Protocol Support of the NPP Cal/Val Process

John Henson[1], J. Zajic [2], J. Feeley [3]
[1]General Dynamics Information Technology, [2]Integrity Applications Incorporated, [3]Aerospace Corporation

The NPP Government Resource for Algorithm Verification, Independent Test, and Evaluation (GRAVITE) Data Storage and Distribution (DSD) Segment currently hosts VIIRS, CrIS, and ATMS proxy data, and will be sized to hold up to 180 days worth of NPP mission data, ancillary data, auxiliary data, correlative data and match-up datasets post-launch. These data are made available to the Cal/Val community via the GRAVITE Transfer Protocol (GTP), designed to provide a single, consistent interface for data discovery and retrieval among one or more disparate information storage systems.

# DATA DISCOVERY

The GTP client allows users quick access to the vast storage catalog maintained by GRAVITE. Users can sort ascending and descending by the time the file was cataloged. Additionally, there are roughly 30 metadata fields which the user can filter on using a simple query string.

For example, to retrieve the 10 latest cataloged files in the system, the user would execute the following:

```
[myuser@mycomputer ~]$ gtp list newest 10
23534742 GCRSO_met_d20101206_t1448090_e1448390_b12345_c20101206171602729724_grav_dev.h5 0d723ca60a7942a8dd441e850f864920 46792
23534740 SVM02_ter_d20101206_t1445001_e1446100_b0000-1_c20101206171602399574_grav_dev.h5 bdbd75fbee3c83231806db3e8df97696 12399597
23534739 SVM02_aqu_d20101206_t1417338_e1418437_b45704_c20101206171601572722_grav_dev.h5 d0fe64f02e714acf68156eafb5688361 12399597
23534738 SVM13_aqu_d20101206_t1406557_e1408056_b45704_c20101206171600136183_grav_dev.h5 ed8b1f8662bcc1a2633cba434b56af62 22219616
23534737 SVM01_ter_d20101206_t1445001_e1446100_b0000-1_c20101206171559917171_grav_dev.h5 0294a5d29d9321428a9fa01281a45bd9 12399597
23534736 btpostcache_23533660_5a3e79cbbc7b5ec571be4757b8484469_1444864
23534735 btprecache_23533660_c47451e54715776b010abff35b043a2d_2888704
23534734 SVM01_aqu_d20101206_t1417338_e1418437_b45704_c20101206171559265868_grav_dev.h5 f5d0fb0c5474bb8546bd33f2dbfba80f 12399597
23534733 SVM02_aqu_d20101206_t1405448_e1406547_b45704_c20101206171559262026_grav_dev.h5 5638995563c508ccdf234bfc9c004ab9 12400141
23534732 SVI01_aqu_d20101206_t1326451_e1327550_b45704_c20101206171557714402_grav_dev.h5 aee89c67dbb6e806c91383607d4a3cf8 41892133
```

The GTP client also provides a more novel approach to data discovery: the "virtual directory" hierarchy system. A user can specify any order of metadata items as virtual directories in a virtual tree. The tree can be navigated using familiar commands such as "ls" and "cd." For example, to filter by "day" or "night" granules and of a specific type, the user would perform the following:

```
[myuser@mycomputer ~]$ gtp keyorder daynight,type
Key order: [daynight, type]
[myuser@mycomputer ~]$ gtp ls
night_to_day/
day_to_night/
night/
day/
[myuser@mycomputer ~]$ gtp cd night/SVM02_TERRA/
[myuser@mycomputer ~]$ gtp ls
SVM02_ter_d20101206_t1730265_e1731364_b00000_c20101206185902916012_grav_dev.h5 12400069 night 2010-12-06 05:30:26
SVM02_ter_d20101206_t1733592_e1735091_b00000_c20101206185854941224_grav_dev.h5 12400069 night 2010-12-06 05:33:59
SVM02_ter_d20101206_t1731374_e1732473_b00000_c20101206185845346601_grav_dev.h5 12400069 night 2010-12-06 05:31:37
SVM02_ter_d20101206_t1732483_e1733582_b00000_c20101206185838890470_grav_dev.h5 12400069 night 2010-12-06 05:32:48
SVM02_ter_d20101206_t1735101_e1736200_b00000_c20101206185808417771_grav_dev.h5 12400069 night 2010-12-06 05:35:10
SVM02_ter_d20101206_t1736210_e1737309_b00000_c20101206185413936723_grav_dev.h5 12400069 night 2010-12-06 05:36:21
SVM02_ter_d20101206_t1737319_e1738418_b00000_c20101206185406593529_grav_dev.h5 12400069 night 2010-12-06 05:37:31
SVM02_ter_d20101206_t1559270_e1600369_b00000_c20101206175901036114_grav_dev.h5 12400069 night 2010-12-06 03:59:27
SVM02_ter_d20101206_t1531053_e1532152_b00000_c20101206175229297859_grav_dev.h5 12400069 night 2010-12-06 03:31:05
SVM02_ter_d20101206_t1532162_e1533261_b00000_c20101206175200843092_grav_dev.h5 12400069 night 2010-12-06 03:32:16
SVM02_ter_d20101206_t1533271_e1534370_b00000_c20101206175007078470_grav_dev.h5 12400069 night 2010-12-06 03:33:27
SVM02_ter_d20101206_t1534380_e1535479_b00000_c20101206175006147167_grav_dev.h5 12400069 night 2010-12-06 03:34:38
SVM02_ter_d20101206_t1535489_e1536588_b00000_c20101206174933209275_grav_dev.h5 12400069 night 2010-12-06 03:35:48
...
SVM02_ter_d20101106_t1536523_e1538022_b00000_c20101106173730666297_grav_dev.h5 12400069 night 2010-11-06 03:36:52
SVM02_ter_d20101105_t0833187_e0834286_b00000_c20101106172551859535_grav_dev.h5 12400069 night 2010-11-05 08:33:18
SVM02_ter_d20101105_t0832078_e0833177_b00000_c20101106172530288743_grav_dev.h5 12400069 night 2010-11-05 08:32:07
SVM02_ter_d20101106_t1552141_e1553240_b00000_c20101106171809843544_grav_dev.h5 12400069 night 2010-11-06 03:52:14
```

# DATA RETRIEVAL

Any file available in the catalog may be retrieved using the client as well:

```
[myuser@mycomputer ~]$ gtp download file 23534740
23534740 [==========================================] 100%   3440.51 KB/s 00:03.604
```

The GTP client can be scripted or run from any common programming language or shell environment, thus enabling batch operations and downloads. Parallel downloads are also possible, permitting users to maximize their bandwidth.

## GRAVITE Metadata Catalog

- Metadata parameters extracted from ingested and generated data
- Parameters are maintained in data storage catalog
- Parameters are indexed and searchable
- Provides filtering capability for GTP

## Existing GTP Server Implementations

- GRAVITE DSD
  - Primary source for GRAVITE-stored data
  - Currently provides access to GRAVITE-generated NPP proxy SDRs and sets of spacecraft test data.
  - Data served since March 2010: 147.24TB
- SDS RIP Server
  - Hosts Retained Intermediate Products (RIPs) from IDPS for NASA Science Data Segment (SDS) retrieval.
- Archivist Server
  - Hosts RIPs for CLASS archival

## General Overview

The GTP system satisfies the need for users to easily and quickly discover and retrieve satellite instrument test data, ancillary data, auxiliary data, truth data, xDRs, etc. While the data may be stored in various places differentiated by system architecture, accessibility and physical location, the user may access these data in a consistent manner. As such, GTP provides a uniform interface to these systems thus reducing the learning curve and integration complexity, and ultimately increasing productivity.

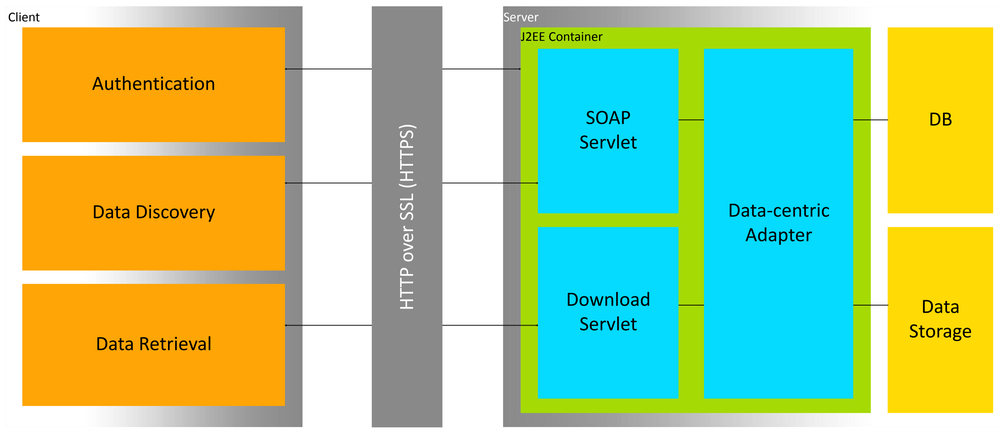
GTP takes advantage of Java's cross-platform compatibility, allowing it to run on many different systems with little code modification. (Since the GTP client interfaces entirely with the server using SOAP, the server could conceivably be written in any other language, such as ASP or Ruby.)

There are two primary uses of GTP. One is to provide a way to discover what data is available, and the other is the ability to actually retrieve that data and save it to a local disk.

As a command-line tool, the GTP client is intended to be used by individual users, shell scripts, compiled code and schedulers (e.g. the Linux cron daemon). The interaction between the client and server occurs between a daemon-like program called GTPProxy and the GTP server via standard SOAP web service requests.

The purpose of the GTPProxy program is to maintain state and authentication information while allowing the GTP client to run in a "one-off" fashion returning control to the command prompt after each request. This is the key to the GTP flexibility.

## High-Level Component Diagram




## GTP Client Capabilities

- Discovery of data in remote storage system
  - Show newest or oldest data
  - Customizable Date Formats
  - List view
    - Configurable output
    - Query by any metadata parameters maintained in the catalog
    - Simple query string language
  - Dynamic Virtual Directory Structure
    - User-defined virtual directory structure mimicking ftp command line interface
- Download of data from system
- Software is scriptable for automated lookup and retrieval of data

## GTP Server Capabilities

- Compatible with any standard relational database server through Java's JDBC interface.
- Core database structure required for GTP operation.
- Application-specific functionality may be incorporated
- On-demand data parsing, retrieval and generation